

Content Moderation Policy

Task

Your task is to create the content moderation policy of a fictive digital platform of your choice. The content moderation policy will represent 70% of your group grade for this task.

Part 1 (Fictive platform): You will first need to outline the a) mission, b) philosophy, and c) functioning of your platform in *around 500 words*.

Part 2 (Main part): For the main task, you will need to write the content moderation policy (also called community guidelines) of your platform. The policy must outline the types of content that are allowed and not allowed on your platform (e.g., spam and adult content). Your content moderation policy should be both comprehensive and in line with the mission, philosophy, and functioning of your platform (see part 1). This section should consist of *around 2000 words*.

Part 3 (Reflection): In the final part, you are required to critically evaluate the advantages and disadvantages of your content moderation policy. You will also need to spell out the considerations that have influenced your decisions (use examples). Finally, you will need to critically assess potential practical issues when implementing your content moderation policy (e.g., costs, edge cases, automation...). Importantly, you should assume that your content is aimed at ordinary users. You are *not* expected to include any legal terminology ("legalese"). This section should consist of *around 500 words*.

The total length of the assignment (three parts together) should not exceed 3000 words!

Video: You will need to create a video of 10 minutes maximum where each group member speaks for at least for one minute. The video should first briefly summarize your fictive platform as well as your content moderation policy. The main part of the video should pertain to a) your critical assessment of your considerations and challenges when drafting the policy and b) outline how the policy connects to the contents of the course. Beside the quality of your content, we will also grade the quality of the delivery. The video will represent 30% of your group grade for this task.

Reference Content Moderation Policies

For your task, you can use the following existing content moderation policies as a *reference*:

1. Pinterest: <https://policy.pinterest.com/en/community-guidelines>
2. Quora: <https://help.quora.com/hc/en-us/articles/360000470706-Platform-Policies>
3. Vimeo: <https://vimeo.com/legal/policies/community-guidelines>
4. Medium: <https://policy.medium.com/medium-rules-30e5502c4eb4>

Groups

You are expected to work in *groups of 4 to 6 people*.

Use of AI Tools

The use of AI tools is permitted for the following purposes:

- Brainstorming ideas for your fictive platform
- Getting feedback on the format and content of *your own work*

Using AI tools is *explicitly not permitted* to write the entire or parts of the task! You are also required to declare the use of AI tools in accordance with ETH's recommended declaration of use:

https://library.ethz.ch/en/researching-and-publishing/scientific-writing-at-eth-zurich/plagiat-und-kuen_stliche-intelligenz-ki.html

Performance Evaluation

Your content moderation policy will represent 70% of your group grade for this task and will be evaluated based on the following criteria:

Criteria (Weight)	Levels of Performance				
	Not Sufficient	Sufficient	Good	Great	Excellent
Completeness (30%)	Major relevant topics are missing.	Most key topics are covered, but some obvious areas are omitted or only superficially addressed.	All essential topics are covered adequately.	All essential topics are covered in detail, showing consideration for edge cases.	All essential topics are comprehensively covered with nuance, foresight, and clear definitions of scope.
Content Quality (30%)	Moderation choices are unclear, illogical, or ineffective.	Moderation choices are understandable but may lack clear justification or practicality.	Moderation choices are generally sound and justified, but may have minor inconsistencies or oversights.	Moderation choices are sound, well-justified, and practical for the platform's context. The Policy entails some creative solutions.	Moderation choices are exceptionally sound, clearly justified, and anticipate potential issues. The Policy entails particularly creative solutions.
Consistency (30%)	Terms are used inconsistently; clauses are incoherent/contradicting; the Policy contradicts the platform goals.	Some inconsistencies in terminology; clauses sometimes contradict; alignment with platform goals partially exist.	Policy clauses are mostly consistent; alignment with platform goals is generally clear but could be stronger.	Policy clauses are consistent; clear alignment between the policy details and overall platform goals.	Policy clauses are precise and terms are used consistently throughout; strong, explicit alignment with platform goals.
Formal Requirements (10%)	Numerous errors in language, formatting, or structure.	Some errors in language/formatting; structure may be confusing.	Minor errors in language/formatting; structure is mostly clear.	Very few errors in language/formatting; structure is clear and logical.	Polished language, correct formatting, sound tone for target audience, and clear, logical structure.