Outline & Abstract on Right to Content Moderation

Salvin Chowdhury

March 26, 2025

1 Requirements of the Outline & Abstract

The purpose of this paper is to craft the outline and abstract with regards to the right to content moderation. We look over the outline requirements and the structure of the paper, and then develop an argument for and against the thesis statement.

1.1 Outline Requirements

The requirements of the paper are to have a title, an abstract that culminates in a thesis statement, a rough outline, and a list of three properly formatted sources as listed below:

- Reading from the class
- Entry from the Stanford Encyclopedia of Philosophy
- A source of choosing

1.2 What a Good Philosophy Paper Should Look Like

A good philosophy paper should present a clear thesis and argue for it. The argument needs to be persuasive and presented in a well-ordered, logical fashion. We then consider the strongest possible objections to the view and offer rebuttals. Basic fallacies should be avoided, and appropriate quotations should be used to clarify points.

1.3 Structure of Paper

The paper should have the following structure:

- An introduction and conclusion
- A body that presents the arguments/evidence for the thesis
- One or two strong arguments against the thesis

1.4 Sources of Research

Avoid using Wikipedia, YouTube, and any random website. Examples of good sources are peer-reviewed sources, the Stanford Encyclopedia of Philosophy, Google Scholar, JSTOR, Philosopher's Index, and books.

1.5 Abstract Structure

The first paragraph should be focused heavily on. Avoid using general introduction statements. Be sure to set a brief introduction by letting the reader know the topic and why the topic is relevant. The reader should be given a play-by-play account of how the essay is going to unfold and include a thesis statement.

1.6 Assumptions of the Reader

Assume that the reader is a jerk and overly hostile. The job is to convince the reader while respecting their intelligence. Always give the opposition the best reading and don't assume the reader will do the same. Be sure to explain the concepts fully, and ensure that anyone should be able to read the paper and understand it.

1.7 Crafting the Thesis Statement

The thesis statement should come as the last sentence of the first paragraph. The thesis statement should be true or false.

2 Planning Out The Paper

In this section, we plan out the arguments regarding the right to moderate content. In this draft, we try to figure out the title, the abstract, as well as the kind of sources that can be used to make the arguments.

2.1 Deciding the Name of the Title

When it comes to deciding the name of the title, we can figure this out by deciding what we want to talk about. With regards to content moderation, we would like to look at who should be able to moderate content online, as well as what ethical theories should be used to determine which content can be removed or not. With such ideas in mind, a fair title would be:

"Who should decide the rules for content moderation and should it be automated using Artificial Intelligence?"

2.2 The Parties Involved & Ethical Theories

To bring the title to life, we introduce ethical theories and parties that should be concerned with the rules for content moderation, as well as which party should be in charge of moderating content. We support our ideas with research as we do a deep dive into the world of content moderation.

2.3 The Parties Involved

Over the course of the modern internet, we have seen different attempts at moderating content on the internet. We attempt to take a look at the collective history of how content moderation has happened on the internet, as it where we can deduce the parties that can be involved.

2.3.1 A Rape in Cyberspace (1993)

LambdaMoo is a multi-user dimension program that is open to the public. The program allows users to create and design the interaction space and context. A character, named Bungle, took control of two other characters, Legba and Starspinner, and perform sadistic actions on them. After the incident, the users gathered and made four arguments:

- The technolibertarians argued that rape in cyberspace was a technical inevitability and that a solution would be to use defensive software tools
- The legalists argued that Bungle could not be "toaded" since the MOO had no explicit rules at all, they proposed the establishment of rules and virtual institutions to exercise the control required
- Another group of users believed that only the programmers, or wizards have the power to implement rules
- The anarchists wanted to see the matter resolved without establishment of social control

The LambdaMoo situation provides a insight into the arguments that may be presented when it comes to moderating the harmful behavior in the virtual space. This is because, similar to unpleasant interactions on the internet, what Bungle did may seem like it is unfair to real rape victims to equate what happens to them with the experiences of LambdaMOO pariticipants who witnessed a representational rape.

The counter argument here is that it seems appropriate that to say both that Bungle did something wrong, and the real person controlling Bungle did something wrong. Both engaged in a form of violent sexual behavior. Here are some of the key takeaways from this:

- Virtual behavior can have real psychological and physical consequences. The controllers of virtual characters have responsibilities for those consequences.
- The controller of virtual characters are generally unknown to the flesh controllers of other virtual characters, and this affects the nature of responsbility in virtual environments. This makes it important to have rules specifying behaviors that are and are not allowed in a virtual environments
- The developers and controllers of a virtual environment interface have responsibilites to the flesh controllers of other virtual characters.
- If justice is to be a goal in a computer-mediated environment, then rules should be explicit. When freedom becomes license, harm is more likely to occur

2.3.2 Seacrhing for a Leviathan in Usenet (1992)

Usenet is the largest computer conferencing network in the world. It's users can send private messages to one another via electronic mail. Despite it's large size, Usenet has no central authority which monitors access or control. All control id exercised at the site level. Sites determine whether to provide access to users or whether they want to provide a "feed" or connection to a potential site.

Usenet has a two-dimensional nature, its creation of an explicit language to describe its "physical" reality, its interference in the transfer of the social strucutre from the external world, and its ability to compensate for the lack of a complete social structure by developing a parallel structure to that of the external world.

Due to this nature, Usenet users face the following difficulties

- They are unable to bring their real-life social behaviors and structures into the online environment because written communication doesn't capture everything about how people interact in the real world.
- They suffer from a deprivation of "subtleties". Users limited to written communication are denied the full range of verbal and non-verbal cues customary to interpersonal communication.

To deal with this, Usenet has a parallel method for conveying norms and traditions which is known as "netiquette". The term implies that it is "network etiquette" and helps to reinforce the standards of behavior that users might miss from the lack of non-verbal cues.

When it comes to users on Usenet, they can be seen to have a personae. Personae, lacking physical form, doesn't require physical sustenance, but is dependent on three essential conditions:

- **Condition One:** it is the continued association between the user and the persona. The loss of the user's access to Usenet severs the association to their persona.
- **Condition Two:** a visible demonstration of presence. While Usenet has great utility to a passive user, the lack of interaction with other users doesn't create a persona
- **Condition Three:** the pariticipation is continuous. A persona belonging to a user who is unwilling to continue to pariticipate will continue to exist until the memory of that existence is forgotten by the other users.

Given the paralle between personae and Hobbes' "persons", it is possible to establish a parallel between Leviathan and Usenet. On the subject of power, Hobbes says:

> **"Natural power is the eminence of the faculties of body, or mind; as extraordinary strength, form, prudence, arts, eloquence, liberality, nobility. Instrumental are those powers, which acquired by these, or by fortune, are means and instruments to acquire more."**

According to Hobbes, people have different kinds of power, like physical strength, intelligence and skills. But, on Usenet, where people interact through text, physical strength and appearance don't matter. "Strength" is Usenet can refer to how well someone can write aggressive messages in discussions. As a result, a person's influence comes from their ability to argue effectively.

2.3.3 Content Moderation using AI

Automating content may often be justified due to the sheer size of content on the internet. It is desirable to use AI due to the immense amount of data , the relentlessness of violations and need to make judgments without human moderators making them.

However, the problem is that the AI tools simply compare new posts to previously flagged content. This means they're great at catching duplicates but not great at identifying new forms of harmful speech. Such AI tools also struggle with understanding context, sarcasm and cultural meanings. AI systems also make mistakes that can unfairly impact marginalized groups.

As such systems work based on patterns in large amounts of data, they unintentionally reinforce existing biases. For example, they may wrongly flag certain communities more often while failing to protect them from real harm.

2.3.4 Fighting Hate Speech, Silencing Drag Queens?

This article discusses about how AI tools used for moderating content don't fully understand the context, and this may negatively impact the LGBTQ community. For example, some LGBTQ people, especially drag queens, use words that AI systems might label as toxic, however they may be actually playful or empowering. However, because such words are used as insults in general conversations, AI tools may assume them as being always offensive.

2.3.5 Government & Content Moderation

The CEO of Cloudflare, Matthew Prince, suggested that the government should be more involved in deciding what speech is allowed online. But the article argues that it is a bad idea because private companies have been in charge of regulating speech on their platforms. If the government had more power over online content, it could threaten free speech and independence of tech companies.

Tech companies are working on handling harmful content. Even though their efforts aren't perfect, they are better suited for the job. If the government started controlling what content is allowed, it could become a political tool to suppress speech that official don't like. This would be dangerous as it could lead to censorship based on political agendas rather than fairness.

3 Outline Drafts

In this section, we attempt to build an abstract in the form of two different drafts. We make sure to set a brief introduction by leeting the reader know the topic and why the topic is relevant.

3.1 Draft One: The Right to Content Moderation

In 1991, a user named ”Bungle” had logged into LambdaMOO, a multi-user dimension program that allows users to create their own characters and interact with others using just text. Bungle had designed a subprogram called ”Voodoo Doll” which allowed the user to take control of two other user’s characters, who were ”Legba” and ”Starspinner”. Bungle took control of both characters and performed sadistic actions on them, such as having them eat pubic hair and have intercourse with one of the victims. It wasn’t over until when another character had used a subprogram to freeze Bungle’s actions. Such an action had caused an intense stir of anger, enormous enough for all users to get together and discusss Bungle’s fate. Some had argued that there needs to be better software protections, some argued that he couldn’t be removed because there were no explicit rules to begin with and others believed that it was up to the moderators to make the decision. While there was a intense debate, the moderators eventually, under immense pressure, decided to remove Bungle. Such a story in the 90’s is a mirror to today’s environment on the Internet. As the debate to who has the right to remove and moderate content rages on, the different societal views on what may seem morally wrong and right greatly differ at the same time. This makes it a great challenge for those in charge, whether it is private companies, governmental institutions or even artificial intelligence, on whether they should have the right to hold the reins of controlling the free speech on the internet.

Problems: The abstract sounds more of an introduction than a abstract. Focus on making a summary of the paper.

3.2 Draft Two: The Right to Content Moderation

In the 90’s, LambdaMOO, a mutli-user dimension (MUD) program, had experienced from one of the earliest cases of ”virtual rape”. Such an incident had shaken the community, prompting discussions as to what the correct action should be and who should execute it. Such a case is a mirror to today’s environment on the Internet. As the debate to who has the right to remove and moderate content rages on, the different societal views on what may seem morally wrong and right greatly differ at the same time. This makes it a great challenge for those in charge, whether it is private companies, governmental institutions or even artificial intelligence, on whether they should have the right to hold the reins of controlling the free speech that goes on to grow at a exponential rate on the internet. This paper explores cases similar to LambdaMOO and decisions made by different groups on free speech. We argue that it is private companies, who should hold the reins of controlling the free speech on the internet due to a lack of political bias, however we argue that it is better to have humans moderating content rather than automating it using Artificial Intelligence (AI).

Problems: The abstract needs a better and refined flow from transitioning from the past to the present.

References