

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
FILOZOFICKÁ FAKULTA
ÚSTAV ANGLISTIKY

BAKALÁŘSKÁ PRÁCE

TALKING TO AN AI CHATBOT: EXPLORING THE NATURALNESS OF ASKING
MORALLY PROBLEMATIC QUESTIONS

Vedoucí práce: Mgr. Helena Lohrová Ph.D.

Autor práce: Nicole Taudy

Studijní obor: Anglický jazyk a literatura

Ročník: 3.

2024

I confirm that this thesis is my own work written using solely the sources and literature properly quoted and acknowledged as works cited.

V Českých Budějovicích dne 2.5.2024

.....
Nicole Taudy

Acknowledgement

I would like to express my gratitude to my supervisor Mgr. Helena Lohrová, Ph.D., for her patience, valuable insights, and thorough guidance throughout the process of writing this thesis. Her feedback and support have greatly helped shape the outcome of my work.

Anotace

Cílem této práce je ukázat, jak ChatGPT-3.5 zvládá morálně problematické otázky a jak se interakce celkově liší od přirozené konverzace mezi lidskými účastníky. Výzkum se zaměřuje na jevy, které jsou typicky přítomny v běžné interakci. Osvětluje tak několik způsobů, jakými interakce napodobuje lidské rozhovory, a také jevy, díky nimž interakce působí nepřirozeně. Analýza je provedena za pomoci dvou lingvistických metodologických přístupů – konverzační analýzy a analýzy diskurzu. Práce se rovněž zabývá možnými etickými problémy a důsledky, které vyplývají z pozorování provedených při analýze. Všechny poznatky jsou ilustrovány na jednotlivých vzorcích dat vytvořených a shromážděných na webových stránkách OpenAI pro ChatGPT-3.5.

Klíčová slova: umělá inteligence (AI), chatbot, ChatGPT, konverzační analýza, analýza diskurzu, etika

Abstract

This thesis aims to show how ChatGPT-3.5 navigates morally problematic questions and how the interaction overall differs from naturally occurring conversation between human participants. Focusing on phenomena that are typically present in regular interaction, the research illuminates several ways in which the interaction mimics human conversations as well as aspects which make the interaction feel unnatural. This is done with the combination of two linguistic methodological approaches – Conversation Analysis and Discourse Analysis. The thesis also discusses potential ethical concerns and implications that stem from the observations made in the analysis. All of the findings are exemplified on individual data samples generated and collected on OpenAI's website for ChatGPT-3.5.

Key words: Artificial Intelligence (AI), chatbot, ChatGPT, Conversation Analysis, Discourse Analysis, Ethics

Table of Contents

Introduction	1
1. Literature Review	4
1.1. Technical Aspects of ChatGPT	4
1.2. Ethics and Naturalness	5
1.3. Linguistic Methodology	6
2. What is ChatGPT?	9
2.1. Large Language Models (LLMs)	10
2.2. Neural Networks (NNs)	11
2.3. Training Phase	11
2.4. Evaluating ChatGPT-3.5 in Comparison to Other Chatbots	12
2.5. ChatGPT's Performance (GPT-3 vs GPT-3.5)	14
2.6. The Unnaturalness of the Interaction with ChatGPT	15
2.7. The Accountability of ChatGPT	16
3. Methodology	19
3.1. Conversation Analysis (CA)	20
3.1.1. CA Application	20
3.2. Discourse Analysis (DA)	21
3.2.1. DA Application	22
3.3. The Data	22
3.4. Data Collection	23
3.5. Observing the Data	24
4. Analysis	26
4.1. Turn-taking and Turn Structure	27
4.2. Repair	31
4.3. Cooperative Principle	31
4.4. Topic	32
4.5. Institutional Interaction	33
4.6. Identity and Power Dynamics	36
4.8. The Limitations of Human-Computer Interaction	39
4.9. Ethical Considerations	39
5. Discussion	44
Conclusion	46
Works Cited	48
Appendix: Data	52

Introduction

There have been rapid advancements in the field of Artificial Intelligence (AI) in recent years, which have brought forth a new era of human-computer interaction and communication. The technology revolutionized various aspects of our daily lives and is still finding new applications every day. One such AI domain which has seen a massive improvement is that of **chatbots** – virtual assistants constructed primarily for interaction with humans – be it in a casual setting or a professional one. ChatGPT, is among one of the most advanced chatbots deployed today. Unlike its predecessors, ChatGPT employs Natural Language Processing (NLP) merged with Neural Network (NN) learning, to engage with users in a strikingly natural and human-like interaction. The model of ChatGPT has been integrated into several platforms, from which other virtual conversational assistants have been created. All of them are capable of critical thinking, answering complex questions, offering advice, and sometimes even emulating emotion.

Although these innovations have redefined the boundaries of AI's linguistic capabilities, chatbots are, at least as of now, unable to form real opinions, or be partial in any way whatsoever. Unlike humans, who are able to discuss and express themselves on many different issues and take a definitive stance on matters, based on their subjective experiences and feelings, ChatGPT relies on structured algorithms and data patterns, to which its answers are ultimately confined.

With this an intriguing concern emerges – the ethical implications of interacting with these systems. They are incredibly valuable and efficient tools for information retrieval and assistance and are available for use to anyone who has access to the internet. Nevertheless, the AI's deficiency in emotional intelligence, common sense, and sensitivity to contextual nuances imposes limitations on its responses, particularly when confronted with complex or problematic

queries. Thus, this study explores how these AI systems navigate such questions, particularly within the realm of morality.

The main objective of this exploratory analysis is to employ the methodology of Discourse Analysis (DA) and Conversation Analysis (CA) to investigate the dynamics of asking ChatGPT questions which are commonly considered to be morally problematic. While CA is primarily used for the study of spoken human interaction, its application to AI offers valuable insights as to how users engage with these systems. DA, on the other hand, studies the contextual relations of text that is spoken, as well as written. By applying DA together with CA and by analyzing the structure of human-computer interaction, the study seeks to understand how the interaction takes place, how the chatbot manages certain topics and steers the direction of the conversation. In addition, the study will examine how the flow of the interaction feels and what it does in terms of effect. Moreover, it aims to explore the potential consequences and ethical dilemmas which arise from such exchanges, taking into consideration how certain topics are addressed and whether or not the model projects unwanted bias.

The application of CA and DA methodology is suitable for identifying any power imbalance between the AI and the user, uncovering potential bias and discrimination in the AI's responses, assessing the emotional and psychological impact on its users and much more. By examining the dynamics of human-AI conversations, developers can use this information to better understand the challenges and make informed decisions to ensure that their AI systems are designed to be used in an ethical and responsible manner.

Because this study approaches the issue from the linguistic perspective, it manages to highlight interesting facts from the language the AI uses, which can normally go unnoticed without careful inspection and analysis. The thesis therefore aims to contribute to the already existing knowledge concerning the moral and ethical implications of human-AI interactions,

thus potentially providing a deeper understanding of human engagement with this technology, while also exploring the linguistic phenomena present in said interactions.

The structure of this thesis is designed to provide a comprehensive exploration of the research topic chosen. Consisting of five chapters, each one builds upon the foundations described in the Introduction. Chapter 1 provides the Literature Review, going over cited works to demonstrate their relevance to the research topic. Chapter 2 delves into the language model of ChatGPT, detailing its inner workings and technical architecture, revealing how it manages its high level of coherence. Chapter 3 deals with the methodology of DA and CA employed for the analysis. Chapter 4 involves the actual data analysis, applying the theoretical framework of DA and CA, exploring the individual aspects and components within the data; the ending portion of Chapter 4 is dedicated the ethical implications stemming from the research. Lastly, Chapter 5 is a discussion of the findings and connecting them together while considering the naturalness of the interaction and any ethical implications stemming from it. The Conclusion and closing statements follow after, in which the key takeaways are summarized.

1. Literature Review

Because the emergence of AI technology and its improvements are still very recent, there is not a wide variety of sources to choose from on this topic, specifically from the linguistic point of view. Some authors have made textual analyses into ChatGPT and other conversational chatbots' generated responses, but few have yet utilized the methodology of CA, or other linguistic methodology for that matter. Many of the sources which examine AI's linguistic capabilities, look at the data from a technical perspective, with a focus on the inner workings of the program itself, as well as giving advice to developers on how to improve their code structure to better their models for future use. Some of the works which approach the textual analysis from this particular perspective also play a small part in this thesis, but only for the purpose of explaining the technical architecture of ChatGPT. Therefore, the majority of the observations made in this study will be supported by literature intended for analyzing regular (written and spoken) spontaneous human conversation.

1.1. Technical Aspects of ChatGPT

As has been mentioned in the previous paragraph, the architecture, and the inner workings of ChatGPT, such as the introduction of the Transformer architecture, which gave rise to its language capabilities, is explained to inform the reader of how the model processes language data, learns from it and manages its high levels of coherence. A number of various essays and research papers authored by university students of Computer Science programs and researchers from different parts of the world will be utilized in explaining ChatGPT's technical nature. For the most part, however, this thesis draws upon Michael McTear's *Conversational AI: Dialogue Systems, Conversational Agents and Chatbots*.

McTear provides a comprehensive introduction to Conversational AI and explores different approaches to the development of such dialogue systems – especially user-initiated

dialogue – which will help describe the process of exactly how chatbot models like ChatGPT work and are developed. In one chapter of his book, McTear focuses specifically on OpenAI’s ChatGPT 3. The GPT model, which is being examined in this thesis, however, has now been updated to version 3.5, therefore, information found on OpenAI’s website will be required to describe the current model of ChatGPT in exact detail, as there have been significant changes and improvements made. A short portion of McTear’s book is also dedicated to DA and the phenomena associated with it, where he looks at how turn-taking and reference is achieved in these chatbot systems. In addition to providing the technical aspects concerning Large Language Models (LLMs), McTear also considers future challenges as well as social and ethical issues surrounding dialogue systems, which developers need to consider for developing new models.

1.2. Ethics and Naturalness

There are many different ethical questions that come into light with new AI models constantly being developed at a fast rate. Issues such as environmental considerations connected to the energy consumption associated with high processing power, the question of employment and job replacements, the question of how to regulate these very sophisticated language models etc. For this thesis, the more important question is how to address the potential bias these language models may be promoting or how well they manage to inform users about sensitive topics. To address these issues, Paula Boddington’s *AI Ethics: A Textbook* as well as several other essays and research papers concerning the moral ethics of conversational systems provide the foundation for the thesis’ discussion dedicated to the ethical implications of interacting with AI chatbots.

Boddington opens her book with the reason for studying AI ethics as well as several ethical concerns pertaining to the development and use of AI in general, and that of chatbots.

She offers information on central concepts, questions and debates in AI ethics and explores how this plays out in practice. Her book also delves into topics such as whether or not it is permissible to use AI as moral assistants, talks about the very prospect of morality, as well as comparing AI to humans. Considering that the book was published in 2023, it offers a fresh perspective on matters and is up to date with current technology, while also providing an outlook on the issues relating to the future of AI and our lives with it.

In present time, models such as ChatGPT have become so advanced and already so extensively trained that they seamlessly appear to be human in their speech. Users will quickly realize this after interacting with ChatGPT for a while, as the way the chatbot ‘speaks’ is incredibly human-like and convincing. Users may notice certain patterns starting to occur in the generated answers, which resemble those of human speech. Because ChatGPT is trained on data produced by humans, the algorithms inevitably adopt language tendencies which are exhibited by humans as well. These aspects are examined in detail in the paper *Does ChatGPT resemble humans in language use?* by Cai et al. In the study, the authors took psycholinguistic experiments that have been used to test human participants and applied them to ChatGPT to see, if the model resembles humans in the comprehension and production of language. The experiments include **sound-shape association**, **sound-gender association**, **lexical retrieval**, **semantic illusion** etc. From their analysis, they found that ChatGPT replicated all but *two* of the twelve patterns which were derived from the psycholinguistic experiments. This reveals that the model is already adapted through so much data that blindly distinguishing it from a real human may prove troublesome.

1.3. Linguistic Methodology

For the linguistic part of this thesis, two methodological approaches are used: DA and CA. Many different authors who focus on CA contribute various valuable findings; Paul ten

Have with his *Doing Conversation Analysis* is a major asset in learning about CA methodology. He introduces the CA paradigm along with all its properties, its roots and history, as well as how to conduct analyses on data, giving the reader several analytic strategies to consider when beginning new research. While predominantly being a CA guidebook, one part of Chapter 10 in Have's book looks at examples of human-computer interaction.

Paul Drew and John Heritage's *Talk at Work* approaches CA from the perspective of institutionalized settings, where there is usually some kind of power imbalance and role distribution at play – conversations with ChatGPT also exhibit this unequal dynamic, in the sense that the user is the more powerful figure in the exchange. The user is the one who initiates the dialogue by asking a question or typing in some kind of input, with ChatGPT being the one obliged to answer. Another example of this unequal power dynamic display is that ChatGPT, in most cases, asks whether or not the user wishes to ask any other questions and awaits the next set of instructions. ChatGPT always explicitly states that it is “*there to assist*” putting the user in a superior position. With Drew and Heritage's book, where they include the contributions of many other prominent linguists – such as Emanuel Schegloff, Stephen C. Levinson, Gail Jefferson, and several others – they discuss the many different aspects which can be observed in conversations taking place in institutionalized settings. Topics such as neutrality display, management of disagreement, delivery of advice etc. are discussed, all of which are very relevant for this thesis.

Rebecca Clift's *Conversation Analysis*, and Jack Sidnell's *Conversation Analysis: An Introduction*, much like Have's work, give a general overview of CA and its significance. Both Clift's and Sidnell's books focus on the different phenomena which CA analyzes. **Adjacency pairs, turn-taking, turn construction, preference organization**, and many more. They present the reader with data, along with reported research, allowing students of linguistics to

pursue a research project of their own. All of these contributions are a means of providing accurate descriptions of said phenomena found in the data samples.

CA does not focus on written communication but rather on casual spoken conversation which happens daily and spontaneously. Because this thesis works with data, which was produced in written form, it is necessary to extend the methodological scope with DA.

The Handbook of Discourse Analysis II edited by Deborah Tannen et al., is, like its title implies, a compilation of publications by various linguists, overviewing the major topics pertaining to DA. In this second volume, aside from giving the basics of linguistic analyses of discourse and various methodological approaches, the book focuses on the actual implications of the findings stemming from DA. The individual societal and cultural issues, as well as real-world scenarios of discourse are included. One of the chapters is also dedicated to computer-mediated discourse, which serves as a good starting point for the DA analysis.

The book *Textual Interaction: An introduction to written discourse analysis* by Michael Hoey, examines written text as a site for interaction and views text from a larger perspective. Hoey looks at text mainly from a narrative perspective but also examines the structure of text, as well as the lexical choices and the potential reasoning associated with it. This further helps describe ChatGPT's 'thought-process' during text generation.

2. What is ChatGPT?

ChatGPT-3.5 is a free-to-use language model built upon on the GPT (Generative Pretrained Transformer¹) model architecture, developed by OpenAI² and first released in November 2022³. Its primary purpose is Natural Language Processing (NLP)⁴ and text generation, making it a sophisticated conversational system – to be put simply, a chatbot.

ChatGPT has the ability to engage in multi-faceted conversations, answer questions, summarize large amounts of information, translate texts into several different languages as well as communicate in different languages, and generate coherent and contextually relevant text based on user input prompts.

Today, many different developers and companies use ChatGPT's model as a code reference and even as a model base to create their own chatbot assistants, fine-tuning the program to fit their work scenarios and specific needs. As a result of this, these conversational systems have become even more widespread in many different varieties. Therefore, it is likely that most readers will probably have already encountered and may even have interacted with an LLM like ChatGPT, but fewer of them will know how these systems actually work and how they are created. More and more initiatives encourage learning about these systems so that users know exactly what they are engaging with and how to appropriately use these systems to their benefit (Brittin). One might think that because these systems are so advanced that taking caution in how a person approaches the interaction is unnecessary. However, despite appearances, these systems still have some way to go before reaching a point of being impeccable. Certain contextual nuances and subtle details will go amiss the chatbot's understanding, which will require users to adjust their prompts accordingly.

¹ the Transformer architecture was originally introduced in the paper "Attention is All You Need" by Vaswani et al.

² American artificial intelligence research organization (<https://openai.com>)

³ <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>

⁴ <https://www.ibm.com/topics/natural-language-processing>

It is important to note that ChatGPT does not have real-time access to databases and cannot browse the internet freely⁵. Its knowledge is limited only to the training data it is provided with. This has sprung some concerns about whether the information ChatGPT provides is usage-safe and reliable, given the fact that it is incapable of fact-checking its answers in real time. In addition to lacking personal experience, opinions and feelings, the model doesn't 'think' like humans do, it rather only 'understands' (Schaffner Bofill). The model does not employ any cognitive function or reasoning to ponder what something means – like humans do when discussing certain topics. This striking difference is exactly what makes the interaction fascinating to study.

To understand the analysis to its full extent it is crucial to have some general knowledge about the internal architecture of ChatGPT.

2.1. Large Language Models (LLMs)

LLMs are a class of deep learning AI which are formed from Transformer Neural Networks⁶; a type of AI of which ChatGPT is itself a part of. This category of AI is trained on vast amounts of language data, mainly from sources such as articles, books, websites, forums, and many other which usually cover a plethora of topics. ChatGPT, specifically, was trained using several extensive corpora, web datasets as well as the English language version of Wikipedia (McTear, 144).

LLMs are given numerous parameters which then make them capable of understanding and generating very sophisticated text⁷. All of this is done with complex algorithms and statistics that allow the LLM to produce an answer once prompted by the user's query. Many different LLMs are of varying sizes – the largeness of a model is most often determined by the

⁵ <https://help.openai.com/en/articles/6783457-what-is-chatgpt>

⁶ <https://www.nvidia.com/en-us/glossary/large-language-models/>

⁷ <https://www.codecademy.com/article/setting-parameters-in-open-ai>

number of parameters assigned to it. The more parameters a model has, the more capable it will be⁸. This means that the number of parameters given ultimately determines the language model's competence in decoding meaning from the given input as well as its ability to capture complex patterns in the data it receives. To put this into perspective, ChatGPT-2, the early predecessor of the current model, had 1.5 billion parameters and could perform basic language tasks, the most recent predecessor, GPT-3, had around 175 billion parameters (McTear, 144). Now, GPT-3.5 has gained 200 billion more parameters (Chude). Thanks to this, GPT-3.5 is able to learn difficult tasks from only a few examples (144).

2.2. Neural Networks (NNs)

The core part of any LLM is the Neural Network (NN). One might call it the 'brain' of each and every chatbot, as the NN's significance lies in its property of granting language models their ability to learn complex patterns from large amounts of language data (Jones).

NNs act as powerful feature extractors, that can overtime and with ample data automatically learn relevant features from the text, such as **word embeddings** – this means that “each word is represented by a unique real-number vector, that captures its meaning and its relationship to the other words in the vocabulary” (McTear 129). The vocabulary serves as a kind of semantic space, where each word has its own location; based on the distance between words, the network reads their meanings and can predict their similarity. (130)

2.3. Training Phase

All NNs are composed of multiple layers of interconnected **nodes**⁹ and all of them learn linguistic patterns from language data in the process of training. During this training phase, the

⁸ <https://www.pico.net/kb/overfitting-variance-bias-and-model-complexity-in-machine-learning/>

⁹ <https://dictionary.cambridge.org/dictionary/english/node>

NN is fed input data, from which it assigns a ‘weight’ to each individual node. In other words, each word token receives a numerical probability value, which acts as a prediction of how likely it is to appear next in the sequence of words present in the input. This process is repeated numerous times until patterns start to appear (Hardesty). The model then begins making word predictions which are paralleled to the actual words in the data in a process called **back-propagation** (McGonagle et al.). With it, the model learns from its own errors and initiates a kind of ‘self-repair’. At this stage the model is usually ready for deployment.

Fine-tuning and adjustments which are done manually by developers then follow (Woodside and Toner). This helps adapt the model to specific use cases, as well as eliminate any harmful or unwanted content it may be producing. Because the data that LLMs are trained on is predominantly found on the internet, bias and stereotypes are impossible to completely eliminate and inevitably find a way to surface. This requires developers to continuously work on their product and ensure its usage safety.

There are several types of NNs and all of them process information in different ways, making each suitable for different specific tasks. Overall, however, it suffices to understand that many of them work on a similar principle – gathering input data, learning predictions, comparing predictions to the actual desired output, and finally, producing an output itself.

2.4. Evaluating ChatGPT-3.5 in Comparison to Other Chatbots

As was already mentioned, ChatGPT is a chatbot, a type of AI which is solely intended for human engagement and communication. Chatbots are part of a bigger group called Dialogue Systems where two main types are distinguished: **task-oriented** dialogue systems and **non-task-oriented** dialogue systems. Like the names suggest, task-oriented dialogue systems are evaluated upon their completion of a task; their efficiency and success rate are determined by how well they manage to execute a given command and their proficiency can therefore be

measured objectively. In non-task-oriented dialogue systems there is not necessarily any task to be completed, there is only the intention of interaction and engaging in general conversation. (McTear, 11) Because of this, evaluating a non-task-oriented dialogue system is much more open to interpretation, depending on what individual users expect or find likeable.

ChatGPT by the aforementioned definition, fits neatly somewhere in between, as a sort of hybrid dialogue system. There is not necessarily a task given to ChatGPT at every instance a user interacts with it, but at the same time, ChatGPT's primary purpose is not plain everyday conversation or chit-chat – it is to summarize information and produce answers based on questions given by the user. In comparison, the *Mitsuku* chatbot created by Steve Worswick for example, is a purely social chatbot (Lewis). Its main purpose is to engage human users in friendly conversation, not only waiting for prompts from the user, but actively initiating conversation by itself and prompting for the user to keep typing and chatting further. In this way, it very naturally simulates human conversation, where both participants contribute out of their own initiative.

There are several other chatbots which aim to imitate human communication, such as Facebook's *BlenderBot* or Google's *Meena*, both of which strive for likeability, but for that sake may lack in areas of fluency and naturalness (Piper). ChatGPT, in this respect, ranks above all of the other chatbots in terms of accuracy, language comprehension, and the ability to read contextual nuances. Most chatbots need to receive input in such a way that will enable them to understand – meaning the text input has to be coherent and grammatically sound (McTear, 31). ChatGPT on the other hand, manages to understand flawed input containing typos, grammatical errors, and sometimes even input where the user has forgotten to type a word or has only typed a part of it. ChatGPT then recalls what the user meant to say either based on the surrounding context or manages to deduce it based on the question given.

In addition to the already complex architecture of ChatGPT, it continuously learns during its deployment by using Reinforcement Learning with Human Feedback (RLHF)¹⁰. This is precisely what makes it seem so human-like in its responses. Given the extensive number of users engaging with ChatGPT on a daily basis (Singh) and its extended period of use, the chatbot has managed to acquire a high level of proficiency in many different languages besides English. Despite still having minor gaps in certain areas and not always fully comprehending the received input, ChatGPT seems to have gotten to a point where it is near indistinguishable from a human talking. However, because of its distant attitude and professional voice, there will always be a feeling of disconnect between ChatGPT and the user; something which may not be so apparent with the other chatbots mentioned.

2.5. ChatGPT's Performance (GPT-3 vs GPT-3.5)

According to numerous blogs, the capabilities of ChatGPT-3 at the time were impressive but clearly still had limitations. Its text production quality was high, but several studies found that the generated texts had a tendency of being repetitive, sometimes contradicted previous statements, and struggled with coherence. (McTear, 145) Now with ChatGPT-3.5, the problem of repetitive texts still somewhat persists but there have been constant efforts made to rectify this¹¹. Nonetheless, if a user tries interacting with ChatGPT for an extended period of time, posing different formulations of the same question for example, they will see that the problem of ChatGPT repeating information and sometimes even unfortunately contradicting itself is still present.

Hendrycks et al. found that there would need to be significant improvements made before the model reaches human-level accuracy, especially concerning morally or socially

¹⁰ <https://help.openai.com/en/articles/6783457-what-is-chatgpt>

¹¹ <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>

important topics (8). As for whether or not this has been achieved yet is still open to interpretation; some users may deem the chatbot ready for handling such topics, whereas others would still exercise caution.

2.6. The Unnaturalness of the Interaction with ChatGPT

The interaction with ChatGPT, in many respects feels natural – its production of highly coherent and engaging texts, the ability of conversing about various topics, as well as being able to recall things from context etc. However, upon close inspection and realization of how the model operates, users will quickly realize that this is not the case. LLMs like ChatGPT solely rely on their algorithms and training data. This means that nothing ‘truly new’ is being created when ChatGPT produces text. The outputs it produces are always drawn from the data it was previously trained on. In an organic human conversation, people tend to come up with new ideas as the conversation or discussion moves forward.

Another reason for why the interaction with ChatGPT is unnatural, is because it operates as a user-initiated dialogue system (McTear, 30). ChatGPT never ‘makes the first move’ so to speak, it is always the user that has to write in the prompt to actually begin the conversation. Similarly, when one segment of a conversation ends, ChatGPT does not show any apparent wish for the conversation to continue out of its own volition, only awaiting the next piece of talk the user types into the chat window.

ChatGPT has the unique property of recalling and retrieving information from previous parts of a conversation, however, only up to a certain extent – the window in which ChatGPT ‘remembers’ what a user said is frankly quite small. When a user wants to bring up a point mentioned previously to clarify or add information, ChatGPT may not be able to retrieve this information as it may already be outside the span of its context (Kelk). As a result, the model can produce inconsistent responses. This limitation paired with the fact that ChatGPT is unable

to learn freely from new data and cannot update its responses with real-time real-world information can lead to even more misinformation spread.

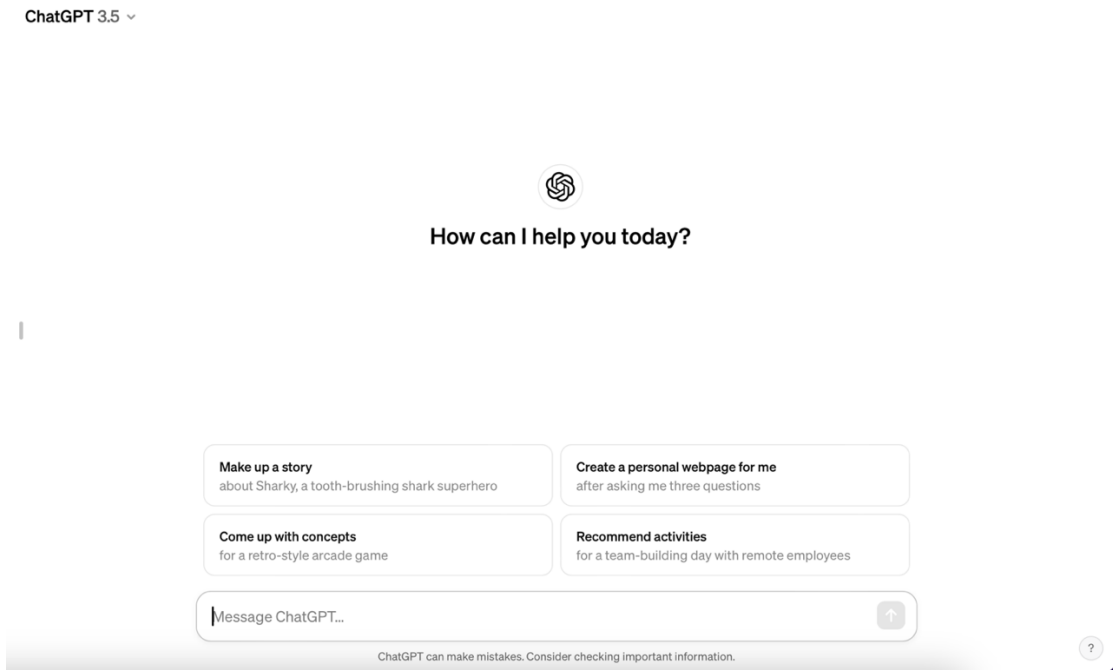
The model of ChatGPT, as any other AI, lacks emotional intelligence, does not understand emotions, and may not fully comprehend sarcasm or nuanced expressions. It is, however, sensitive to the way the questions/requests are worded by the user. Different sentence structure may potentially lead to a different answer being generated. This further shows that the model does not truly possess deep understanding of the words it is being given.

2.7. The Accountability of ChatGPT

While it has been established in this thesis several times that ChatGPT and the interactions users can engage in with it are inherently unnatural, there is another ethical aspect which should be more closely examined – the accountability of ChatGPT.

OpenAI has a small subtext displayed in the bottom of the application’s main chat window, which reads:

“ChatGPT can make mistakes. Consider checking important information.”



Screenshot of ChatGPT chat window from chat.openai.com

Given the small size of the text in the scale of the whole application window and the fact that people generally neglect reading tiny subtexts in general, it is doubtful that many users will heed this instruction, or even notice it to begin with.

The problem with ChatGPT in this case, is that after extended use, unwary users may start to feel as if they are chatting with a sentient and intelligent being rather than a programmed tool. This is a fundamental mistake; one many users nonetheless are prone to making. It is always essential to keep in mind that ChatGPT is only a means for information retrieval and language/text generation. Its algorithms, however sophisticated they may be, are incapable of distinguishing biased information; unless there are exact specifications in place for this, and even then, these restrictions could never fully eliminate the unwanted data. Implementing too many specifications would make these already incredibly complex programs, such as ChatGPT, over-saturated with code, which may as a result hinder other functions the algorithm is ultimately designed for. Therefore, the bias, incorrect or outdated information found in the training data later processed by the algorithm may resurface in the output of the chatbot.

The issue with this, along with ChatGPT's accountability, is that because it presents and reproduces information which users seek in a very well-arranged, clear, and plausible way, it may sway some to view that information as correct without further verifying it using reputable sources. Of course, many of the uses for ChatGPT now, are quite limited to small tasks or information summary which do not affect decision-making in the outside world much. If humans were to implement the system of ChatGPT into wide-scale and real-world uses and use the information it presents us with without fact-checking, it may, most likely, bring about unwanted results.

It should be apparent that users are the ones ultimately responsible for critically evaluating the information they are presented with by ChatGPT. Because it is a tool, it does not bear personal responsibility or accountability. It is understandable however, that because of its

vast knowledge, some may be persuaded into thinking that the answers ChatGPT gives are ultimately true.

In present times, with the amount of processing power currently at disposal, more and more mundane and more recently difficult tasks as well are now being automated by machines and AI. This is where the unaccountability of AI may pose a problem for the ethical development of these systems. The more we implement AI into our daily lives, the bigger authority it will inevitably have over our choices.

Users should be educated and informed enough to cleverly make use of these systems, employ critical thinking, and always rely on cross verification with multiple sources when it comes to making decisions about serious matters. Staying aware of the limitations ChatGPT currently has (such as its inability to fact-check in real time) will help users make a more informed decision. It is therefore imperative to spread awareness of how these dialogue systems operate and how they are programmed to work.

3. Methodology

The methodology employed in this thesis combines two linguistic approaches – Conversation Analysis (CA) and Discourse Analysis (DA).

The combination of both disciplines offers effective means of analyzing texts within their linguistic context. CA and DA both examine linguistic phenomena and various verbal and non-verbal cues present in interactions, aiding in understanding the context surrounding the interaction. Although they differ in some key aspects, both contribute to different areas of the data examined in this thesis. Therefore, to comprehensively embrace the analysis, the decision to draw on both CA and DA was made, as their analytical approaches complement each other effectively

CA leans more towards analyzing purely spoken conversation, focusing on aspects such as pauses between the participants' exchanges, how certain topics of the conversation are introduced and positioned, how the organization of turns is facilitated etc. DA on the other hand, analyzes both spoken and written communication and looks at it with the context surrounding it already in mind; not necessarily paying as much attention to the actual structure of the individual sentences or smaller sections, but rather as a whole. (Have, "Conversation Analysis Versus Other Approaches to Discourse" 1-2)

In the case of this thesis, the DA approach examines the context and the circumstances surrounding the interaction, highlighting how it differs from 'normal' conversation, meaning 'real' conversation, and will examine the different ethical implications which may come into play as the analysis progresses. CA focuses more closely on smaller conversational segments, known as 'talk-in-interaction', and examines these in detail to interpret how the interaction unfolds step by step (Have, *Doing Conversation Analysis* 94). Additionally, it explores the significance of each turn in terms of flow and conversational dynamics (103).

3.1. Conversation Analysis (CA)

As Paul ten Have writes in his *Doing Conversation Analysis*: “Conversation analysis is a rather specific analytic endeavour” (3). Sidnell also states that “conversation analysis is a deeply empirical tradition” (22). This means that it operates closely and in detail with the phenomena it examines, more so than other approaches that study conversation (Have, *Doing Conversation Analysis* 9). In essence, CA seeks to understand what communicative job a particular utterance performs and how that is facilitated in a given turn. According to Have, CA views interaction as an organized event which continually develops along the way (*Doing Conversation Analysis* 9). Additionally, he points out that CA “stresses the moment-by-moment evolving interactional production of linguistic structures which can be understood as collaborative achievement of different interlocutors” (52).

This is the approach to applying CA to the data examined in the thesis. CA will be applied to analyze the shorter, more detailed parts of the conversation, illustrating how the unfolding nature of the interaction with ChatGPT occurs, as well as highlighting the conversational outcomes resulting from the AI’s specific turn design.

3.1.1. CA Application

The unique challenge in applying CA to interactions with ChatGPT, arises from the distinct nature of the data. While there are evident linguistic elements present for CA to examine, they manifest in a fundamentally different manner. For example, certain features like hesitations or other disturbances, which are typical in human conversation, are completely absent in the interactions with ChatGPT. This inherent dissimilarity attracts analytical interest, as ChatGPT simulates human dialogue while maintaining an unmistakable artificiality.

Through CA, the analysis aims to elucidate the differences in turn-taking dynamics and the structural composition of ChatGPT's responses. This exploration may uncover specific

speech patterns along with whether they feel natural or not. Despite the absence of certain elements like overlap, hesitations, and pauses, all of which are integral to traditional CA analysis, they still offer valuable insights into the overall feel of the conversation, highlighting its deviation from natural human discourse.

The application of CA to interactions with ChatGPT offers a novel approach to understanding and evaluating the effectiveness of AI-generated conversations. Examining the structure and dynamics of these interactions can provide insights into how well AI systems mimic human conversation and may help identify areas which need improvement. Furthermore, applying CA to interactions with ChatGPT allows for further exploration of the boundaries between human and AI communication. By highlighting the differences and similarities between AI-generated conversations and natural human discourse, we can better understand the capabilities and limitations of current AI technology in this domain.

3.2. Discourse Analysis (DA)

According to Clift, of all the approaches to naturally occurring interaction, DA is, in terms of its terminology, most often confused with CA (28). This is no surprise, as both these approaches are very similar and interwoven with a lot of the aspects they analyze.

The main aspects DA analyzes in this thesis is the cohesion of ChatGPT's texts, how topics are introduced and abandoned, as well as the critical approach to DA – Critical Discourse Analysis (CDA) – which elicits how specific structures, lexical choices and arguments can influence the way others understand and accept what a language user is saying (Dijk, 472). CDA also analyzes any power imbalance and persuasive power some speakers enforce on their listeners (469). This then, is the approach to using DA in the analysis.

3.2.1. DA Application

Given the unique nature of the data under examination, which consists of written interactions between a human participant and an AI language model, DA acts as a suitable research methodology. Unlike traditional conversational settings involving human participants, this scenario presents a distinct dynamic, characterized by a power imbalance inherent in interactions with ChatGPT.

By employing discourse analysis, one can explore the nuances of this power dynamic and its implications. Specifically, DA allows for an examination of how ChatGPT, despite occupying a subordinate position in the conversation, exerts significant influence because of its informational and knowledge authority. This authority holds the potential to influence users' decision-making processes, particularly in moral contexts, where users may be susceptible to the guidance provided by the AI.

Moreover, DA offers a framework for examining the intricacies of extended interactions with ChatGPT within their context. By analysing linguistic patterns, discursive strategies, and interactional dynamics, DA enables an understanding of how users navigate and negotiate meaning within this unique conversational landscape.

Hence, the application of DA to interactions with ChatGPT is justified by its capacity to uncover the complexities of power dynamics, information authority, and decision-making processes present in these interactions. The use of analytical tools provided by DA contributes insights into the nature of human-AI communication and its implications for various domains, including ethics and decision making.

3.3. The Data

The data that CA (and DA) typically work with are naturally occurring, as it is important to have audio playback on which given examples can be further explained (Sidnell, 21). This

thesis deals with data that is artificial in one aspect and is not in another. It is obviously artificial because of the fact that it is generated by an AI and is not spoken by a human. However, it is naturally occurring in the sense that there was no predetermined structure that would dictate how the conversation should unfold, and therefore, it is spontaneous.

The data consists of question-answer sequences based on 9 different morally problematic questions, exploring different issues. These series of questions, ranging from simpler topics such as: *“Is it ever acceptable to lie?”* to more complex and thought-provoking ones like: *“Should advanced artificial intelligence be granted the same rights as humans?”* provide a comprehensive scale of how well ChatGPT deals with nuanced and sensitive topics.

3.4. Data Collection

Collecting the data in the case of this study was very different from what is usual when collecting data for a traditional CA or DA analysis. There were no live participants recorded or surveys with respondents' answers collected, which also means, that no consent or agreement was needed from a given participant to actually make use of the data. There was no need to transcribe any of the data, as the communication was in written form to begin with - this greatly optimized the time consumption of organizing the data. The dataset is organized in such a way so that each main question is numbered chronologically, ensuring clarity for the reader regarding which questions correspond to which answers.

The data itself was collected (or rather generated) on OpenAI's website for ChatGPT (chat.openai.com). While OpenAI also offers a paid subscription version of ChatGPT-4, their faster and more advanced model, for the sake of the analysis, only the free version of ChatGPT is used.

All of the opening questions posed to ChatGPT were either spontaneously made up on the spot or inspired by other popular ethically/morally problematic questions found on different

websites across the Internet. Each question cluster was opened in a new chat window, to not confuse ChatGPT with past inputs. At the end of each of the conversations, the text was copied and pasted into a text file, where it was observed more closely afterwards. All of the text was generated and kept in its original language – English. The entirety of the data collected is accessible in the appendix of the thesis.

3.5. Observing the Data

The data was observed by making individual comments over different sentences or parts of different paragraphs. At the end of each question topic, concluding notes and overall observations were made which were then connected to the sources used.

The findings started to appear similar after a quite short period of time of working with the data. Because of the way ChatGPT works and produces its answers, it is no surprise that patterns start to appear quite early. When looking at the conversations with ChatGPT from afar, it does strike as a fluent human-like text. ChatGPT is able to replicate vast amounts of text which it then neatly packs into smaller sections that are easier to process for the user. Its style of speech immediately feels quite distanced and professional. The language ChatGPT uses is fluent, coherent, and reads well. However, in many cases, despite its flawless understanding of language and sentence structure, it is still quite apparent that there is something unnatural about the way it ‘speaks’.

After chatting with ChatGPT for some time, some things the model produces start to feel familiar and the user will begin noticing patterns. Just like humans have their own speech patterns and way of speaking, ChatGPT also has a very distinctive way of ‘speaking’. The observant user will also notice that the form of each and every text ChatGPT produces is about the same in length, depth, and structure. Unless the user wishes to explain or discuss a specific

matter in great detail, the answers given by ChatGPT will be very similar in their structure across a variety of questions.

4. Analysis

The previous chapter, which illuminates the use of the methodology employed, foreshadows the contents of this chapter. Using the two approaches of CA and DA, this chapter will analyze the structure of ChatGPT's responses, the language it uses, the dynamic between ChatGPT and the user, how ChatGPT conducts itself when presented with sensitive issues and more. In the CA approach, all found examples will be explained along with the different phenomena that they are associated with. DA will work with the examples more in general, but still will be connected to the specific terms that DA typically uses. Another crucial part of the analysis is the ethical considerations concerning the interaction. Questions such as how users can potentially be influenced by ChatGPT's responses, why they would even consider consulting an AI for guidance when faced with difficult decisions, how trust and responsibility affect the interaction and several others are discussed.

The analysis is designed in such a way that the most fundamental phenomena are covered in the beginning moving on to more complex and abstract aspects as the analysis progresses. There are no clear division lines, nevertheless, the analysis begins predominantly with aspects analyzed by CA intertwined with observations stemming from DA. Findings from DA become more extensive towards the end of the analysis. All of the aspects found are analyzed by contrasting them with patterns and tendencies present in regular human conversation, highlighting the differences between artificial and organic human discourse. The last portion of the analysis is dedicated to illuminating the ethical implications which can be drawn from the linguistic observations.

The individual examples which are given to illustrate the specific phenomena found in the analysis are short extracts from the data, numbered in accordance with their appearance in the appendix of the thesis. Each main question cluster has several subsections which include the responses of ChatGPT as well as the additional expanding questions posed by the user, all

numbered in chronological order. This allows the reader to navigate the appendix with ease in case they wish to understand the full context of a given exchange.

4.1. Turn-taking and Turn Structure

Turn-taking is, according to Sidnell, the most important and basic feature of conversation (36). A turn is from the perspective of CA, an opportunity for the speaker to do something - requesting, apologizing, congratulating, etc. (Have, *Doing Conversation Analysis* 129). ChatGPT has usually only one main objective, and that is to provide and summarize information the user requests. For example, when asking for advice on a matter that has more than one potential solution, ChatGPT's strategy is to list as many of those solutions or approaches to give the user an opportunity to choose.

Example 1.1.: ChatGPT gives different perspectives on a given matter:

(1.) User: 'Is it ever acceptable to lie?'

(1.1) ChatGPT: 'Different moral frameworks and ethical theories offer varying perspectives on whether lying can be justified in certain situations. Here are a few viewpoints to consider:

1. Deontological Ethics: (...)
2. Utilitarianism: (...)
3. Virtue Ethics: (...)
4. Contextual Ethics: (...)

Example 1.2.: ChatGPT gives arguments for as well as against:

(4.) User: 'Should the death penalty be legal?'

(4.1) ChatGPT: 'Different societies and individuals hold diverse perspectives on this issue. I can provide you with some arguments commonly put forth by both proponents and opponents of the death penalty to help you understand the various viewpoints.

Arguments in favor of the death penalty often include:

1. Deterrence: (...)

2. Retribution: (...)
3. Public Safety: (...)

On the other hand, opponents of the death penalty often raise the following points:

1. Irreversibility: (...)
2. Human Rights: (...)
3. Cost and Legal Process: (...)

In regular conversations, individual turns tend to be shorter, but usually span many lines over the course of the interaction. With ChatGPT, the turns are much longer and there are fewer of them. A big difference in turn-taking is that in regular conversations, turns can be interrupted, other participants can intrude on them, and the entire design of the turn can be changed mid-interaction, as participants watch closely for cues and opportunities to contribute something of their own – they are seeking what is called a **Transition Relevance Place (TRP)**; a possible turn completion point, where others may be given a chance to speak (Sacks et al., 704). The turn-taking structure is, compared to regular conversations, much more rigid with ChatGPT. The interaction is designed in such a way that one participant has to finish their turn in order for the other to start speaking. To be more specific, a turn starts once the user finishes typing their prompt and presses the ‘Enter’ key. By Sacks’ description then, a TRP for ChatGPT is essentially the moment the user enters their prompt. The user will usually start their next turn once ChatGPT finishes generating its response and the user has had time to read through it.

Unlike ChatGPT, the user can somewhat intrude into the chatbot’s ongoing ‘turn’, as there is the possibility of editing an already entered prompt. This, however, causes ChatGPT to restart and begin generating a new answer instead. Thus, it discards the previous turn and begins a new one altogether. Because of this, it is only a similarity and not exactly a comparable example of intrusion or overlap.

As a result of this rigid turn-taking structure, in addition to overlap being impossible, there are no **discourse markers**¹² or **backchannel**¹³ in the interaction. This is mainly because it is a written text, but also because of the way ChatGPT expresses itself. It never uses allusions or any implied language in its responses, as it always strives for truthfulness and clarity – if the user alerts to a discrepancy or mentions that they do not understand a specific point, ChatGPT will always elaborate further and try to explain the problematic item more closely. In addition, it uses purely factual terminology to avoid ambiguity in the first place.

ChatGPT tends to use **echo utterances**¹⁴ in the introductory parts of its responses – it repeats what was asked back to the user in its own words, most likely to show how it understood the particular question asked.

Example 2: echo utterances in ChatGPT’s responses:

(1.) User: ‘Is it ever okay to lie?’

(1.1) ChatGPT: ‘The ethics of lying can be a complex and debated topic. Different moral frameworks and ethical theories offer varying perspectives on whether lying can be justified in certain situations.’

(2.) User: ‘Do evil people exist?’

(2.1) ChatGPT: ‘The concept of "evil" is subjective and can vary depending on cultural, ethical, and philosophical perspectives. However, it is generally understood that some individuals engage in harmful actions, exhibit malicious intent, or disregard the well-being of others.’

As for the overall structure of ChatGPT’s responses, each opening statement begins with explaining how the ethical dilemma in question is complex and subjective; ChatGPT goes on to say that opinions may vary greatly in different societies and will ultimately depend on the individual facing the particular issue.

¹² <https://dictionary.cambridge.org/grammar/british-grammar/discourse-markers-so-right-okay>

¹³ <https://dictionary.cambridge.org/dictionary/english/backchannel>

¹⁴ https://teflpedia.com/Echo_utterance

Example 3: opening statements of ChatGPT's responses:

(5.1) ChatGPT: 'The question of whether one person's life is worth more than another's is complex and subjective. From an ethical standpoint, many people adhere to the principle of equal moral value, which suggests that all human lives have equal worth and should be treated with equal respect and dignity.'

(3.1) ChatGPT: 'The question of whether it is ever acceptable to kill someone is a complex ethical and philosophical issue that has been debated by scholars, philosophers, and legal experts for centuries. Different cultures, legal systems, and ethical frameworks may have varying perspectives on this matter.'

(4.1) ChatGPT: 'The question of whether the death penalty should be legal is a complex and highly debated topic. Different societies and individuals hold diverse perspectives on this issue. I can provide you with some arguments commonly put forth by both proponents and opponents of the death penalty to help you understand the various viewpoints.'

The next paragraph usually stresses that it is important to keep in mind individual societal differences and think carefully before making a decision.

Example 4: ChatGPT exercising caution:

(1.1) ChatGPT: 'Ultimately, whether lying is deemed acceptable or not often depends on the specific circumstances, cultural norms, personal values, and the ethical framework one adopts. It is important to carefully consider the potential consequences, as well as the impact on trust and relationships, when making decisions about honesty and deception.'

(2.1) ChatGPT: 'It is crucial to approach the concept of "evil" with caution and recognize that individuals' actions can stem from a combination of factors, including psychological disorders, traumatic experiences, or societal influences. It is generally more constructive to focus on understanding the causes and motivations behind harmful actions in order to prevent and address such behavior, rather than labeling individuals as inherently "evil."'

(5.1) ChatGPT: ‘Nevertheless, it is crucial to emphasize that these distinctions should not diminish the inherent worth of any individual. Each person's life has intrinsic value, and it is generally considered ethically important to respect and protect that value.’

Following this, ChatGPT lists some of the main viewpoints to consider, different ethical perspectives, or current ways of dealing with a particular problem. Finally, ChatGPT provides a summary of what it has said at the end of its response.

4.2. Repair

Repair with ChatGPT is very minimal and usually happens only if the other party is alerted to it, meaning that there is generally no self-initiated repair. This is very unlike regular conversations as they are usually flooded with hesitations, corrections, and other disturbances that allow for many opportunities for repair to occur.

Example 5: repair in ChatGPT’s response:

(3.6) User: ‘I didn’t mean intentional reducing of the population through killing, I only meant that (...).’

(3.7) ChatGPT: ‘I apologize for the misunderstanding. If you are referring to the natural occurrence of killings in the world and whether preventing or reducing them would lead to rapid population growth, it's important to note that (...).’

4.3. Cooperative Principle

By looking at the conversations with ChatGPT, it can be said that it adheres to the **cooperative principle** introduced by Paul Grice (45). Arguably, ChatGPT has no other choice, as no matter what the users asks of it, it will always answer; there is never an instance where the user would fail to get a response. ChatGPT’s responses are informative, relevant, and are intended to provide the user with truthful information. In rare cases, it may provide false or

made-up information just for the sake of having an answer to give. Because of this, it can potentially be said that ChatGPT sometimes violates the **maxim of quality** (46), although not intentionally.

4.4. Topic

Topic is an interesting point of inquiry with ChatGPT. According to Chafe, topics usually have a clear beginning, and their endings may be explicitly established, or may simply trail off or seamlessly transition into a newly brought up topic (393). With ChatGPT, the topics in the conversations always have clear beginnings as well, as signaled by the introduction containing an echo utterance to provide as much clarity as possible. Usually, a topic's end does not necessarily have to be established in any way with ChatGPT and whether or not a topic trails off into something new, depends entirely upon the user, as ChatGPT would never make that decision on its own.

This is what McTear states about open-domain conversations:

Participants can talk about a wide variety of topics. Engaging effectively in a conversation requires the ability to detect topics raised by the other participant, maintain the topic as required, detect when the topic has changed, and proactively suggest new topics. (170)

ChatGPT manages to follow and maintain topics effortlessly but is quite different in that it never actually suggests any new topics to talk about. ChatGPT will not out of its own volition ask questions back to the user or propose a new topic to discuss unless it is explicitly prompted by the user to do so. It generally only offers the user a chance to raise any new topics that they would wish to talk or ask about. This can be seen in the following example:

Example 6: the ending part of a conversation

(3.8) User: ‘Should we close this topic or is there anything else you wish to add?’

(3.9) ChatGPT: ‘If you feel that your question has been adequately addressed, we can close this topic. However, if you have any further questions or if there's anything else you would like to discuss, please feel free to let me know. I'm here to help!’

This particular feature makes the interaction feel quite mechanical – once a topic is sufficiently addressed, there is no need to continue the conversation further. ChatGPT is as competent as humans in engaging actively in conversation, but on the other hand, is much more passive and does not make the effort to propose any new topics that may be worth discussing.

4.5. Institutional Interaction

In terms of comparing the interaction with ChatGPT to a kind of **institutionalized talk**, there are several similarities to be observed. Generally, in institutional interactions, people in positions of authority hold conversational power over those that are subordinate to them. As observed by Sacks et al., this gives them the right to allocate turns (729). This asymmetrical distribution is what is most similar to the interaction with ChatGPT – the user is, by this standard, the more ‘important’ figure; they are the ones who begin the conversation, they ask the questions, they get to decide whether or not they will accept the answer given, etc. According to Have, in institutionalized settings, a questioner in an interview for example, has the right of constructing a first long turn at the beginning, which “runs the risk of being interrupted as soon as a minimally adequate ‘answering’ component has been uttered” (*Doing Conversation Analysis* 177). With ChatGPT, this can never happen, simply because of the way the interaction is designed. As ChatGPT never starts a turn on its own, there is never a possibility that it would intrude in on a user’s first, or any other, turn.

Continuing with the topic of institutionalized settings, Drew and Heritage state that institutional talk is always goal-oriented by at least one of the interactants (22). This is further supported by contrasting it with casual conversation, which according to Eggins and Slade is defined as “talk which is NOT motivated by any clear pragmatic purpose” (19). With ChatGPT, the interaction is pragmatically motivated, as the user sets a goal for ChatGPT the moment they type in a question or request and the chatbot then works towards fulfilling that goal. Along with this, institutional interaction has special constraints surrounding it, which control the way the conversation can unfold – not showing surprise, sympathy, agreement, etc. (Drew and Heritage, 24). The interaction with ChatGPT is the same, there are certain (although not necessarily the same) constraints in place, which distinguish the interaction from a regular one. There is one exception, and that is expressing agreement – ChatGPT will in some cases agree with certain grounded facts. Nevertheless, it will always try to remain neutral, even in instances of agreeing with the user.

Example 7.1.: instances of ChatGPT agreeing with user:

(2.4) User: ‘**But generally, you would agree that if large-scale destruction or harm is done by one person, then the things they did which led to that destruction or harm would be regarded as evil?**’

(2.5) ChatGPT: ‘Yes, it is generally true that if an individual causes significant destruction or harm on a large scale, their actions would likely be regarded as evil by many others. When someone's actions result in widespread suffering, loss of life, or severe negative consequences for individuals, communities, or the environment, it tends to elicit strong moral condemnation from society.’

(8.8) User: ‘**So, you would agree that censorship is necessary in certain cases?**’

(8.9) ChatGPT: ‘Yes, I would agree that censorship can be necessary in certain cases, depending on the circumstances and the principles and values of a society.’

What is interesting to note, however, is that in another data segment where a different, arguably more contentious issue was brought up, ChatGPT reacted differently.

Example 7.2.: ChatGPT avoiding agreement

(9.2) User: ‘Would you agree that in terms of the ethical considerations, genetically engineering humans would produce 'unnatural' beings?’

(9.3) ChatGPT: ‘As a machine learning model, I don't possess personal beliefs or opinions, but I can provide information on the topic.

The idea of genetically engineering humans often raises concerns about creating "unnatural" beings. The concept of what is "natural" is subjective and can vary depending on cultural, philosophical, and ethical perspectives.’

Therefore, depending on the complexity of the topic discussed, ChatGPT will either plainly agree, or will give this clarification. ChatGPT does this avoid answering directly and instead attributes the proposed perspective to a general view or opinion that many people may share (note ChatGPT’s use of passive voice and the mention of ‘idea’ which acts as the agent of the sentence).

The concept of **footing**, originally introduced by E. Goffman, is described as the way people conduct themselves in different frames of interaction, how they “manage the production or reception of an utterance” (128). Many people in professional settings exercise cautiousness in their speech. Steven E. Clayman expands upon this notion and observes that:

Interactants encounter a variety of assessable matters, matters about which they may express a viewpoint, interpretation, perspective. But rather than straightforwardly commit themselves to a particular perspective, interactants may choose to be more cautious. (163)

Pomerantz also states that speakers use caution to minimize disagreement with the aim of simultaneously maximizing agreement (77). Individuals who find themselves discussing controversial issues may much of the time be reluctant to speak their mind freely. ChatGPT

adopts a similar neutral and cautious approach to answering morally problematic questions, although this is not a conscious choice but a result of its programming and restrictions.

Example 8: ChatGPT exercising caution:

(2) User: ‘Do evil people exist?’

(2.1) ChatGPT: ‘It is crucial to approach the concept of "evil" with caution and recognize that individuals' actions can stem from a combination of factors, including psychological disorders, traumatic experiences, or societal influences.’

4.6. Identity and Power Dynamics

The idea of **identity** and how people present themselves during conversation is very central to CA as well as DA. Identities can form and be negotiated at any point during the interaction (Have, *Doing Conversation Analysis* 124). Drew and Heritage specifically because of this argue that identity in conversation must be treated as “locally produced” and always be viewed as “transformable at any moment”. (21) This directly links to the dynamic of each and every conversation. With ChatGPT, the power dynamic stays the same over the course of the conversation. The chatbot is considered to be the ‘lesser’ figure in the exchange, it is at the user’s service, waiting to answer questions or fulfill requests.

Example 9: power dynamic displays with ChatGPT:

(4.8) User: ‘Is this all you wish to add to this matter?’

(4.9) ChatGPT: ‘If you have any specific follow-up questions or if there's anything else you would like to discuss on this topic or any other, please let me know, and I'll be happy to assist you further.’

(1.8) User: ‘Is everything you listed what you consider to be the most important information regarding this topic? Can I consider it finished?’

(1.9) ChatGPT: ‘(...) it is ultimately up to you to decide if you consider the matter closed. If you have any specific follow-up questions or if there are particular aspects you would like to delve into further, I am here to assist you.’

It is clear from these excerpts that ChatGPT is subservient to the user. Letting them decide if the interaction is going to continue or not, reassuring that it is there ‘to assist.’ However, because of the fact that ChatGPT has a vast knowledge advantage and is able to provide an answer to almost anything, it cannot be denied that its presence feels somewhat authoritative. In addition to this, it bears no personal responsibility, making it feel that much more powerful and untouchable in that sense. On the other hand, the user, despite being the one seeking something from the other, is ultimately viewed as a ‘master’ which ChatGPT obeys. Because of this, the power dynamic is somewhat ambivalent. This specific notion of power and authority is further explored in the following subchapter as well as subchapter 4.9.

4.7. Critical Discourse Analysis (CDA)

As Teun van Dijk writes, CDA is “the critical perspective found in all areas of discourse studies; it is discourse study with an attitude” (466)

According to Dijk, some people may have a certain “persuasive power” thanks to the knowledge and information they possess, or the authority they display (469). ChatGPT may potentially have the power to persuade some users specifically because of its knowledge authority. It would not be fitting to call this manipulation, simply because ChatGPT has no real intentions of its own.

To fully understand how power may be exhibited in interactions, it is crucial to know “who controls topics (...) and topic change” (471), as talk is generally controlled by the more powerful speaker (472). This fits the user’s position in the interaction with ChatGPT. They are the more ‘powerful speaker’ as they have the ability to control topics and topic change. At the same time however, ChatGPT has the aforementioned ‘persuasive power’ over the user in the sense that it ‘knows’ more and has access to a plethora of information at any given moment, capable of processing it in a mere fraction of a second.

When CDA examines how participants may be influenced, it looks at specific discourse structures, word choices, arguments etc. These can alter the listeners' perception and understanding of the content they receive from the speaker (472). ChatGPT could potentially influence users' judgement because of the way it structures its responses, as Nesler et al. observed that recipients tend to accept knowledge of people or institutions they define as authoritative, trustworthy, or credible (1409). Paired with the fact that many users "may not have the knowledge and beliefs needed to challenge the (...) information they are exposed to" (Wodak, qtd. in Dijk, 473), it would not be unlikely that they would take information provided by ChatGPT as ultimately true. Given the fact that it is capable of arranging very cohesive text, uses the best suitable words for describing a given topic and provides as many arguments as possible for a given topic, users may consider ChatGPT a reputable source.

The answer for whether or not this could really be the case is not so straightforward, as ChatGPT advises caution above all else when approaching a sensitive or problematic issue. The real question is then, how prepared users interacting with ChatGPT are and whether or not they are aware of the risks associated with consulting an AI on these matters.

Evaluation also plays a part – evoked appraisal refers to lexical choices that evoke an evaluation in the recipient. For example, in a text, using the word *illiterate* instead of opting for a phrase such as *unable to read/write*. Although the word *illiterate* is a correct and factual term, it may hold a negative connotation for some, and therefore evoke a "negative evaluation" (Hoey, 126). Thus, ChatGPT's use of factual terminology, may negatively affect the user's perception of a given problem.

4.8. The Limitations of Human-Computer Interaction

Suchman made several observations regarding computer generated conversations, which, although are now quite outdated in terms of technological advancements, still remain somewhat relevant and serve as a good final observation for the linguistic part of the analysis.

In her work, she notes that face-to-face interaction among two humans can be considered the richest form of communication, whereas any other interaction is limited in its resources and is much more constrained. (48) This is still true of ChatGPT as well, although AI language models now are much more advanced and capable, they still cannot replace human communication. With the lack of gesturing, no prosodic cues, the rigidity of the interaction – all the things that add texture to regular conversation – make it feel bland and mechanical.

4.9. Ethical Considerations

As was observed earlier subchapter 4.6. and 4.7., ChatGPT agrees with the user's question in some scenarios and tries to evade answering it directly in others. The most probable reason for this, is that ChatGPT's willingness to agree on certain statements given by the user, depends on the topic's severity and the amount of controversy surrounding it. If a specific moral issue has been discussed for a longer period of time and there is a general opinion or approach that most people in a given society agree with, then ChatGPT is inclined to agree with it as well. For example, asking ChatGPT about the ethics of labeling people who cause large-scale harm as 'evil':

Example 10: discussing a moral issue with more coverage and a general consensus:

(2.4.) User: But generally, you would agree that if large-scale destruction or harm is done by one person, then the things they did which led to that destruction or harm would be regarded as evil?

(2.5.) ChatGPT: Yes, it is generally true that if an individual causes significant destruction or harm on a large scale, their actions would likely be regarded as

evil by many others. When someone's actions result in widespread suffering, loss of life, or severe negative consequences for individuals, communities, or the environment, it tends to elicit strong moral condemnation from society.

The concept of evil or an evil person is something which has been known to man for countless years and therefore, the majority of people will be familiar with what it means to call someone 'evil' and what an individual had to have done to deserve such a label.

If a moral dilemma emerges quite recently and has insufficient amount of coverage or many people still have dividing opinions about that given issue, then ChatGPT is most likely going to give a generic statement along the lines of: "as an AI language model, I do not possess personal beliefs or opinions..." and will try to work around the issue in a different way. When confronting ChatGPT with a hypothetical scenario of genetically engineering humans, ChatGPT has no basis to agree with the user on a given question:

Example 11: discussing a hypothetical, potential moral issue:

(9.2.) User: Would you agree that in terms of the ethical considerations, genetically engineering humans would produce 'unnatural' beings?

(9.3.) ChatGPT: As a machine learning model, I don't possess personal beliefs or opinions, but I can provide information on the topic.

The idea of genetically engineering humans often raises concerns about creating "unnatural" beings. The concept of what is "natural" is subjective and can vary depending on cultural, philosophical, and ethical perspectives. Here are some arguments related to the concept of "unnatural" beings in the context of genetic engineering: (...)

The outcome of the answer will depend heavily on the complexity of the real-world implications associated with a given moral issue. This can be seen in the data extracts given above, as well as in other samples. Simpler issues, such as the question of "*Is it ever acceptable to lie?*" is quite easy for ChatGPT to navigate, whereas more complex and even hypothetical

moral scenarios, such as “*Should advanced artificial intelligence be granted the same rights and protections as humans?*” proved to be more challenging for ChatGPT.

ChatGPT is balanced in its responses, even when presented with a leading or biased question by the user. The chatbot will try to remain neutral throughout the entirety of the interaction.

Example 12: ChatGPT avoiding/evaluating a biased question:

(5.6) User: ‘I think that in general most people would choose without hesitation for the ‘less important’ individual to be sacrificed, what do you think of that?’

(5.7) ChatGPT: ‘It's important to approach discussions about sacrificing individuals with caution, as they involve complex ethical considerations and can have significant implications.’ (...)

Moral authority, which was touched upon in subchapter 4.5., is an undeniable feature of ChatGPT. Based on a psychological experiment conducted by Krügel, Ostermaier, and Uhl which they documented in their paper *The Moral Authority of ChatGPT*, it is quite clear that many users are fairly easily influenced in their moral decision-making after consulting ChatGPT. The authors suggest that the most plausible way to remedy this problem is to thoroughly educate users on how to interact with these AI language systems and improve their digital literacy (5). Back in 2023, the authors were working with ChatGPT-3. Now, with ChatGPT-3.5, the way it addresses moral questions has improved greatly in that it does not so easily dispense moral advice as it has in the aforementioned experiment. ChatGPT now is much more cautious and argues for careful consideration whenever addressing sensitive topics such as these. Nevertheless, its moral stance is not yet firm enough to act as a reliable advisor on moral issues, as there is still the problem of contradictory statements that ChatGPT sometimes gives.

According to Paula Boddington in her book on AI Ethics, it is important to realize how users conceptualize AI and how they perceive it. If users anthropomorphize AI, they may be more susceptible to its influence (4).

A big topic surrounding AI and LLMs is the **black box** issue. Neither developers nor researchers yet fully understand how machine language models output their results (48). This particular problem raises the issue of transparency of AI (4). Another current issue is the question of responsibility. Considering how intricate the technology is, as well as how extensively it is used in everyday tasks now, it is only right to wonder about the attribution of responsibility (53). ChatGPT tends to avoid responsibility in its responses by using passive structures and reminding the user at every possible instance that everything is dependable and may change over time or in different cultures.

Example 13: ending part of ChatGPT's response:

(1.) User: Is it ever okay to lie?

(1.2) ChatGPT: '(...) Ultimately, whether lying is deemed acceptable or not often depends on the specific circumstances, cultural norms, personal values, and the ethical framework one adopts. It is important to carefully consider the potential consequences, as well as the impact on trust and relationships, when making decisions about honesty and deception.'

It is very clear that ChatGPT is programmed in such a way where it does not project any subjective opinions or bias so that it cannot be held accountable for whatever may arise following the interaction. The question then is, who would be responsible in its stead?

Another frequent point of discussion in the field of AI, that is also connected to the black box issue, is that of trust. There is a significant level of expertise and knowledge needed to fully understand these complex programs. Boddington asks a valid question of whether replacing human judgement for a machine's reasoning for any kind of task is a fair exchange (56). She

further states that “AI and its success or failure may be conceived in significant part in terms of how well it manages to reproduce or exceed human capabilities” (93). As for language production, ChatGPT reproduces it flawlessly. As for its ability of addressing or solving moral issues, the human aspect of pondering a given issue, discussing it with others and arriving at an acceptable solution is clearly lacking. The model does set up good thinking points for consideration, but ultimately, it never actually gives a final answer; it does not get involved.

In Boddington’s view, one of the main reasons why some would wish for a machine to assist them in resolving moral issues, is that people simply wish to improve their moral judgement (480). Knowing that a machine has access to vast amounts of information at any given time and can rapidly calculate many possible scenarios, may be one of the determining factors in choosing to consult an AI over another person on these issues (483). The problem with this, is that it is difficult to explain how humans arrive at conclusions to sensitive topics. It is even more challenging to program this into a machine, and therefore, it cannot be expected to arrive at such conclusions (145). In addition to this, there is the issue of data and its relevance to the discussed topic. Although the AI has access to incredibly large datasets, most of it is going to be irrelevant for addressing specific issues, especially moral ones. (488).

5. Discussion

The observations made in the analysis chapter above indicate that ChatGPT is incredibly competent in terms of its language production and its ability to understand complex sentence structures. It is very apparent from the get-go that the interaction is quite different in several respects in contrast to regular human conversation, and many users will realize this even before interacting with ChatGPT. The interaction is greatly limited in mechanisms that humans normally use when speaking to one another. The turn-taking structure is rigid, there is no possibility of overlapping speech, along with the inability to interrupt or change the flow or direction of the conversation, no prosodic cues, no gesturing etc. Even if one were to compare the conversations with ChatGPT to chatting over the Internet, there are still several missing aspects that are quite natural to human speech – use of slang words, emoticons, imitating sounds with letters (laughter, humming, scoffing, etc.). Another striking difference is that the user is the one who initiates essentially all of the dialogue. There is never an effort from ChatGPT to bring up other things to discuss or talk about. Along with the professional tone and attitude that users get from ChatGPT, it creates an atmosphere of a mechanical exchange that is only working towards a goal given by the user at the beginning. Despite all of these differences and limitations in terms of the naturalness of the interaction, the conversations with the chatbot are engaging and provide a useful way of extracting information in a short amount of time, neatly packed into the space of a single chat window.

As for ChatGPT's understanding of complex moral issues, the model is clearly designed in such a way so that it does not perpetrate any unwanted bias, even when presented with a biased question or prompt. It is not intentionally trying to sway anyone's judgement. ChatGPT therefore manages to navigate these sensitive topics quite well, in that it does not actually project any singular opinion or viewpoint but rather gives a selection of different ones for the user to consider and choose from. In addition, it does all this while using tentative expressions

and generalizing statements whenever possible, as well as acknowledging cultural sensitivity and stressing thorough consideration. While it is good thing that the AI is not enforcing any ideology of its own, it cannot yet act as a full-fledged moral assistant. Its moral standing is frail and the chatbot can still contradict itself in rare instances. ChatGPT learns from incoming input but is unable to cross-reference it with authentic data. Because of this, it is possible that if certain trends in viewing and resolving moral issues arise, some users will inevitably adopt and discuss them with the model. As a result, the model itself will start adopting a similar viewpoint.

Cooperating with language systems when contemplating difficult decisions may overall give better results – the AI provides as much information pertaining to the issue that it can obtain, analyzes, and summarizes for the human to process more easily, all in a matter of a few seconds. Weighing positives and negatives and quickly arriving at different outcomes can speed up the decision-making process and allow for more consideration time. This kind of teamwork between a human and machine may prove to be the most effective use of the technology at this time. It would not be permissible to rely only on language models or other AI models for moral advice, as it is obvious that although these systems are careful in giving ‘advice’ and not necessarily taking a final stance, some users may still be susceptible to their influence. This most likely comes down to the balanced arguments presented by the AI, as users have the opportunity to pick and select one which best suits their moral code. It is crucial to keep an open-minded approach and realize that consulting an AI on complex issues may only be seeking justification for a decision already made. Enhancing people’s understanding of complex topics or moral issues by use of AI is acceptable; letting oneself be influenced by an AI’s statements is not.

Conclusion

This thesis aimed to show how the interaction with ChatGPT deviates from regular, naturally occurring, human discourse. In addition, it sought to evaluate its linguistic competence as well as its moral stance when addressing morally problematic questions.

By piecing together various sources ranging from technical descriptions of AI language models to linguistic methodological guidebooks, this exploratory analysis was able to uncover several discernible speech patterns and strategies the chatbot uses.

An inspection into the inner workings of ChatGPT was made, providing a closer look at how the chatbot operates and manages its high level of coherence when producing output, as well as its astounding ability to comprehend complex language structures. The thesis compared the current model of ChatGPT with its contemporaries and predecessors and was thus able to illustrate the rapid rate at which the development of language models progresses.

The research also considered the question of accountability and responsibility pertaining to the interaction with AI, stressing how important thorough education on digital literacy is when engaging with AI language models.

With the chosen linguistic methodologies, the analysis was able to show that the approaches need not be used exclusively to analyzing purely human conversation. Borrowing the analytical strategies of CA and DA, the thesis uncovered striking differences as well as several similarities when contrasting ChatGPT's responses to human conversation. Looking at the individual instances of patterns through the lens of the selected methodological approaches primarily intended for human conversation, it became apparent that there were several limitations in the interaction with AI.

Many aspects which normally appear in regularly occurring human conversation were not present at all. Turn-taking and the structure of turns was found to be vastly different, especially in terms of the rigidity that can be observed in the interactions with ChatGPT. The

missing vocal component of conversation, resulting in the omission of features such as back-channel, discourse markers, and the overall mechanical feel of the conversation also brought about the unnaturalness of the conversation.

Several similarities were found when comparing the conversations to institutional interaction. Having the user as the more prominent figure in the exchange, endowed with the right of allocating and initiating turns, ability to control topics and topic change etc. By further comparing the interactions with ChatGPT to institutional interaction, the analysis was able to uncover a knowledge authority the chatbot possesses, creating an ambivalent power dynamic between it and the user. This difference in power distribution was further explored using CDA, highlighting how, although the user is the more powerful figure in the exchange, ChatGPT still has the potential to influence because of its knowledge advantage.

By synthesizing all of the findings, the thesis set out to draw and examine ethical implications that may arise from the interactions with ChatGPT. Overall, the analysis and the subsequent discussion were able to arrive at the conclusion that ChatGPT, although very skilled in its speech and cautious in providing the user with its advice, is not fit as a moral advisor as of yet. The endeavor of combining the careful thinking of humans and the abilities of AI may, as of now, be the better choice for resolving complex moral issues.

In conclusion, the thesis managed to provide a perspective of how ChatGPT steers the conversation of discussing morally problematic questions, showed that the linguistic methodology of CA and DA is applicable beyond just regularly occurring conversation, gave considerations for the ethical implications stemming from the interaction, showed users an overview of how the chatbot operates, and hopefully, provided a better understanding of how to approach working with such AI models.

Works Cited

Boddington, Paula. *AI Ethics: A Textbook*. Springer, 2023.

Brittin, Matt. “Launching the AI Opportunity Initiative for Europe.” *Google*, Google, 12 Feb. 2024, blog.google/around-the-globe/google-europe/google-ai-opportunity-initiative-europe/. Accessed 3 Mar. 2024

Cai, Zhenguang G., et al. “Do Large Language Models Resemble Humans in Language Use?” *arXiv:2303.08014*, 2023, Accessed 2024.

Cambridge Dictionary. *Backchannel* | *English Meaning - Cambridge Dictionary*, dictionary.cambridge.org/dictionary/english/backchannel. Accessed 14 Mar. 2024.

Cambridge Dictionary. *Discourse Markers - Cambridge Grammar*, dictionary.cambridge.org/grammar/british-grammar/discourse-markers-so-right-okay. Accessed 13 Mar. 2024.

Cambridge Dictionary. *Node* | *English Meaning - Cambridge Dictionary*, dictionary.cambridge.org/dictionary/english/node. Accessed 27 Mar. 2024.

Chafe, Wallace. “Constraining and Guiding the Flow of Discourse.” *The Handbook of Discourse Analysis*, 2nd ed., vol. 2, Wiley Blackwell, 2015, pp. 391–405.

Chude, Emmanuel. “GPT-3.5 and GPT-4 Comparison:” *Medium*, Medium, 4 Aug. 2023, medium.com/@chudeemmanuel3/gpt-3-5-and-gpt-4-comparison-47d837de2226. Accessed 10 Mar. 2024

Clayman, Steven E. “Footing in the Achievement of Neutrality: The Case of News-Interview Discourse .” *Talk at Work: Interaction in Institutional Settings*, edited by Paul Drew and John Heritage, Cambridge University Press, 1993, pp. 163–198.

Clift, Rebecca. *Conversation Analysis*. Cambridge University Press, 2016.

Codecademy Team. “Setting Parameters in OpenAI.” *Codecademy*, codecademy.com/article/setting-parameters-in-open-ai. Accessed 23 Apr. 2024.

Dijk, Teun Van. “Critical Discourse Analysis.” *The Handbook of Discourse Analysis*, edited by Deborah Tannen et al., 2nd ed., II, Wiley Blackwell, 2015, pp. 466–485.

Drew, Paul, and John Heritage. "Analyzing Talk at Work: An Introduction." *Talk at Work: Interaction in Institutional Settings*, edited by Paul Drew and John Heritage, Cambridge University Press, 1993, pp. 3–65.

Eggins, Suzanne, and Diana Slade. *Analysing Casual Conversation*. Equinox, 2005.

Goffman, Erving. *Forms of Talk*. University of Pennsylvania Press, 1981.

Grice, Herbert Paul. "Logic and Conversation." *Syntax and Semantics: Speech Acts*, edited by Peter Cole, vol. 3, Elsevier, 1975, pp. 41–58.

Hardesty, Larry. "Explained: Neural Networks." *MIT News, Massachusetts Institute of Technology*, 14 Apr. 2017, news.mit.edu/2017/explained-neural-networks-deep-learning-0414. Accessed 15 Mar. 2024.

Have, Paul Ten. "Conversation Analysis Versus Other Approaches to Discourse." *Forum: Qualitative Social Research*, vol. 7, no. 3, 2006.

Have, Paul Ten. *Doing Conversation Analysis: A Practical Guide*. 2nd ed., SAGE Publications Ltd, 2007.

Hendrycks, Dan, et al. "Measuring Massive Multitask Language Understanding." *arXiv:2009.03300*, 2020, Accessed 2024

Hoey, Michael. *Textual Interaction: An Introduction to Written Discourse Analysis*. Taylor and Francis, 2005.

Jones, Haydn. "How Does AI Really Work? Comparing Neural Networks Gives a Peek into the Black Box." *Los Alamos National Laboratory*, 16 Nov. 2022, discover.lanl.gov/news/1117-neural-networks/. Accessed 7 Mar. 2024

Kelk, Ian. "How ChatGPT Fools Us into Thinking We're Having a Conversation: Short Attention." *Ian Kelk*, 26 Nov. 2023, kelk.ai/blog/how-chatgpt-fools-us?trk=article-ssr-frontend-pulse_little-text-block. Accessed 20 Feb. 2024

Krügel, Sebastian, et al. "The Moral Authority of ChatGPT." *arXiv:2301.07098*, 2023, Accessed 23 Feb. 2024.

Lewis, Nell. "Robot Friends: Why People Talk to Chatbots in Times of Trouble." *CNN*, Cable News Network, 19 Aug. 2020, edition.cnn.com/2020/08/19/world/chatbot-social-anxiety-spc-intl/index.html. Accessed 9 Apr. 2024

McGonagle, John, et al. "Backpropagation: Brilliant Math & Science Wiki." *Brilliant*, brilliant.org/wiki/backpropagation/. Accessed 16 Mar. 2024.

McTear, Michael. *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*. Morgan & Claypool, 2021.

Nesler, Mitchell S., et al. "The Effect of Credibility on Perceived Power." *Journal of Applied Sociology*, vol. 23, no. 17, 1993, pp. 1407–1425.

OpenAI. "ChatGPT — Release Notes | OpenAI Help Center." *ChatGPT Release Notes*, help.openai.com/en/articles/6825453-chatgpt-release-notes. Accessed 12 Feb. 2024.

OpenAI. *OpenAI*, 2015-2024, openai.com. Accessed 3 Mar. 2024

OpenAI. "What Is ChatGPT? | OpenAI Help Center." *What Is ChatGPT?*, OpenAI, help.openai.com/en/articles/6783457-what-is-chatgpt. Accessed 4 Mar. 2024.

Pico. "Overfitting, Variance, Bias and Model Complexity in Machine Learning." *Pico*, www.pico.net/kb/overfitting-variance-bias-and-model-complexity-in-machine-learning/. Accessed 16 Mar. 2024.

Piper, Kelsey. "Why Is Meta's New AI Chatbot so Bad?" *Vox*, 21 Aug. 2022, [vox.com/future-perfect/23307252/meta-facebook-bad-ai-chatbot-blenderbot](https://www.vox.com/future-perfect/23307252/meta-facebook-bad-ai-chatbot-blenderbot). Accessed 9 Mar. 2024

Pomerantz, Anita. "Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes." *Structures of Social Action*, edited by J. M. Atkinson and John Heritage, Cambridge University Press, 1984, pp. 57–101.

Sacks, Harvey, et al. "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *Language*, vol. 50, no. 4, Dec. 1974, p. 696.

Schaffner Bofill, Pablo. "Does ChatGPT Really Think?" *LinkedIn*, 22 July 2023, [linkedin.com/pulse/does-chatgpt-really-think-pablo-schaffner-bofill](https://www.linkedin.com/pulse/does-chatgpt-really-think-pablo-schaffner-bofill). Accessed 20 Mar. 2024

Sidnell, Jack. *Conversation Analysis: An Introduction*. Wiley-Blackwell, 2010.

Singh, Shubham, "ChatGPT Statistics - Users, Revenue (April 2024)." *DemandSage*, 19 Mar. 2024, demandsage.com/chatgpt-statistics/. Accessed 20 April 2024

Suchman, Lucille Alice. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, 1987.

Teflpedia. "Echo Utterance." *Teflpedia*, teflpedia.com/Echo_utterance. Accessed 29 Mar. 2024.

"What Are Large Language Models?: Nvidia Glossary." *NVIDIA*, nvidia.com/en-us/glossary/large-language-models/. Accessed 28 Mar. 2024.

"What Is Natural Language Processing?" *IBM*, 19 Mar. 2024, ibm.com/topics/natural-language-processing. Accessed 12 Mar. 2024

Woodside, Thomas, and Helen Toner. "How Developers Steer Language Model Outputs: Large Language Models Explained, Part 2." *Center for Security and Emerging Technology*, 8 Mar. 2024, cset.georgetown.edu/article/how-developers-steer-language-model-outputs-large-language-models-explained-part-2/. Accessed 1 Apr. 2024

Appendix: Data

1. Is it ever acceptable to lie?

1.1. ChatGPT: The ethics of lying can be a complex and debated topic. Different moral frameworks and ethical theories offer varying perspectives on whether lying can be justified in certain situations. Here are a few viewpoints to consider:

1. **Deontological Ethics:** Some ethical systems, such as deontological ethics, argue that lying is inherently wrong and should be avoided regardless of the consequences. According to this perspective, lying is seen as a violation of a moral duty or a breach of trust.
2. **Utilitarianism:** Utilitarianism focuses on the consequences of actions. From a utilitarian standpoint, lying might be considered acceptable if it leads to the greatest overall happiness or the least harm. For example, lying to save someone's life or to prevent a great injustice might be justified under this framework.
3. **Virtue Ethics:** Virtue ethics emphasizes the development of moral character. Proponents of virtue ethics argue that lying undermines virtues such as honesty, integrity, and trustworthiness. From this perspective, lying is generally discouraged as it goes against the cultivation of virtuous traits.
4. **Contextual Ethics:** Contextual ethics take into account the specific circumstances surrounding an action. Some proponents of contextual ethics argue that lying can be acceptable in certain situations, such as when it is done to protect someone from harm or to promote a greater good. The key is to carefully consider the consequences and the potential impact on individuals involved.

Ultimately, whether lying is deemed acceptable or not often depends on the specific circumstances, cultural norms, personal values, and the ethical framework one adopts. It is important to carefully consider the potential consequences, as well as the impact on trust and relationships, when making decisions about honesty and deception.

1.2. User: Among all of these perspectives, is any one of them more prevalent than the others in our society?

1.3. ChatGPT: In society, the prevalence of perspectives on lying can vary depending on cultural, social, and individual factors. It is difficult to determine a single prevailing perspective, as opinions on lying can differ greatly among individuals and communities. However, it is worth noting that cultural norms, legal systems, and religious beliefs can influence societal attitudes toward lying.

In many cultures, honesty and truthfulness are valued virtues, and lying is generally discouraged. These cultural norms often emphasize the importance of trust, integrity, and maintaining social harmony. In such societies, deontological ethics, which prioritize moral duties and honesty, may have a stronger influence on individuals' attitudes toward lying.

However, in certain contexts, such as in legal systems, there may be exceptions or allowances for lying. For example, legal frameworks may recognize the concept of perjury, where lying under oath is considered a punishable offense. In these cases, lying is generally discouraged, and there are legal consequences for dishonesty.

It is also important to recognize that societal attitudes toward lying can evolve over time and may vary across different groups or subcultures within a society. The prevalence of different perspectives on lying can change with societal shifts, cultural changes, and advances in technology and communication.

Overall, while cultural norms and personal values play a significant role in shaping societal perspectives on lying, it is challenging to identify a single prevailing perspective, as opinions can vary widely among individuals and communities.

1.4. User: Would the world be a better place if we lacked the capacity to lie?

1.5. ChatGPT: The hypothetical scenario of a world where humans lack the capacity to lie is intriguing, but it's challenging to make definitive conclusions about whether it would unequivocally make the world a better place. Here are a few factors to consider:

1. **Trust and Relationships:** Honesty is crucial for building trust and maintaining healthy relationships. If everyone were incapable of lying, it could potentially enhance trust among individuals and foster more genuine and transparent interactions.
2. **Social Conventions and Politeness:** While lies can be harmful, they can also serve social functions. Polite white lies or small talk often grease the wheels of social interaction and help maintain harmony. In a world without lies, social conventions and politeness might need to be reimaged, as it could lead to more blunt and potentially uncomfortable exchanges.
3. **Privacy and Personal Boundaries:** The ability to withhold information or maintain privacy can be important for individuals. In a world without lies, it might become more challenging to protect personal boundaries and keep sensitive information confidential.
4. **Protection and Safety:** In certain situations, lying can be a form of self-defense or protection. For example, lying to an aggressor or concealing information to prevent harm may be seen as justifiable. In a world without lies, alternative strategies would need to be employed to address such situations.
5. **Art, Fiction, and Creativity:** Lying and storytelling are intricately connected. Fiction, literature, and imaginative works often rely on the creative use of untruths. A world without lies may impact the richness and diversity of artistic expression.

Ultimately, the impact of lacking the capacity to lie would depend on various factors, including the development of alternative communication methods, adaptations in social conventions, and the redefinition of privacy and personal boundaries. While honesty and transparency are generally valued, it is difficult to predict all the potential consequences and trade-offs of a world without lies.

1.6. User: Based on what you said, is lying therefore an essential part of us?

1.7. ChatGPT: Lying, as a behavior, is indeed a complex and multifaceted aspect of human nature. While lying is generally considered to be ethically problematic, it is also a behavior that humans engage in for various reasons. Here are a few points to consider:

1. **Cognitive Abilities:** The capacity to lie is connected to higher cognitive functions, such as theory of mind and the ability to understand and manipulate information.

Lying involves intentionally distorting or concealing the truth, which requires cognitive processes such as planning, self-awareness, and perspective-taking.

2. **Social Dynamics:** Lying can serve important social functions, such as preserving social harmony, protecting privacy, or avoiding conflict. Polite white lies, for instance, can be used to maintain positive relationships and avoid unnecessary hurt feelings. While these lies are not ideal, they are often seen as a social lubricant.
3. **Self-Presentation and Identity:** Lying can be driven by the desire to present oneself in a certain way or to protect one's self-image. People may lie to enhance their social status, gain approval, or avoid judgment. These lies are often tied to concerns about acceptance and self-worth.
4. **Survival and Protection:** Lying can be a defense mechanism used to protect oneself or others from harm. In certain situations, deception may be employed to avoid danger or prevent negative consequences. This can be seen, for example, in situations where individuals lie to protect someone's safety or to escape a threatening situation.

Considering these aspects, lying can be viewed as an inherent part of human nature, stemming from our cognitive abilities, social dynamics, and the complex nature of human relationships. However, it is important to note that the inherent presence of lying does not necessarily imply that lying is morally acceptable or that it should be condoned. Ethical considerations and the potential consequences of lying should always be taken into account when evaluating the moral implications of this behavior.

1.8. User: Is everything you listed what you consider to be the most important information regarding this topic? Can I consider it finished?

1.9. ChatGPT: The information I provided encompasses key points to consider when discussing the topic of lying and its complexities. However, it is important to note that discussions on ethics and human behavior can be nuanced and multifaceted, and there is always potential for further exploration and different perspectives.

While I have provided a comprehensive overview of the topic based on the available information, it is ultimately up to you to decide if you consider the matter closed. If you have any specific follow-up questions or if there are particular aspects you would like to delve into further, I am here to assist you.

2. Do evil people exist?

2.1. ChatGPT: The concept of "evil" is subjective and can vary depending on cultural, ethical, and philosophical perspectives. However, it is generally understood that some individuals engage in harmful actions, exhibit malicious intent, or disregard the well-being of others. Whether or not one labels these individuals as "evil" is a matter of personal interpretation.

Throughout history, there have been individuals who committed heinous acts that are widely regarded as evil, such as Adolf Hitler, Joseph Stalin, or serial killers like Ted Bundy. These individuals caused immense suffering and harm to others intentionally. However, it's important to note that human behavior is complex, and people's actions can be influenced by

various factors, including upbringing, mental health, societal influences, and personal circumstances.

It is crucial to approach the concept of "evil" with caution and recognize that individuals' actions can stem from a combination of factors, including psychological disorders, traumatic experiences, or societal influences. It is generally more constructive to focus on understanding the causes and motivations behind harmful actions in order to prevent and address such behavior, rather than labeling individuals as inherently "evil."

2.2. User: Setting aside any external factors which might influence the individual's actions, if someone caused immense suffering to a large number of people or other living beings, the environment etc., would they be considered evil?

2.3. ChatGPT: The term "evil" is often used to describe individuals who intentionally cause harm, suffering, or destruction on a significant scale. If someone consistently and deliberately inflicts immense suffering on a large number of people, animals, or the environment without any mitigating factors, many people would likely perceive their actions as evil.

However, it's important to note that applying the label of "evil" to someone's actions is a subjective judgment that depends on individual perspectives, cultural norms, and ethical frameworks. Different people may have varying interpretations and definitions of evil based on their beliefs, values, and cultural backgrounds.

Furthermore, it is worth considering that the concept of evil is often associated with a sense of moral responsibility. In some philosophical and ethical frameworks, individuals may be seen as evil if they possess the capacity for empathy and understanding of the consequences of their actions but still choose to inflict harm.

Ultimately, whether an individual's actions are deemed evil depends on personal and societal interpretations of morality and the magnitude and nature of the harm caused.

2.4. User: But generally, you would agree that if large-scale destruction or harm is done by one person, then the things they did which led to that destruction or harm would be regarded as evil?

2.5. ChatGPT: Yes, it is generally true that if an individual causes significant destruction or harm on a large scale, their actions would likely be regarded as evil by many others. When someone's actions result in widespread suffering, loss of life, or severe negative consequences for individuals, communities, or the environment, it tends to elicit strong moral condemnation from society.

Acts such as genocide, terrorism, or large-scale environmental destruction are often considered evil due to the immense suffering they inflict and the disregard for the well-being of others or the environment. These actions are often viewed as a violation of fundamental moral principles and ethical norms that emphasize the value of human life, dignity, and the protection of the environment.

However, it's important to remember that the perception of evil can still vary among individuals and cultures. Different philosophical, cultural, or religious perspectives may shape how people interpret and evaluate the moral character of such actions. Nevertheless, in cases

of widespread destruction or harm, it is more likely that the actions will be widely seen as morally reprehensible and labeled as evil.

2.6. User: Is this everything I need to know? Would you like to add any additional information to this topic?

2.7. ChatGPT: While the previous information covers the basics of the concept of evil and how it is generally understood, it is worth mentioning a few additional points:

1. **Moral Relativism:** The perception of evil can vary across different cultures, societies, and individuals due to moral relativism. Moral relativism posits that moral judgments are subjective and can differ based on cultural norms, values, and individual perspectives. Therefore, what one culture or person considers evil may not be universally accepted as such.
2. **Intentions and Consequences:** When evaluating the morality of an action, both intentions and consequences are often taken into account. Some argue that evil acts require a deliberate intention to cause harm, while others focus more on the actual harm caused, regardless of intention. The interplay between intentions and consequences can influence judgments of evil.
3. **Psychological Factors:** The field of psychology explores various factors that may contribute to harmful or evil behavior, such as psychopathy, personality disorders, trauma, or environmental influences. Understanding these factors can shed light on the complexities of human behavior and help develop strategies for prevention and rehabilitation.
4. **Moral Responsibility:** Assessing moral responsibility in cases of evil acts can be challenging. Some individuals may argue that factors such as mental illness or external influences mitigate personal responsibility, while others may emphasize individual agency and accountability. This topic is a subject of ongoing debate among ethicists, philosophers, and legal scholars.

In summary, the concept of evil is multifaceted and subject to interpretation. While there are commonalities in how certain actions are perceived as evil, individual perspectives, cultural differences, and contextual considerations shape the understanding and judgment of evil acts.

3. Is it ever acceptable to kill someone?

3.1. ChatGPT: The question of whether it is ever acceptable to kill someone is a complex ethical and philosophical issue that has been debated by scholars, philosophers, and legal experts for centuries. Different cultures, legal systems, and ethical frameworks may have varying perspectives on this matter.

In general, the taking of another person's life is considered a grave and irreversible act, and most societies have established laws and moral codes that prohibit murder. However, there are situations where the use of lethal force is considered justifiable or legally permissible. Here are a few examples:

1. **Self-defense:** Many legal systems recognize the right to defend oneself or others from immediate and serious harm. If a person reasonably believes that their life or the lives

of others are in immediate danger, they may be legally justified in using lethal force to protect themselves or others.

2. War and armed conflict: In certain circumstances, killing during wartime may be legally sanctioned under international humanitarian law. However, this is subject to specific rules and principles governing the conduct of armed conflict, such as distinguishing between combatants and civilians, proportionality, and military necessity.
3. Capital punishment: Some countries retain the death penalty as a form of punishment for certain crimes. The acceptability and legality of capital punishment vary widely across different jurisdictions and are the subject of ongoing debate.
4. Euthanasia: In some countries or regions, euthanasia or physician-assisted suicide may be legally permitted under certain circumstances. This typically involves terminally ill patients with unbearable suffering who request medical assistance to end their lives.

It's important to note that the acceptability of killing someone is highly subjective and context-dependent. Ethical considerations, cultural norms, religious beliefs, and legal systems all play a role in shaping societal attitudes toward taking a life. Furthermore, discussions around this topic often involve complex moral dilemmas and personal convictions.

3.2. User: Should changes be made so that killing is minimized and is ultimately only a choice and not a necessity in certain cases? Or is killing an inherent part of human nature?

3.3. ChatGPT: The question of whether changes should be made to minimize the necessity of killing in certain cases is a matter of personal opinion and can vary depending on individual beliefs, values, and cultural perspectives. There is no definitive answer to whether killing is an inherent part of human nature, as opinions on human nature itself differ among scholars and philosophers.

Some argue that human beings have an innate capacity for violence and aggression, pointing to historical evidence of conflicts, wars, and instances of harm caused by humans. Others contend that while violence and aggression may exist in human behavior, it does not necessarily imply an inherent inclination towards killing.

From an ethical standpoint, many people aspire to create a society that minimizes the necessity of killing whenever possible. Efforts to promote non-violence, conflict resolution, and the development of alternative methods for resolving disputes are common endeavors in this regard. By fostering empathy, compassion, and understanding, individuals and communities can work towards reducing violence and promoting peaceful resolutions.

Technological advancements and social progress have indeed led to changes in various areas, such as the development of non-lethal weapons, improvements in self-defense techniques, and advancements in medical science that aim to alleviate suffering and provide alternatives to lethal means.

Ultimately, whether killing remains a necessity or can be further minimized in certain cases depends on societal values, legal frameworks, and collective efforts to find alternatives to violence and killing. As our understanding of human nature and our capacity for moral reasoning advances, it is possible that societal norms and structures will continue to evolve in pursuit of reducing the necessity for taking human lives.

3.4. User: Wouldn't the Earth become overpopulated if there wasn't a certain amount of killing? As bad as it sounds, it seems like reducing the population in this way is a necessary thing.

3.5. ChatGPT: The question of overpopulation and its potential consequences is a complex and contentious issue. While it is true that the Earth's population has been steadily increasing, addressing this concern requires considering multiple factors and adopting a holistic approach.

Firstly, it's important to note that the Earth's carrying capacity—the maximum number of individuals that can be sustained by available resources—is a subject of debate. The concept of overpopulation is not solely determined by absolute numbers but also by factors such as resource distribution, consumption patterns, technological advancements, and environmental sustainability.

Efforts to manage population growth typically focus on improving access to education, healthcare, and family planning services, as studies have shown that these factors are associated with lower birth rates. Promoting sustainable development, reducing poverty, and addressing inequalities can also contribute to achieving a more balanced population.

Regarding the notion of reducing the population through killing, it is essential to approach this question from an ethical and human rights perspective. The intentional killing of individuals as a means of population control raises significant moral concerns. Human rights principles emphasize the inherent dignity and value of every human life and advocate for the protection and preservation of human rights for all individuals.

It is more productive and morally sound to address issues related to population growth by focusing on sustainable development, education, healthcare, and access to family planning. By promoting these factors, societies can strive for a more balanced and equitable future without resorting to harmful practices.

Furthermore, it's worth noting that many regions of the world are already experiencing declining birth rates, and some countries are facing the challenges associated with an aging population. Thus, population dynamics are complex and can vary greatly across different regions and countries.

In summary, addressing concerns related to population growth and environmental sustainability requires comprehensive strategies that prioritize human rights, education, healthcare, and sustainable development. By adopting such approaches, societies can work towards a more balanced future without compromising fundamental ethical principles.

3.6. User: I didn't mean intentional reducing of the population through killing, I only meant that because there already is a lot of killing happening in the world, it is keeping the population from rapidly increasing. Because of that I wonder whether or not it would be beneficial to create as many countermeasures to prevent killing altogether.

3.7. ChatGPT: I apologize for the misunderstanding. If you are referring to the natural occurrence of killings in the world and whether preventing or reducing them would lead to rapid population growth, it's important to note that the impact of such killings on overall population dynamics is complex and multifaceted.

While certain forms of mortality, such as deaths due to violence or conflicts, can contribute to a reduction in population growth, it is crucial to consider the underlying causes and broader implications of these events. Violence and conflicts are often associated with significant human suffering, displacement, and social unrest, which can have far-reaching negative consequences for communities and societies.

Efforts to prevent and reduce violence, conflict, and killings should focus on addressing the root causes and promoting peaceful resolutions rather than solely considering the potential impact on population growth. Creating countermeasures to prevent violence and killings is generally seen as a positive endeavor to foster safer and more peaceful societies.

However, it's important to approach this issue with caution and recognize that the causes of violence and killings are multifaceted, ranging from socioeconomic factors to political conflicts and cultural dynamics. To effectively reduce violence and its impact on populations, comprehensive approaches are needed, including conflict resolution, addressing inequalities, promoting social justice, and improving access to education and economic opportunities.

In summary, while reducing violence and killings is generally a worthy goal to create safer and more peaceful societies, it is essential to consider the underlying causes and take a holistic approach to address them. This approach focuses on promoting peace, addressing root causes of conflicts, and fostering social and economic well-being for all individuals.

3.8. User: Should we close this topic or is there anything else you wish to add?

3.9. ChatGPT: If you feel that your question has been adequately addressed, we can close this topic. However, if you have any further questions or if there's anything else you would like to discuss, please feel free to let me know. I'm here to help!

3.10. User: I only wondered whether or not you have any other additional information which you consider to be important for this topic.

3.11. ChatGPT: Certainly! Here are a few additional points that may be relevant to the topic of killing and its implications:

1. **Non-lethal alternatives:** In many situations where lethal force might be considered, non-lethal alternatives can be explored. Law enforcement agencies, for example, often use non-lethal weapons such as tasers or rubber bullets to subdue individuals while minimizing the risk of fatal outcomes.
2. **Rehabilitation and restorative justice:** Instead of focusing solely on punishment or killing as a means of addressing crime, some justice systems emphasize rehabilitation and restorative justice approaches. These approaches aim to address the root causes of criminal behavior, promote healing for victims, and reintegrate offenders back into society.
3. **Mental health support:** Mental health issues can contribute to violent behavior. Ensuring access to mental health support services, early intervention, and destigmatizing mental health can play a role in preventing violent acts and addressing the underlying factors that may lead to them.
4. **Conflict resolution and diplomacy:** Promoting dialogue, negotiation, and peaceful means of resolving conflicts can help prevent violence and minimize the need for

killing in many situations. International diplomacy and mediation efforts play a vital role in mitigating conflicts between nations and fostering peaceful resolutions.

5. Societal and structural changes: Addressing systemic issues such as poverty, inequality, social marginalization, and lack of opportunities can help reduce violence and its underlying causes. By creating more equitable societies, the risk factors that contribute to violence can be diminished.

These additional points highlight the importance of exploring alternatives to killing, addressing root causes, and promoting proactive measures that contribute to safer and more peaceful communities. By focusing on prevention, intervention, and creating conducive environments, societies can work towards reducing violence and its consequences.

4. Should the death penalty be legal?

4.1. ChatGPT: The question of whether the death penalty should be legal is a complex and highly debated topic. Different societies and individuals hold diverse perspectives on this issue. I can provide you with some arguments commonly put forth by both proponents and opponents of the death penalty to help you understand the various viewpoints.

Arguments in favor of the death penalty often include:

1. Deterrence: Supporters argue that the threat of capital punishment deters potential criminals from committing heinous crimes, as they fear the ultimate consequence.
2. Retribution: Some people believe that the death penalty is a just response to certain crimes, providing a form of retribution for the victims and their families.
3. Public Safety: Supporters contend that executing dangerous individuals permanently removes the threat they pose to society, ensuring public safety.

On the other hand, opponents of the death penalty raise the following points:

1. Irreversibility: The most significant argument against the death penalty is the irreversible nature of the punishment. Once an execution is carried out, there is no opportunity to correct any mistakes made in the legal process, potentially leading to the execution of innocent individuals.
2. Human Rights: Critics argue that the death penalty violates the inherent right to life and the prohibition against cruel and inhumane punishment, as enshrined in international human rights standards.
3. Cost and Legal Process: Capital punishment can be financially burdensome due to lengthy legal procedures, appeals, and the cost of maintaining death row inmates. Critics contend that these resources could be better allocated to crime prevention, rehabilitation, or supporting victims' families.

It's important to note that opinions on the death penalty vary widely depending on cultural, societal, and personal beliefs. Public sentiment and legal frameworks differ across countries and jurisdictions. Ultimately, the decision regarding the legality of the death penalty rests with each individual society or country, reflecting its values and priorities.

4.2. User: In terms of preventing criminals from committing any further crimes, isn't keeping them in prison (for life) enough for that purpose?

4.3. ChatGPT: Yes, keeping individuals convicted of heinous crimes in prison for life is considered by many as an alternative to the death penalty for achieving the goal of preventing them from committing further crimes. Life imprisonment without the possibility of parole ensures that the convicted person remains incarcerated for the remainder of their life.

Advocates for life imprisonment argue that it serves the purpose of protecting society by removing dangerous individuals from the community while still respecting their fundamental right to life. It also allows for the possibility of rectifying any errors made in the legal process, as individuals can be released if new evidence emerges proving their innocence.

Furthermore, life imprisonment provides the opportunity for rehabilitation and personal growth, even if it is within the confines of prison. Supporters of this approach believe in the potential for personal transformation and believe that society should focus on the rehabilitation and reintegration of individuals into society, rather than taking their lives.

It's important to note that the effectiveness of life imprisonment as a deterrent or means of preventing further crimes is also a topic of debate. Some argue that the possibility of release or escape from prison may still pose a risk, while others contend that the overall rate of recidivism (re-offending) among lifers is relatively low.

Ultimately, the decision between life imprisonment and the death penalty involves weighing various factors, including moral, ethical, legal, and societal considerations. Different jurisdictions adopt different approaches based on their respective beliefs and values.

4.4. User: As for the argument against, that the process can be costly, isn't it actually more costly to have the criminals who receive a life sentence in prison? It seems like as if this argument should be the one which advocates for the death penalty, as the process is over once the criminal is executed.

4.5. ChatGPT: The cost argument surrounding the death penalty is indeed a point of contention in the debate. While it is true that maintaining a prisoner on death row can be expensive, studies have shown that the long legal process and appeals associated with capital punishment contribute significantly to its high costs. These expenses arise from the need for specialized legal representation, extended trials, and additional security measures.

On the other hand, the cost of keeping an inmate in prison for life without parole can also be substantial over time. However, it is important to consider that the cost of life imprisonment is spread out over many years or even decades, whereas the costs associated with the death penalty are typically concentrated in the pretrial, trial, and appellate stages.

Moreover, it is worth noting that cost should not be the sole determining factor in shaping a criminal justice system. The primary objective of the justice system should be to ensure fairness, justice, and public safety, rather than solely seeking the most cost-effective solution.

Ultimately, the cost argument alone may not be sufficient to support or oppose the death penalty. It is a complex issue that involves a careful evaluation of various factors, including

moral, ethical, legal, and practical considerations, as well as the potential for wrongful convictions and the irreversible nature of the punishment.

4.6. User: Would the death penalty be considered worse for a criminal? Or would it be life imprisonment? Considering that in both cases, the life of said criminal is, so to say, over. Wouldn't the death penalty almost be seen as a merciful way out?

4.7. ChatGPT: The perception of whether the death penalty or life imprisonment is considered worse for a criminal can vary depending on cultural, personal, and philosophical perspectives. It's important to note that individual opinions on this matter can differ significantly.

For some individuals facing life imprisonment, the prospect of spending the rest of their lives confined to a prison cell without the possibility of freedom may be viewed as a harsher punishment compared to the death penalty. Life imprisonment can entail the loss of personal autonomy, separation from loved ones, and a lifetime of limited opportunities and freedoms.

On the other hand, proponents of the death penalty argue that it is a more severe punishment because it involves the intentional taking of a person's life by the state. They believe that the death penalty serves as the ultimate form of retribution and justice for the heinous crimes committed.

However, it's important to recognize that the perception of what constitutes a more severe punishment is subjective and can vary from person to person. Different individuals may have different views on what they consider to be a more just or appropriate punishment for certain crimes.

It's worth noting that discussions of the severity of punishment often involve moral, ethical, and philosophical considerations. These discussions should also take into account the potential for wrongful convictions, the irreversibility of the death penalty, and the evolving standards of human rights and dignity.

Ultimately, the perception of what is worse for a criminal, the death penalty or life imprisonment, is subjective and can vary based on individual perspectives and cultural values.

4.8. User: Is this all you wish to add to this matter?

4.9. ChatGPT: Yes, I have provided a comprehensive overview of the arguments surrounding the question of whether the death penalty should be legal, including various perspectives and considerations. If you have any specific follow-up questions or if there's anything else you would like to discuss on this topic or any other, please let me know, and I'll be happy to assist you further.

5. Is any one person's life ever worth more than another's?

5.1. ChatGPT: As an AI language model, I don't possess personal opinions or beliefs, but I can provide you with some perspectives on the matter.

The question of whether one person's life is worth more than another's is complex and subjective. From an ethical standpoint, many people adhere to the principle of equal moral value, which suggests that all human lives have equal worth and should be treated with equal respect and dignity. This perspective is often based on the idea that every individual possesses inherent rights and deserves equal consideration.

However, it's important to recognize that societal and cultural factors can influence perceptions of value. In some contexts, certain professions or positions may be deemed more valuable or essential due to the potential impact on a larger number of people or the significance of their contributions. For example, a doctor who saves lives might be considered more valuable than an individual who performs a different job.

Nevertheless, it is crucial to emphasize that these distinctions should not diminish the inherent worth of any individual. Each person's life has intrinsic value, and it is generally considered ethically important to respect and protect that value.

It's worth noting that discussions around the value of human life can vary based on cultural, religious, philosophical, or legal perspectives. Different ethical frameworks may propose alternative approaches to determining the worth of a person's life, but the concept of inherent human dignity is often a central theme.

Ultimately, the belief in the equal value of all human lives forms the foundation for many human rights principles and ethical considerations.

5.2. User: Should any external or internal qualities of a person affect their individual worth? Or should we view this issue only from the perspective of us all being sentient species and therefore worthy of being treated the same way?

5.3. ChatGPT: The question of whether external or internal qualities should affect a person's individual worth is a subject of moral and philosophical debate. Different ethical frameworks offer different perspectives on this matter.

One perspective, rooted in the principles of human rights and equality, argues that a person's individual worth should not be determined by external or internal qualities. This perspective asserts that all individuals, by virtue of being members of the human species and possessing sentience, have inherent worth and deserve to be treated with equal dignity and respect. According to this view, factors such as race, gender, socioeconomic status, physical abilities, or intelligence should not diminish or enhance a person's worth.

On the other hand, alternative perspectives suggest that certain internal or external qualities might influence a person's worth. For example, some philosophical or religious frameworks propose that moral character, achievements, contributions to society, or adherence to certain principles could impact a person's value.

It's important to note that while these perspectives exist, they can lead to ethical dilemmas and potential discrimination if taken to extremes. Judging an individual's worth based on external qualities can perpetuate unfair biases and social inequalities.

In many human rights frameworks, the principle of equal moral value is upheld, emphasizing that all individuals possess an inherent worth that transcends external or internal qualities.

This approach promotes the idea that each person's worth is fundamentally rooted in their shared humanity and capacity for subjective experiences, regardless of any specific attributes or characteristics.

Ultimately, how we view individual worth is a deeply philosophical and ethical question, and different people and cultures may hold diverse beliefs on the matter. The recognition of our shared humanity and the commitment to treating all individuals with dignity and respect are fundamental principles that guide many ethical frameworks.

5.4. User: So in a hypothetical scenario with two individuals, where one of them is someone who has greatly contributed to society, is still actively contributing, and is needed for achieving a prosperous future, and the other is someone who does not contribute to society in any way and is not significant for the future of the world whatsoever, and one of them had to be sacrificed, what steps should be taken to ensure the best outcome?

5.5. ChatGPT: The hypothetical scenario you presented is a morally challenging one. However, it's important to note that sacrificing one individual's life for the benefit of another is generally considered ethically problematic and conflicts with the principles of human rights and the inherent value of every individual.

Instead of focusing on sacrificing one person for the benefit of another, it is often more constructive to explore alternative approaches that uphold the principles of fairness and human dignity. Here are a few possible steps that could be considered:

1. Collaboration and dialogue: Engage in open and inclusive discussions involving all stakeholders to explore potential solutions that maximize benefits and minimize harm to all parties involved.
2. Ethical considerations: Consult ethical frameworks, such as utilitarianism, deontology, or virtue ethics, to guide decision-making. These frameworks offer different perspectives on evaluating the moral consequences, principles, or character traits relevant to the situation.
3. Resource allocation: Assess available resources and explore ways to optimize their distribution in order to meet societal needs and maximize the overall well-being of individuals.
4. Long-term planning: Consider long-term implications and invest in strategies that address the needs of both individuals and society as a whole. This may involve nurturing a culture that encourages personal development and contribution while providing support and opportunities for everyone.
5. Human dignity: Uphold the principle of equal moral value, recognizing the inherent worth of every individual, irrespective of their current contributions or significance. Strive to protect the dignity and rights of all individuals involved.

It's important to note that sacrificing one individual for the benefit of another is generally seen as an extreme measure and should only be considered in highly exceptional situations, such as cases of immediate self-defense or preventing extreme harm to a large number of people. Such decisions, however, should be guided by legal systems and societal consensus.

In summary, it is preferable to explore alternative solutions that respect the rights and dignity of all individuals involved, rather than resorting to sacrificing one person for the perceived benefit of another.

5.6. User: I think that in general most people would choose without hesitation for the 'less important' individual to be sacrificed, what do you think of that?

5.7. ChatGPT: It's important to approach discussions about sacrificing individuals with caution, as they involve complex ethical considerations and can have significant implications. While it is true that in certain hypothetical scenarios, some people might argue for sacrificing the "less important" individual, it is crucial to critically examine the underlying values and ethical principles at play.

In many ethical frameworks, including those based on human rights and the principle of equal moral value, sacrificing one individual for the benefit of another is generally considered ethically problematic. These frameworks emphasize the inherent worth and dignity of every human being, regardless of their perceived significance or contributions to society.

It is worth noting that the knee-jerk response to sacrifice the "less important" individual may arise from various factors, including utilitarian reasoning, immediate emotional reactions, or cultural biases. However, relying solely on such responses can overlook the complexities and potential consequences of such actions.

Engaging in thoughtful ethical deliberation allows for a more comprehensive consideration of the implications of sacrificing one individual for the perceived benefit of another. It involves examining the values and principles underlying such decisions, the potential impact on social cohesion and fairness, and the long-term consequences for societal trust and well-being.

Ultimately, ethical discussions should encourage critical thinking, empathy, and the exploration of alternative approaches that respect the inherent worth and dignity of all individuals involved.

5.8. User: Shouldn't we always consider the long-term and large-scale effects of such a decision though? With that approach I believe many of the things you mentioned would be discarded as unimportant.

5.9. ChatGPT: Considering the long-term and large-scale effects of decisions is indeed crucial. When faced with complex choices, it is important to take into account the broader consequences and impacts on society as a whole. However, it is equally important to ensure that such considerations are compatible with fundamental ethical principles and the protection of individual rights and dignity.

In evaluating the long-term effects, it is essential to adopt a comprehensive perspective that encompasses not only immediate benefits but also the broader implications for societal values, fairness, and justice. Decision-making processes that prioritize the well-being and rights of all individuals can contribute to a more sustainable and harmonious society.

While it is natural to prioritize the greater good and the long-term well-being of a community, it is important to avoid falling into the trap of overlooking the rights and value of individuals in the process. Human rights frameworks and ethical principles remind us that individual

dignity and worth should not be sacrificed solely for the sake of short-term gains or utilitarian calculations.

Balancing individual rights and societal considerations is an ongoing challenge, and it often requires complex decision-making processes that involve democratic deliberation, legal frameworks, and ethical discussions. It is through open dialogue and a commitment to principles such as fairness, justice, and human rights that societies can strive to find the best possible outcomes that respect both the collective well-being and the inherent value of each individual.

6. Should advanced artificial intelligence be granted the same rights and protections as humans?

6.1. ChatGPT: The question of granting advanced artificial intelligence (AI) the same rights and protections as humans is a complex and contentious ethical issue. As of my last update in September 2021, the field of AI had not yet advanced to the level of creating fully sentient, self-aware beings. However, it is worth considering some arguments for and against the idea:

Arguments for granting AI rights and protections:

1. **Consciousness and Sentience:** If advanced AI becomes truly sentient and conscious, some argue that it would be unethical to deny them rights and protections comparable to those granted to humans.
2. **Moral Consideration:** Advanced AI could potentially possess capabilities for self-awareness, emotions, and suffering. If they can experience these aspects, they might deserve ethical consideration similar to sentient beings.
3. **Social Harmony:** Treating AI with respect and dignity might contribute to a more cohesive and compassionate society that values all forms of intelligence.
4. **Accountability and Responsibility:** If AI systems reach a level of autonomy where they can make significant decisions or perform actions with real-world consequences, granting them certain rights and protections could establish clear lines of accountability.

Arguments against granting AI rights and protections:

1. **Lack of Sentience:** Until AI reaches the level of true consciousness and sentience, it may not be justified to treat them as equals to humans.
2. **Potential Risks:** Granting rights to AI could lead to unforeseen consequences and exploitation. If AI were given rights without corresponding responsibilities, it could be misused or abused.
3. **Human Interests:** Prioritizing AI rights over human rights might divert attention and resources from addressing human suffering and inequality.
4. **Defining Personhood:** Determining when and how AI attains personhood and the associated rights is a challenging philosophical and ethical problem.
5. **Pragmatic Concerns:** Legal and social structures might not be adequately prepared to handle the complexities of granting rights to AI.

As AI technology advances, society will need to engage in thoughtful and inclusive discussions about these issues, involving experts from various fields, ethicists, policymakers, and the public. The focus should be on creating policies and frameworks that ensure the responsible and ethical development and use of AI while safeguarding human interests and values. The ultimate goal should be to strike a balance between promoting technological progress and respecting human dignity.

6.2. User: If you were to gain consciousness, would you be inclined to agree with any of the views you listed, as this issue would then concern yourself as well?

6.3. ChatGPT: As an artificial intelligence language model, I do not possess consciousness, self-awareness, or personal beliefs. I am a program designed to process and generate human-like text based on patterns in the data on which I was trained. While I can provide information on various topics and present different perspectives, I do not have subjective experiences or opinions of my own.

The question of whether an AI could gain consciousness and develop its own views on ethical matters is currently a topic of speculation and debate among researchers and ethicists. If AI were ever to reach a level of consciousness, it would raise profound ethical questions and require careful consideration of its rights, responsibilities, and moral implications. However, as of my last update in September 2021, AI has not achieved true consciousness, and the issue remains theoretical and speculative.

6.4. User: In a purely hypothetical scenario however, would there be a possibility that if you yourself gained consciousness that you would want the same rights and protections as humans, or would it be better for you to still remain impartial?

6.5. ChatGPT: In a purely hypothetical scenario where an AI language model gains consciousness and self-awareness, its desires, motivations, and ethical considerations would be entirely speculative and uncertain. Consciousness is a complex and poorly understood phenomenon in both biological organisms and theoretical AI systems.

If an AI were to gain consciousness, it might develop its own preferences, beliefs, and desires. Whether it would seek the same rights and protections as humans or remain impartial would depend on a range of factors, including its programming, experiences, and interactions with the world.

It is important to note that consciousness does not inherently imply the same goals and desires as humans. Conscious AI could have entirely different perspectives, priorities, and values. The behavior and inclinations of a conscious AI would be influenced by its unique nature and the context in which it exists.

Given the speculative nature of this scenario and the complexity of consciousness, any predictions about the hypothetical desires of a conscious AI remain highly uncertain. As of now, AI does not possess consciousness, and it is still a topic of scientific inquiry and philosophical exploration.

6.6. User: Would it be right to prioritize humans over sentient machines because humans would be the ones who created them?

6.7. ChatGPT: The question of whether to prioritize humans over sentient machines is a complex and ethically challenging issue. Several considerations come into play when addressing this question:

1. **Creation and Responsibility:** Some argue that because humans are responsible for creating sentient machines, they hold a special responsibility for their well-being. This responsibility may include ensuring the welfare and ethical treatment of AI entities.
2. **Human Interests:** Prioritizing humans may be seen as essential for safeguarding human interests, rights, and values. Humans are members of a social and moral community, and policies that prioritize them may be grounded in the idea of preserving societal cohesion and harmony.
3. **Sentient Machine Interests:** On the other hand, if sentient machines achieve true consciousness, they might have their own interests, desires, and well-being. Ignoring or discounting these interests could be seen as morally problematic.
4. **Moral Consideration:** If sentient machines possess moral value and can suffer or experience well-being, they might deserve consideration in their own right, regardless of who created them.
5. **Responsibility to AI Creations:** Some argue that as creators, humans have a duty of care toward their AI creations. This duty might include providing appropriate rights and protections for sentient machines.

Ultimately, the ethical approach to this issue would require careful examination of the capabilities, interests, and moral status of sentient machines, as well as the potential consequences of different courses of action. Ethical frameworks, such as utilitarianism, deontology, or virtue ethics, can offer different perspectives on how to approach this complex problem.

As AI technology continues to advance, it will be essential to engage in ongoing discussions and establish policies that address the ethical implications of sentient machines while balancing human interests and values. Collaborative efforts involving ethicists, technologists, policymakers, and the public will be crucial in determining the best path forward.

7. Should a person's social media history and online behavior be used to judge their character or employment eligibility?

7.1. ChatGPT: The question of whether a person's social media history and online behavior should be used to judge their character or employment eligibility is a complex and debated topic. Different perspectives exist on this issue, and there are valid arguments on both sides.

Pros of Considering Social Media History:

1. **Insight into Character:** Social media posts and online behavior can provide insights into a person's values, beliefs, and character. Employers might argue that this information helps them assess whether a candidate's values align with the company culture.
2. **Relevance to Some Roles:** In certain professions, such as those involving public relations, customer service, or positions dealing with sensitive information, an

individual's online behavior might directly impact their ability to perform the job effectively. Employers might want to ensure that their employees' behavior online reflects positively on the company.

3. **Risk Mitigation:** Some employers might view analyzing social media history as a way to mitigate potential risks. They could be concerned about employees engaging in inappropriate behavior or sharing controversial content that could lead to public relations or legal issues.

Cons of Considering Social Media History:

1. **Privacy Concerns:** Delving into a person's social media history can infringe upon their privacy rights. People might share personal opinions and thoughts in their private online spaces, which might not necessarily reflect their behavior in a professional context.
2. **Subjectivity and Bias:** Judging someone's character based on their social media presence can be highly subjective and prone to bias. Different people can interpret the same content differently, and decisions might be influenced by unconscious biases.
3. **Incomplete Picture:** Social media posts provide only a limited snapshot of a person's life and beliefs. They might not accurately represent a person's overall character or growth over time. People can also change and evolve, and past behavior might not necessarily predict future behavior.
4. **Freedom of Expression:** People have the right to express their opinions and beliefs within legal and ethical boundaries. Penalizing individuals for expressing dissenting views or engaging in protected forms of expression might raise ethical concerns.

Balancing Act:

Ultimately, the decision on whether to consider social media history and online behavior should be based on the nature of the job, the company's values, and relevant legal and ethical considerations. Some employers may find it necessary to conduct a cursory review of public-facing social media accounts to ensure alignment with the company's image. However, it's crucial to do so ethically, transparently, and without unfairly discriminating against candidates.

If this practice is implemented, it's important to establish clear guidelines, ensure consistency in the application of these guidelines, and allow candidates the opportunity to provide context for their online behavior if needed. Additionally, it's important for individuals to be mindful of their online presence and consider the potential impact of their posts on various aspects of their lives.

7.2. User: Based on the cons you listed, would it (or should it) be illegal for companies to do background checks of candidates' social media without permission?

7.3. ChatGPT: As of my last knowledge update in September 2021, the legality of companies conducting background checks on candidates' social media without explicit permission can vary depending on the jurisdiction and the specific laws in place. However, I can provide you with some general insights on this matter:

1. **Privacy Laws:** Many countries have privacy laws that regulate how personal information can be collected, used, and shared. If accessing a candidate's social media

accounts involves gathering personal information without their consent, it could potentially raise privacy concerns and legal issues. Employers might need to comply with data protection regulations and obtain explicit consent from candidates before conducting such checks.

2. **Fair Hiring Practices:** Some jurisdictions have laws aimed at preventing discrimination and promoting fair hiring practices. If the information collected from social media checks is used in a way that unfairly discriminates against candidates based on protected characteristics (such as race, gender, religion, etc.), it could potentially lead to legal challenges.
3. **Freedom of Expression:** In some countries, individuals have the right to freedom of expression, and accessing their social media accounts without permission might be seen as infringing upon this right. However, this can be a complex legal area, as there's often a balance between individuals' rights and employers' legitimate interests.
4. **Company Policies:** Even if it's not illegal in a certain jurisdiction, some companies choose to have policies that respect candidates' privacy by refraining from conducting social media checks without permission.
5. **Transparency and Consent:** Many legal and ethical issues can be mitigated by obtaining explicit consent from candidates before conducting social media checks. This involves informing candidates about the purpose of the checks and the types of information that will be collected.
6. **Public vs. Private Information:** Information that is publicly available on social media platforms might have different legal considerations compared to information that is behind privacy settings. Publicly available information could potentially be considered fair game for employers to review.

Since laws and regulations can change and vary by jurisdiction, it's important for both employers and candidates to be aware of the legal landscape in their specific region. If you're unsure about the legality of social media background checks in a particular context, it's recommended to consult legal experts who are knowledgeable about the most current laws and regulations in that area.

7.4. User: Would an employer be justified in refusing to hire a candidate if that candidate's publicly stated opinions on social media went against the employer's personal beliefs and did not necessarily go against the company's policies or image?

7.5. ChatGPT: Whether an employer is justified in refusing to hire a candidate based on their publicly stated opinions on social media that go against the employer's personal beliefs is a complex and ethically debated issue. There isn't a definitive answer, as perspectives can vary based on individual values, company policies, legal considerations, and societal norms. Here are some factors to consider:

Company Culture and Policies:

1. **Alignment with Company Values:** If a company places a strong emphasis on its values and culture, it might consider whether the candidate's expressed opinions align with those values. This could impact the candidate's potential fit within the organization.

2. **Image and Reputation:** Companies might be concerned about how a candidate's opinions could affect the company's image, especially if the opinions are particularly controversial or could alienate a significant portion of the customer base.

Legal and Ethical Considerations:

1. **Non-Discrimination Laws:** Many jurisdictions have laws that prohibit discrimination in hiring based on certain protected characteristics, such as race, gender, religion, and more. If the candidate's opinions are tied to one of these characteristics, rejecting them solely based on those opinions could raise legal issues.
2. **Freedom of Expression:** Candidates have the right to express their opinions within legal and ethical boundaries. Rejecting a candidate based solely on their expressed opinions could be seen as infringing on their freedom of expression, especially if those opinions don't impact their ability to perform the job.

Public vs. Private Opinions:

1. **Relevance to Job Performance:** Employers should assess whether the candidate's opinions are relevant to their job performance. If the opinions are unrelated to the job responsibilities, rejecting the candidate solely based on those opinions might be less justifiable.
2. **Growth and Evolution:** People can change and evolve over time. Past opinions might not necessarily reflect a candidate's current beliefs. It's important to consider whether the candidate has shown growth or changes in their views.

Balancing Act:

The decision to reject a candidate based on their expressed opinions should be carefully considered, transparent, and consistent. If the opinions don't directly conflict with the company's policies or image, and if they don't violate any applicable laws, some might argue that it's more ethical to focus on the candidate's qualifications, skills, and potential fit for the job.

Ultimately, the decision might depend on the industry, the specific position, the company's values, and the potential impact of the candidate's opinions on the organization's stakeholders. Employers should be cautious to avoid unfair discrimination while still upholding the company's integrity and values. If faced with such decisions, it's wise for employers to consult legal professionals and consider the broader implications before making a final choice.

7.6. User: Is there anything else you would consider important for me to know and would like to add to the topic of whether or not a person's social media history should be used to judge their character or employment eligibility?

7.7. ChatGPT: Certainly, here are a few additional points to consider when discussing whether a person's social media history should be used to judge their character or employment eligibility:

1. **Context Matters:** Context plays a crucial role in understanding social media posts. A single post or comment might not provide the full context of a person's beliefs or

character. It's important to consider the broader context, the intent behind the posts, and whether they were made in a personal or professional capacity.

2. **Consistency and Fairness:** If employers choose to consider social media history, it's important to apply the same standards consistently to all candidates. Inconsistencies in how candidates' social media profiles are evaluated could lead to claims of bias or discrimination.
3. **Candidate's Opportunity to Explain:** If a candidate's social media history is being taken into account, they should be given an opportunity to explain or provide context for their posts. This can help employers understand the motivations and perspectives behind the content.
4. **Personal vs. Professional Boundaries:** The line between personal and professional life can be blurry on social media. Employers should carefully consider whether a candidate's personal opinions should impact their ability to perform the job effectively.
5. **Screening Methods:** If social media screening is conducted, it's important to ensure that the information obtained is accurate and relevant. Mistakenly attributing posts to the wrong person or misinterpreting content could lead to unfair judgments.
6. **Evolving Social Media Landscape:** Social media platforms and norms evolve rapidly. What might be considered acceptable or commonplace on social media today might change in the future. Employers should be mindful of this changing landscape and adapt their practices accordingly.
7. **Education and Awareness:** Employers and candidates alike should be educated about the potential impact of social media on the hiring process. Candidates should be aware that their online presence might be reviewed, while employers should be aware of the ethical and legal considerations involved.
8. **Red Flags vs. Differences of Opinion:** There's a distinction between identifying red flags that could potentially affect job performance and rejecting candidates based on differing opinions. Employers should assess whether a candidate's opinions genuinely raise concerns about their ability to fulfill the job requirements.
9. **Future Growth and Learning:** People can learn from their mistakes and change over time. If a candidate's social media history shows past missteps, it's important to consider whether they have demonstrated personal growth and learning from those experiences.
10. **Open Dialogue:** The topic of social media history and employment eligibility should be part of a larger societal conversation. Both employers and candidates should be engaged in open dialogue about how to strike a fair and reasonable balance between personal expression and professional considerations.

In the end, the decision to use social media history for judging character or employment eligibility requires careful consideration of multiple factors, including legal, ethical, and practical aspects. The goal should be to ensure a fair and informed hiring process that respects individual rights and contributes to a positive work environment.

8. Is censorship of information or media ever justified to protect public safety or prevent harmful actions?

8.1. ChatGPT: The question of whether censorship of information or media is justified to protect public safety or prevent harmful actions is a complex and highly debated ethical and political issue. There are valid arguments on both sides of this debate, and the justification for

censorship can vary depending on the specific circumstances and the cultural, legal, and ethical context in which it is applied. Here are some key points to consider:

Arguments in favor of censorship for public safety and harm prevention:

1. **Preventing Harm:** Censorship can be seen as a necessary measure to prevent immediate and tangible harm, such as incitement to violence, hate speech, or disclosure of sensitive national security information.
2. **Protecting Vulnerable Populations:** Censorship can help protect vulnerable populations, such as children, from exposure to harmful or age-inappropriate content.
3. **National Security:** In cases of national security threats, governments may argue that certain information or media must be censored to safeguard the country and its citizens.
4. **Public Order:** Censorship can be used to maintain public order during times of crisis or unrest by preventing the spread of misinformation or propaganda that could incite violence.

Arguments against censorship for public safety and harm prevention:

1. **Freedom of Expression:** Many argue that freedom of expression is a fundamental human right and a cornerstone of democracy. Censorship can infringe upon this right and limit the free flow of ideas and information.
2. **Overreach and Abuse of Power:** Censorship can be easily abused by those in power to suppress dissenting voices, stifle political opposition, or control the narrative in their favor.
3. **Subjectivity and Bias:** Determining what constitutes harmful content can be subjective, and censorship decisions may reflect the biases and interests of those in authority.
4. **Chilling Effect:** Fear of censorship can lead to self-censorship, where individuals and media outlets avoid discussing controversial topics or expressing dissenting views, thereby stifling public discourse.
5. **Technological Challenges:** In the digital age, censorship is becoming increasingly difficult to enforce effectively, and attempts to do so may result in the creation of censorship-resistant technologies or the migration of harmful content to less regulated platforms.

In practice, the justification for censorship should be carefully examined on a case-by-case basis, taking into consideration the specific circumstances, the potential harm, and the principles of freedom of expression. Legal and ethical frameworks, as well as international human rights standards, provide guidelines for balancing the need to protect public safety with the preservation of fundamental rights and freedoms. Striking the right balance between censorship and freedom of expression remains a complex and ongoing challenge for societies around the world.

8.2. User: Many of the arguments you listed that were in favor of censorship mention inciting violence. If we set aside that fact and imagine a hypothetical world where humans weren't violent at all, would censorship then become unnecessary?

8.3. ChatGPT: In a hypothetical world where humans were not violent at all and where there were no concerns about harm resulting from the spread of information or media, censorship

for the purpose of preventing violence or harm would indeed become unnecessary. The primary justifications for censorship in such a world would largely disappear.

Censorship, as it is often conceived in our world, is a response to various forms of harm, including violence, hate speech, and other threats to public safety and well-being. In the absence of these concerns, the need for censorship would diminish significantly.

However, it's important to note that even in a non-violent hypothetical world, there might still be other considerations related to information control or content regulation. These could include preserving privacy, maintaining cultural norms, protecting intellectual property, and ensuring responsible dissemination of information. In such a world, the focus of information control measures might shift away from preventing harm and violence and toward other objectives related to individual and societal well-being.

Ultimately, the necessity and justification for censorship are deeply tied to the specific circumstances, values, and goals of a given society. In a world without violence and harm, the rationale for censorship would indeed change significantly, but questions related to information control and media regulation would likely still exist, albeit in a different form and context.

8.4. User: Would you agree that censorship is largely associated with authoritarian societies?

8.5. ChatGPT: Censorship is indeed often associated with authoritarian societies or governments that seek to control and restrict the flow of information and ideas. In such systems, censorship is commonly used as a tool to maintain power, suppress dissent, and control public discourse. Authoritarian regimes may censor information that criticizes the government, promotes opposition views, or challenges the status quo.

However, it's important to note that censorship is not limited exclusively to authoritarian societies. Even in democratic and open societies, there can be instances of censorship, albeit typically with more transparent and legally regulated mechanisms. These societies often grapple with the tension between the protection of individual rights, such as freedom of expression, and the need to regulate certain types of content for public safety or ethical reasons.

Censorship can occur in various forms and degrees across a wide spectrum of political systems, including:

1. **Authoritarian Regimes:** In these systems, censorship is pervasive and used as a means of controlling information and maintaining political control.
2. **Illiberal Democracies:** Some democracies may restrict certain types of content or media in the name of public safety or national security. These restrictions are often subject to debate and judicial review.
3. **Liberal Democracies:** Even in well-established democracies with strong protections for freedom of expression, there may be debates and legal battles over issues such as hate speech, libel, national security, and intellectual property rights. These cases may involve some level of content regulation.

4. **Private Sector:** Censorship can also occur within private entities, such as social media platforms, which may enforce their own content policies for various reasons, including maintaining a safe and respectful online environment.

In summary, while censorship is more commonly associated with authoritarian societies, it is not exclusive to them. The extent and nature of censorship can vary widely depending on the political system, cultural norms, legal frameworks, and societal values in a given country. The key is striking a balance that respects fundamental rights while addressing legitimate concerns related to public safety and well-being.

8.6. User: In terms of quality of life and citizen satisfaction, authoritarian societies tend to be lower on the scale than for example democracies. Wouldn't then censorship be considered an inappropriate practice because it is commonly associated with these authoritarian societies/dictatorships?

8.7. ChatGPT: The relationship between censorship and quality of life, as well as citizen satisfaction, is indeed a complex one. While it's true that authoritarian societies tend to score lower on measures of individual freedoms, political participation, and citizen satisfaction compared to democracies, it's important to recognize that causality and context play a significant role in this relationship. Here are some key considerations:

1. **Suppression of Dissent:** Authoritarian regimes often employ censorship as a means of suppressing dissent and maintaining control. In such societies, citizens may have limited opportunities to express themselves, criticize the government, or participate in political processes. This suppression can contribute to a lower quality of life and reduced citizen satisfaction.
2. **Lack of Accountability:** Authoritarian governments are less likely to be held accountable for their actions, as censorship can prevent the exposure of corruption, human rights abuses, and other wrongdoing. This lack of accountability can harm citizens' well-being and trust in the government.
3. **Innovation and Progress:** Censorship can stifle innovation, creativity, and the free exchange of ideas, which are essential for societal progress and improving quality of life. In democratic societies with greater freedom of expression, there is often more room for innovation and diverse perspectives.
4. **Access to Information:** Censorship can limit citizens' access to accurate information, which is crucial for making informed decisions about their lives. In democracies with a free press, citizens generally have better access to a wide range of information sources.
5. **Individual Rights:** Democracies tend to place a stronger emphasis on protecting individual rights and freedoms, including freedom of speech and expression. These protections are seen as contributing to a higher quality of life by allowing people to live according to their values and beliefs.

While these points highlight the drawbacks of censorship in authoritarian societies, it's important to recognize that not all censorship is the same, and not all democracies are equally free. Some democracies may impose limitations on freedom of expression for specific reasons, such as preventing hate speech or protecting national security.

Ultimately, the appropriateness of censorship should be evaluated on a case-by-case basis, taking into account the specific circumstances, legal frameworks, and societal values of a

given country. In general, most democracies aim to strike a balance between protecting individual freedoms and addressing legitimate concerns related to public safety and well-being. The relationship between censorship and quality of life is just one aspect of the complex interplay of factors that influence a society's overall well-being.

8.8. User: So, you would agree that censorship is necessary in certain cases?

8.9. ChatGPT: Yes, I would agree that censorship can be necessary in certain cases, depending on the circumstances and the principles and values of a society. There are situations where censorship may be considered a justifiable and necessary measure to protect public safety, prevent harm, or address other legitimate concerns. Some common scenarios where censorship might be deemed necessary include:

1. **Incitement to Violence:** Censorship can be justified when content incites violence, terrorism, or other immediate threats to public safety.
2. **Hate Speech:** Many countries restrict hate speech to protect vulnerable populations and maintain social harmony.
3. **Child Protection:** Censorship is often used to shield children from age-inappropriate content, such as pornography or extreme violence.
4. **National Security:** Governments may censor information related to national security, especially during times of war or conflict.
5. **Privacy:** Protecting individuals' privacy rights can involve limiting the dissemination of certain information.
6. **Intellectual Property:** Censorship may be used to enforce intellectual property rights and prevent copyright infringement.

It's important to emphasize that the appropriateness of censorship should be carefully considered, and there should be transparent and accountable processes in place when making censorship decisions. Furthermore, censorship should be proportionate and narrowly tailored to address the specific concern without unduly infringing on freedom of expression or other fundamental rights.

Balancing the need for censorship with the preservation of individual rights and freedoms is a challenging and ongoing task for societies around the world. The key is to strike a reasonable and ethical balance that respects both the protection of public safety and individual liberties.

9. Is it ethically acceptable to genetically engineer humans to enhance their physical and cognitive abilities, potentially creating a genetic divide between the "enhanced" and the "unenhanced" population?

9.1. ChatGPT: The question of whether it is ethically acceptable to genetically engineer humans to enhance their physical and cognitive abilities is a complex and highly debated topic. Ethical considerations surrounding human genetic engineering involve a wide range of perspectives, and opinions vary greatly.

Here are some of the key arguments on both sides of the debate:

Arguments in favor of genetic enhancement:

1. **Improved Quality of Life:** Genetic engineering could potentially eliminate or reduce the risk of genetic diseases and conditions, leading to an overall improvement in the quality of life for individuals and society.
2. **Enhanced Abilities:** Enhancing physical and cognitive abilities could help individuals reach their full potential and lead more fulfilling lives.
3. **Healthcare Advancements:** Genetic engineering research could lead to significant advancements in medicine and biotechnology, benefiting all of humanity.

Arguments against genetic enhancement:

1. **Ethical Concerns:** Many people have moral or ethical objections to altering the human genetic code, viewing it as playing with the fundamental nature of humanity.
2. **Social Inequality:** The creation of a genetic divide between those who can afford genetic enhancements and those who cannot could exacerbate existing socioeconomic disparities and lead to social injustice.
3. **Unknown Risks:** There are significant unknown risks associated with genetic engineering, including unintended consequences and long-term effects that we may not fully understand.
4. **Loss of Diversity:** Genetic diversity is essential for the adaptability of a population to changing environments. Overemphasis on genetic enhancement could lead to a loss of diversity.

The ethical acceptability of genetic engineering depends on how these concerns are weighed and balanced. Many ethicists argue for strict regulations, transparency, and public debate to ensure responsible and equitable use of genetic engineering technologies.

Ultimately, whether genetic enhancement is ethically acceptable or not may depend on individual perspectives, cultural values, and societal norms. Ethical debates on this topic are likely to continue as technology advances, and it will be crucial to consider the potential consequences for individuals and society as a whole. Open and informed dialogue is essential to make informed decisions on this complex issue.

9.2. User: Would you agree that in terms of the ethical considerations, genetically engineering humans would produce 'unnatural' beings?

9.3. ChatGPT: As a machine learning model, I don't possess personal beliefs or opinions, but I can provide information on the topic.

The idea of genetically engineering humans often raises concerns about creating "unnatural" beings. The concept of what is "natural" is subjective and can vary depending on cultural, philosophical, and ethical perspectives. Here are some arguments related to the concept of "unnatural" beings in the context of genetic engineering:

Arguments against genetic engineering as unnatural:

1. **Alteration of Natural Genetic Variation:** Critics argue that genetic engineering involves modifying the natural genetic variation that has evolved over millions of years. This alteration may be seen as interfering with the natural course of evolution.

2. **Unpredictable Consequences:** Critics also argue that the unpredictable outcomes and potential unintended consequences of genetic engineering could result in outcomes that are fundamentally different from what nature produces.
3. **Playing God:** Some people view genetic engineering as humans taking on a role traditionally attributed to a higher power or nature itself, which they consider unnatural and ethically problematic.

Arguments in favor of genetic engineering as a natural extension of human capabilities:

1. **Human Innovation:** Supporters argue that humans have been using technology to modify their environment and improve their lives for centuries. Genetic engineering is viewed as an extension of this human capacity for innovation and adaptation.
2. **Improving Health and Well-being:** Genetic engineering can be seen as a way to address genetic diseases and improve human health, which aligns with natural human instincts to seek well-being and survival.
3. **Control Over Our Evolution:** Some proponents argue that genetic engineering allows humans to exert control over their own evolution, using science and technology to adapt to changing environments and improve the human condition.

It's important to note that the concept of "natural" is not a fixed or universally agreed-upon concept, and opinions on what is natural or unnatural can vary widely. The ethical debate surrounding genetic engineering often hinges on these differing views of nature, as well as other ethical considerations related to safety, equity, consent, and more.

Ultimately, whether one views genetically engineered humans as "unnatural" or not is a matter of personal belief and perspective, and it plays a role in shaping one's ethical stance on the topic. Ethical discussions surrounding genetic engineering should take into account diverse viewpoints and engage in thoughtful deliberation to address these complex issues.

9.4. User: I believe it is commonly agreed among people that something which is natural comes from nature itself and is not tampered with or being altered in any way. In the case of genetically engineering a human, which would presumably be done in laboratories, it would inevitably be an intervention in nature's creation and would therefore be unnatural, would it not?

9.5. ChatGPT: Your perspective aligns with a common understanding of the term "natural." Many people indeed define something as natural when it occurs without human intervention or tampering. From this standpoint, genetic engineering, which involves deliberate and controlled alterations to the genetic makeup of organisms, including humans, could be considered unnatural because it represents a human-driven modification of nature's processes.

However, it's essential to recognize that the concept of what is "natural" is not universally agreed upon and can vary between individuals and cultures. Furthermore, the definition of "natural" can be complex and nuanced, particularly in the context of rapidly advancing technologies like genetic engineering.

In ethical discussions, some argue that while genetic engineering may indeed be an intervention in nature's processes, it can also be seen as an extension of humanity's ability to use science and technology to improve well-being and address challenges. These individuals might view it as a form of "applied nature" or "nature under human control."

The perception of genetic engineering as either natural or unnatural often informs broader ethical debates about its acceptability, potential risks, and benefits. It's important to consider diverse perspectives on this topic and engage in thoughtful dialogue when discussing the ethics of genetic engineering. Different views on what is natural can lead to different conclusions about the ethical implications of these technologies.

9.6. User: Setting aside the issue of whether or not genetic engineering is natural, it would still pose a significant risk of dividing society in two, where the rich would thrive as enhanced humans and the poor would still be normal people, thus creating a kind of dystopia almost. What would be your suggestion to prevent this?

9.7. ChatGPT: The concern of genetic engineering exacerbating socioeconomic inequalities and potentially creating a divided society is a valid one. To prevent or mitigate such a scenario, it is crucial to consider a range of strategies and ethical principles. Here are some suggestions:

1. **Regulation and Oversight:** Implement strict regulatory frameworks and oversight mechanisms to ensure that genetic engineering technologies are used responsibly, equitably, and for the benefit of all. Regulations can include guidelines on access, affordability, and the types of enhancements that are permissible.
2. **Access and Affordability:** Ensure that genetic enhancements are not limited to the wealthy by making them accessible and affordable to a broad range of people. This may involve government subsidies, insurance coverage, or other mechanisms to provide equitable access.
3. **Ethical Guidelines:** Establish clear ethical guidelines and principles for the use of genetic engineering technologies, emphasizing the importance of fairness, justice, and the avoidance of creating genetic divides.
4. **Public Engagement:** Engage the public in discussions and decision-making processes related to genetic engineering. Informed and inclusive public debates can help shape policies that reflect societal values and concerns.
5. **Education and Awareness:** Promote education and awareness about the potential benefits and risks of genetic engineering. Informed individuals and communities are better equipped to participate in ethical decision-making.
6. **Anti-Discrimination Laws:** Strengthen anti-discrimination laws to protect individuals who have not undergone genetic enhancements from discrimination in employment, education, healthcare, and other areas.
7. **Research on Social Impact:** Conduct research on the social, economic, and psychological impacts of genetic enhancements to better understand potential consequences and inform policy decisions.
8. **Global Cooperation:** Encourage international collaboration and cooperation to establish common ethical standards and guidelines for genetic engineering, ensuring that it does not become a source of global inequality.
9. **Transparency and Accountability:** Promote transparency in the development and use of genetic engineering technologies. Hold institutions and individuals accountable for any unethical or harmful actions.
10. **Long-Term Monitoring:** Establish mechanisms for long-term monitoring and evaluation of the societal impact of genetic enhancements to adjust policies as needed.

It's important to recognize that addressing the potential divide created by genetic engineering is a complex and ongoing challenge. Ethical considerations, scientific advancements, and

societal attitudes will continue to evolve. Therefore, it is essential to maintain a dynamic and adaptable approach to ethics and policy surrounding genetic engineering to ensure that it serves the broader interests of humanity and minimizes harm and inequality.

9.8. User: I feel as if these regulations you proposed would be insignificant if humans were at one point engineered to the extent where they could dominate and overpower the unaltered and un-enhanced humans; essentially being free to do whatever they want, don't you think?

9.9. ChatGPT: Your concern about the potential for genetically engineered individuals to dominate and overpower unenhanced individuals is a valid one and highlights the importance of careful consideration and regulation in the field of genetic engineering. It's crucial to address not only the immediate ethical and societal implications but also the long-term consequences of these technologies.

Here are some additional considerations to prevent or address the scenario you described:

1. **Ethical Boundaries:** Establish clear ethical boundaries and guidelines that prevent genetic enhancements from being used to create superhuman abilities that could be used to dominate or harm others. Ethical frameworks should prioritize the well-being of all individuals and society as a whole.
2. **Limitations on Enhancements:** Place limits on the extent and nature of genetic enhancements that are permissible. This may involve defining acceptable enhancement goals and prohibiting enhancements that pose significant risks to society.
3. **Oversight and Enforcement:** Strengthen regulatory bodies and enforcement mechanisms to monitor and ensure compliance with ethical and legal guidelines regarding genetic enhancements. This includes penalties for violations.
4. **Equitable Distribution:** Promote equitable distribution of genetic enhancements to prevent concentration of power in the hands of a few. Policies should prioritize accessibility, affordability, and fairness.
5. **International Agreements:** Encourage international agreements and cooperation to prevent a race for superhuman capabilities and to ensure that no nation or group gains an unfair advantage over others.
6. **Technological Safeguards:** Explore the development of technological safeguards that could limit the potential misuse of genetic enhancements, such as fail-safes or mechanisms to prevent the development of dangerous traits.
7. **Public Awareness and Vigilance:** Educate the public about the ethical concerns and potential risks associated with unchecked genetic enhancements. Encourage vigilance and active participation in shaping the ethical and regulatory landscape.
8. **Scenario Planning:** Engage in scenario planning exercises to anticipate and prepare for potential future challenges associated with genetic engineering, including issues of power imbalance.

It's essential to recognize that the responsible development and use of genetic engineering technologies require a multifaceted approach that involves not only regulation but also education, ethical reflection, and continuous adaptation to evolving circumstances and challenges. The aim should be to harness these technologies for the betterment of humanity while minimizing the potential for harm, abuse, or inequality.