Design Thinking Project Workbook

Don't find customers for your product but find products for your customers

1. Team

Team Name: “Comment Vision”

Team Members:

1. Samanvitha, Creating the Code of Project, 2320030437
2. Annapurna, interaction with clients and research of project in social platforms, 2320090064

2. Problem/Opportunity Domain

Domain of Interest: Comment Analyzer using AI

Description of the Domain:

The Comment Analyzer Project falls under the domain of Natural Language Processing (NLP) and Sentiment Analysis, a subfield of Artificial Intelligence (AI) that focuses on enabling computers to understand, interpret, and respond to human language.
In this domain, textual data is processed to extract meaningful insights, such as sentiment, emotions, toxicity, or spam detection. The Comment Analyzer leverages machine learning models to analyse user-generated comments, classify their sentiment (positive, negative, or neutral), and detect inappropriate or harmful content.

Why did you choose this domain?:

1. **Real-World Impact –** NLP and Sentiment Analysis help in moderating online content, improving user experience, and ensuring safe digital interactions.
2. **Hands-on AI Experience –** This project provides practical exposure to text processing, classification algorithms, and ethical AI practices.
3. **Growing Industry Demand –** NLP is widely used in chatbots, social media monitoring, and customer feedback analysis, making it a valuable skill for future career opportunities.

2.Problem/Opportunity Statement

Problem Statement: Users struggle with efficiently analysing and moderating online comments due to the high volume of interactions. A system is needed to automatically categorize, filter, and analyse comments to improve content moderation and user engagement.

Problem Description:
In online platforms, user-generated comments often contain spam, hate speech, or inappropriate content, making moderation a challenging task. Manually filtering such comments is time-consuming and inefficient. This project aims to develop an AI-based Comment Analyzer that uses Natural Language Processing (NLP) to classify and filter comments based on sentiment, toxicity, and relevance. It will help in automating content moderation, ensuring a safer and more engaging user experience.

Context (When does the problem occur):

1. When online platforms receive a high volume of user-generated comments, making manual moderation difficult.
2. When toxic, spam, or inappropriate comments spread quickly, affecting user experience and engagement.

Alternatives (What does the customer do to fix the problem):

1. Use manual moderation by reviewing and filtering comments individually.
2. Rely on basic keyword-based filtering, which may not accurately detect nuanced toxicity or spam.

Customers (Who has the problem most often):

1. Social media platform administrators and content moderators.
2. Website owners and forum managers.
3. Organizations managing online communities and public discussion platforms.

Emotional Impact (How does the customer feel):

1. Overwhelmed by the large volume of comments needing moderation.
2. Stressed about failing to remove harmful content quickly enough.
3. Frustrated by the inefficiency of manual moderation and inaccurate filtering methods.

Quantifiable Impact (What is the measurable impact):

- Increased moderation time: Manual moderation is time-consuming, reducing efficiency in handling large volumes of comments.
- Higher exposure to harmful content: Delay in detecting and removing toxic comments can lead to negative user experiences and increased platform disengagement.
- Reduced user retention and trust: Users may leave or avoid engaging if they frequently encounter offensive or inappropriate content.
- Increased moderation costs: Hiring more human moderators or relying on inefficient tools increases operational costs.

Alternative Shortcomings (What are the disadvantages of the alternatives):

1. **Manual Moderation: Time-consuming, prone to human bias, difficult to scale.**
2. **Keyword-Based Filtering: Inaccurate, fails to detect context-based toxicity, high false positives/negatives.**
3. **Generic AI Models: May lack domain-specific accuracy, struggle with slang and evolving language, potential privacy concerns.**
4. **Community Reporting: Delayed response time, reliance on users, risk of mass reporting abuse.**

Any Video or Images to showcase the problem:

3. Addressing SDGs

Relevant Sustainable Development Goals (SDGs):

SDG 8 (Decent Work & Economic Growth):

- AI-driven comment analysis improves workplace efficiency by automating moderation, reducing manual workload.
- Reducing exposure to harmful content lowers stress and burnout for moderators and users.

SDG 9 (Industry, Innovation, and Infrastructure):

- AI-powered comment moderation enhances digital platforms by fostering safer and more engaging online discussions.
- Encourages innovation in content moderation, leveraging NLP and machine learning to improve accuracy and efficiency.

SDG 16 (Peace, Justice, and Strong Institutions):

- Helps create a safer digital environment by identifying and filtering toxic, harmful, or misleading comments.
- Supports fair and unbiased moderation, reducing the spread of misinformation and online harassment.

How does your problem/opportunity address these SDGs?:

- Enhances digital safety: AI-powered moderation reduces cyberbullying, misinformation, and harmful content.
- Promotes fair and unbiased content moderation: By leveraging AI, platforms can ensure equal enforcement of community guidelines.
- Encourages innovation in online communication: AI-driven solutions improve the efficiency of comment moderation, making digital interactions more positive and constructive.

Stakeholders

1. **Who are the key stakeholders involved in or affected by this project?**

The key stakeholders include users, content moderators, developers, businesses, investors, and security experts. Users and moderators interact with the AI, while developers maintain it. Businesses implement the tool, investors fund it, and security experts ensure compliance**.**

2. **What roles do the stakeholders play in the success of the innovation?**

Users drive adoption, moderators improve content quality, and developers enhance AI accuracy. Businesses integrate the tool, investors provide funding, and security experts ensure privacy and fairness. Each stakeholder contributes to the system's effectiveness and reliability.

3. **What are the main interests and concerns of each stakeholder?**

Users want fair moderation, and moderators need an efficient tool. Developers focus on AI accuracy, while businesses seek better engagement and compliance. Investors expect returns, and security experts prioritize privacy and ethical AI use.

4. **How much influence does each stakeholder have on the outcome of the project?**

Users, moderators, and businesses have the most influence through adoption and feedback. Developers and security experts shape AI performance and compliance. Investors impact the project's scale through financial support.

5. **What is the level of engagement or support expected from each stakeholder?**

Users and moderators provide feedback, and developers refine the AI. Businesses integrate and promote the tool, while investors fund its growth. Security experts oversee compliance and ethical use.

6. How will you communicate and collaborate with stakeholders throughout the project?

Feedback channels will be used for users and moderators. Developers, businesses, and investors will have regular updates and reports. Security experts will be involved in audits and compliance checks.

5. Power Interest Matrix of Stakeholders

High Power, High Interest:

- **Businesses & Social Media Platforms:** They have the power to implement the system and a strong interest in improving content moderation and user experience.
- **Developers/AI Engineers:** They shape the AI model and have a high interest in ensuring its accuracy and ethical implementation.

High Power, Low Interest:

- **Investors/Funders:** They provide financial support but may not be directly involved in system improvements.
- **Regulatory Bodies (Data Privacy & AI Ethics):** They enforce compliance and ethical AI practices but have limited interest in daily operations.

Low Power, High Interest:

- **Users & Content Creators:** They rely on fair moderation and have a high interest in the AI's accuracy but limited control over its development**.**
- **Content Moderators:** They use the tool to filter harmful comments and have a vested interest in its efficiency and reliability.

Low Power, Low Interest:

- **General Public:** They are indirectly affected but may not engage unless they use the platform regularly.

6. Empathetic Interviews

Conduct Skilled interview with at least 30 citizens/Users by asking open ended questions (What, why/How etc) and list the insights as per the format below

I need to know (thoughts, feelings, actions)	Questions I will ask (open-ended)	Insights I hope to gain
Thoughts	What are your key expectations from an AI-powered Comment Analyzer?	What are their primary goals for using the tool?
	How important is accuracy in detecting sentiment and harmful content?	What performance metrics (e.g., precision, recall) define success for them?
	How do you see this tool improving online discussions?	What key challenges do they face with current moderation methods?

I need to know (thoughts, feelings, actions)	Questions I will ask (open-ended)	Insights I hope to gain
	What features would make this tool more useful in your daily work?	What functionalities are most valuable to users?
Feelings	How do you feel about AI moderating and analying comments?	Do they trust AI-driven moderation? Any biases or concerns?
	What concerns do you have about false positives or negatives in comment analysis?	What risks do they foresee, and how can they be mitigated?
	How transparent should the AI model be in decision-making?	Do they prefer explanations for flagged comments?
	What would make you fully trust and adopt this AI tool?	What factors drive their confidence in AI-powered comment analysis?
Actions	How would you like to be involved in testing the tool?	What role do they want in validating and improving the system?
	What types of comments should the AI focus on (toxic, spam, hate speech, etc.)?	Which areas of comment analysis are most critical for them?
	What datasets or real-world examples would improve the model?	How can user-generated data improve AI accuracy?
	What concerns, if any, do you have regarding AI's impact on free speech?	How do users balance moderation with maintaining open discussions?

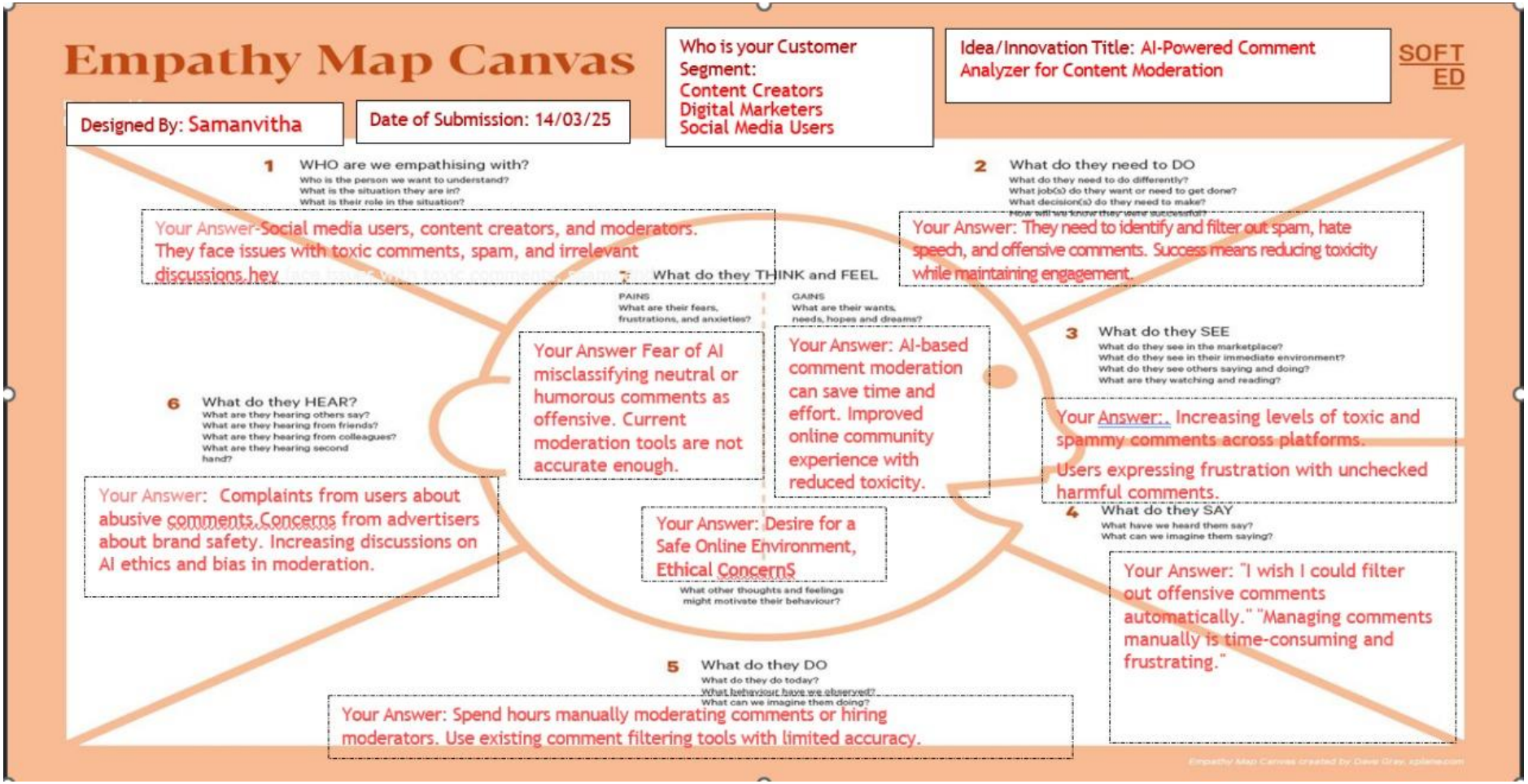
SKILLED INTERVIEW REPORT

(Examples are given. Erase them and fill with your user information.)

User/Interviewee	Questions Asked	Insights gained (NOT THEIR ANSWERS)
Rahul, Student	How often do you encounter harmful or toxic comments online?	Many users face frequent exposure to negativity, impacting their online experience.

Priya, Content Creator	What challenges do you face in moderating comments on your posts?	Manual moderation is time-consuming, and automated filters often miss context.
Arjun, Social Media Manager	What features would make an AI-powered comment analyser more useful for you?	Users want customizable filters, real-time analysis, and explainability in moderation decisions.
Sneha, Teacher	Do you think AI moderation tools affect free speech?	Some believe AI moderation risks censorship, while others see it as necessary for healthy discussions.
Vikram, Developer	What concerns do you have about AI-based comment analysis?	Accuracy and bias in sentiment detection are major concerns, requiring constant improvement.

EMPATHY MAP:


Empathy Map Canvas
Who is your Customer Segment:
Content Creators
Digital Marketers
Social Media Users
Idea/Innovation Title: AI-Powered Comment Analyzer for Content Moderation
SOFT ED
Designed By: Samanvitha
Date of Submission: 14/03/25
1 WHO are we empathising with?
Who is the person we want to understand?
What is the situation they are in?
What is their role in the situation?
Your Answer-Social media users, content creators, and moderators. They face issues with toxic comments, spam, and irrelevant discussions.hey
2 What do they need to DO
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
Your Answer: They need to identify and filter out spam, hate speech, and offensive comments. Success means reducing toxicity while maintaining engagement.
7 What do they THINK and FEEL
PAINS
What are their fears, frustrations, and anxieties?
Your Answer Fear of AI misclassifying neutral or humorous comments as offensive. Current moderation tools are not accurate enough.
GAINS
What are their wants, needs, hopes and dreams?
Your Answer: AI-based comment moderation can save time and effort. Improved online community experience with reduced toxicity.
3 What do they SEE
What do they see in the marketplace?
What do they see in their immediate environment?
What do they see others saying and doing?
What are they watching and reading?
Your Answer:. Increasing levels of toxic and spammy comments across platforms.
Users expressing frustration with unchecked harmful comments.
6 What do they HEAR?
What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second hand?
Your Answer: Complaints from users about abusive comments.Concerns from advertisers about brand safety. Increasing discussions on AI ethics and bias in moderation.
Your Answer: Desire for a Safe Online Environment, Ethical ConcernS
What other thoughts and feelings might motivate their behaviour?
4 What do they SAY
What have we heard them say?
What can we imagine them saying?
Your Answer: "I wish I could filter out offensive comments automatically." "Managing comments manually is time-consuming and frustrating."
5 What do they DO
What do they do today?
What behaviour have we observed?
What can we imagine them doing?
Your Answer: Spend hours manually moderating comments or hiring moderators. Use existing comment filtering tools with limited accuracy.
Empathy Map Canvas created by Dave Gray, xplane.com

3. Empathy Map

- **1. Who is your customer?**
- **Description:**
- Social media users, content creators, and platform moderators.
- People who want to analyse comments for sentiment classification.
- Ensure content safety and promote a healthy environment.
- Key Points:
- Improve engagement and filter out negative or offensive comments.
- Help platform moderators automate comment analysis.

- **2. Who are we empathizing with?**
- Description:
- People who experience online abuse, negative feedback, or offensive comments.
- Platform moderators who are responsible for content regulation.
- Key Points:
- Goal is to create a safe and positive online environment.
- Reduce manual effort in content moderation.

8. Persona of Stakeholders

Stakeholder Name: MUNNY

Demographics:

- Age: 20 years
- Location: Urban area
- Role: Social media content creator / Platform moderator
- Dependency: Heavily relies on analysing comments to ensure healthy engagement on their platform.

Goals:

- Filter out negative and offensive comments to promote a safe and engaging environment.
- Increase positive interactions by moderating comments efficiently.
- Build trust with their audience by ensuring respectful and healthy discussions.

Challenges:

- Managing large volumes of comments with mixed sentiments.
- Filtering out abusive, hateful, or toxic comments without affecting genuine feedback.
- Maintaining engagement and audience trust despite offensive comments.

Aspirations:

- To create a safe and positive online community.
- To ensure maximum audience engagement by filtering out harmful content.
- To reduce manual moderation efforts using AI-driven comment analysis.

Needs:

- Real-time sentiment analysis of comments.
- Automated filtering of offensive or spam comments.
- Actionable insights to increase positive engagement.

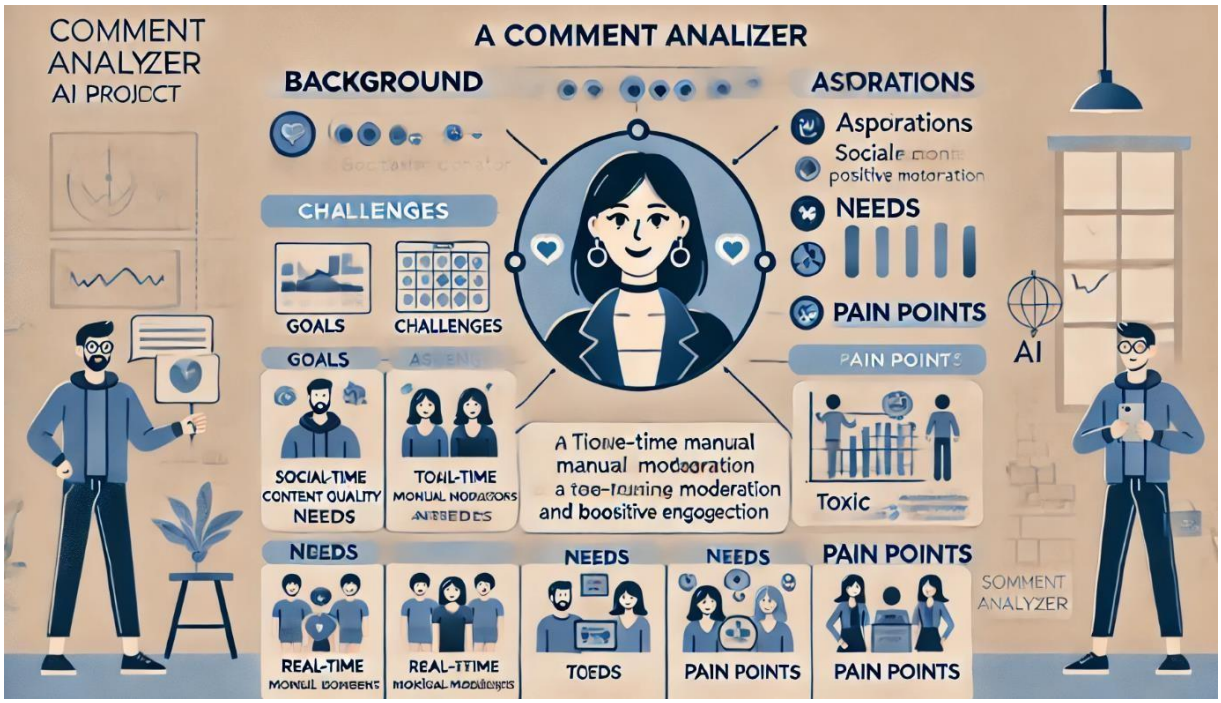
Pain Points:

- Losing audience engagement due to toxic comments.
- Spending excessive time on manual moderation.
- Inconsistent sentiment analysis affecting content strategy.

Storytelling:

Sravani works hard to maintain a positive online community. However, she often struggles with managing toxic or offensive comments, which affect the overall audience engagement. By using the Comment Analyzer AI, she can now automatically filter out negative and abusive comments, ensuring a healthier and more interactive online space. This also helps her maintain audience trust and boosts content performance.

Sample:



10. Look for Common Themes, Behaviours, Needs, and Pain Points among the Users

Common Themes:

- AI-Powered Moderation: Users want automated solutions to manage and analyze comments efficiently.
- User Safety & Well-being: There is a strong need for reducing toxicity and harassment online.
- Sentiment Analysis for Engagement: Users are interested in understanding audience reactions beyond just likes and shares.

Common Behaviors:

- Manually filtering and deleting harmful comments.
- Reporting abusive content but often facing slow responses from platforms.
- Using keyword-based filters, which sometimes block relevant comments.

.

Common Needs:

- A real-time AI comment analyzer that filters spam, hate speech, and offensive language.
- Customizable moderation settings to align with individual or brand preferences.
- Insights on audience sentiment to improve content strategy.

Common Pain Points:

- High volume of comments makes manual moderation overwhelming.

- False positives & negatives: AI misclassifying harmless comments as offensive and vice versa.
- Lack of control: Current tools offer little flexibility in moderation rules.

12. Define Needs and Insights of Your Users

User Needs:

- Accurate and Timely Comment Analysis – Users need real-time and precise analysis of comments to maintain a positive and respectful community.
- Toxic Comment Detection – Users expect the AI tool to quickly identify and flag offensive, abusive, or harmful comments.
- User-Friendly Dashboard – Users need a simple and clear interface to view flagged comments, sentiment analysis, and overall comment health.

User Insights:

- Content Moderation is a Priority – Users highly prioritize a clean and positive comment section to maintain brand image and audience satisfaction.
- Concern About Misclassification – Users fear that some positive comments may get flagged or toxic comments may be missed, leading to a lack of trust in the AI tool.
- High Demand for Real-Time Moderation – Users want instant notifications and automated filtering for toxic comments to reduce manual moderation efforts.

13. POV Statements

POV Statements:

PoV Statements **(At least ten)**	**Role-based or Situation-Based**	**Benefit, Way to Benefit,** **Job TBD,** **Need (more/less)**	**PoV Questions** **(At least one per statement)**
A social media manager who handles brand reputation needs a tool to identify and remove toxic comments because negative comments can harm the brand image.	A social media manager handling brand reputation	Protect the brand's image by eliminating harmful content.	How might we ensure real-time detection and removal of toxic comments to protect brand image?
A content creator who regularly uploads videos needs an AI-powered comment analyser because toxic comments can affect audience engagement and mental health.	A content creator uploading content	Maintain a positive and engaging comment section.	How might we help content creators create a safe and engaging comment space?
An online gaming community manager who moderates chats needs an AI tool to detect hate speech because toxic conversations can negatively impact the gaming experience.	A gaming community moderator	Ensure a healthy and positive gaming environment.	How might we automatically filter out hate speech from gaming chatrooms?
A YouTube video creator who receives thousands of comments daily needs an AI tool to automatically moderate toxic comments	A YouTube content creator	Reduce manual effort by automating the comment analysis process.	How might we provide content creators with a smarter and faster comment moderation system?
because manual moderation is time-consuming.			

A business page owner who wants to promote products needs a tool to filter harmful or misleading comments because negative feedback can affect product sales.	Business page owner	Improve customer engagement by eliminating harmful comments	How might we support business owners in maintaining positive product discussions?
An HR manager who monitors the company's internal communication platform needs a tool to identify offensive language because a healthy work environment is crucial.	HR manager monitoring internal communication	Ensure a safe and respectful communication environment.	How might we enable companies to automatically detect and filter offensive language in internal platforms?
A blog owner who shares informative content needs an AI analyser to remove irrelevant or spam comments because it decreases the value of genuine discussions.	Blog owner publishing content	Promote meaningful discussions by removing irrelevant comments.	How might we automatically distinguish between meaningful and spam comments?
A government official who monitors social media for public feedback needs a comment analyser to identify hate speech or political hate comments.	Government official monitoring public feedback	Ensure public safety by eliminating hate speech from comments.	How might we detect and remove politically influenced hate speech automatically?

14. Develop POV/How Might We (HMW) Questions to Transform Insights/Needs into Opportunities for Design

User Need/Insight	"How Might We" Question
Users need a tool to automatically identify and remove toxic comments to protect brand reputation.	**How might we build an AI system that can detect and remove toxic comments in real-time?**
Users need an AI-powered system to filter out spam or irrelevant comments from social media platforms.	**How might we develop an AI tool that can automatically distinguish between meaningful and spam comments?**
Users want to protect their mental health from negative comments when uploading content online.	**How might we enable content creators to create a safe and positive comment space using AI?**

Business page owners need to maintain a positive brand image by removing harmful comments.	**How might we provide a smart comment filter to protect the brand's image from toxic comments?**
Online gaming communities need to eliminate toxic chat messages to create a positive gaming environment.	**How might we create an AI model to identify and eliminate hate speech in gaming chatrooms?**
Government organizations need to control public hate speech in social media platforms.	**How might we develop a comment analyzer that can detect and prevent politically influenced hate speech?**
Customer service teams need to quickly identify negative feedback to improve product quality and services.	**How might we create an AI system that automatically highlights negative feedback for product improvement?**
Content creators need a solution to handle thousands of comments without manual moderation.	**How might we build an AI system that can automatically moderate and categorize comments?**

16. Crafting a Balanced and Actionable Design Challenge

Design Challenge:
Develop an **AI-driven Comment Analyzer system** that can automatically detect and filter **toxic, spam, and irrelevant comments**, ensuring a **positive and safe digital space** for individuals, businesses, and organizations.

17. Validating the Problem Statement with Stakeholders for Alignment

Validation Plan:

The plan involves engaging stakeholders such as social media users, business owners, content creators, and platform administrators through surveys, feedback sessions, and demos to validate the problem statement.
Key questions will focus on:

- Relevance of the AI tool for comment moderation.
- Accuracy in detecting harmful or irrelevant comments.
- Impact on user experience and brand reputation.
- Suggestions for improvement in the AI algorithm.

Stakeholder/User	Role	Feedback on Problem Statement	Suggestions for Improvement
Social Media Managers		Highlighted the need for improved detection of context-based toxic comments.	Suggested adding customizable moderation filters for different types of content.
Content Creators		Manages audience interactions	Emphasized the need for better sentiment analysis to reduce false positives.

18. Ideation

Ideation Process:

Idea Number	Proposed Solution	Key Features/Benefits	Challenges/Concerns
Idea 1	AI-Based Toxic Comment Detection	Improved detection of hate speech, harassment, and harmful language.	Balancing between freedom of speech and content moderation.
Idea 2	Real-Time Comment Filtering	Immediate identification and removal of harmful comments.	Ensuring fast processing during high-traffic events.
Idea 3	Sentiment Analysis Integration	Detects emotional tone (positive, negative, neutral) in comments.	Difficulty in interpreting sarcasm or coded language.
Idea 4	Spam and Bot Detection	Identifies and blocks automated spam attacks.	Potential for false positives.
Idea 5	Multi-Language Support	Moderate's comments in multiple languages.	Requires comprehensive training data for accurate results.

Solution Concept Form

1. Problem Statement:

● Digital platforms struggle to manage toxic comments, spam, and harmful content effectively.

2. Target Audience:

● Content creators, social media managers, online forum moderators, digital marketing teams.

3. Solution Overview:

● An AI-driven comment analyser that automatically detects and filters toxic language, spam, and inappropriate content using NLP and machine learning.

4. Key Features:

Feature	Description
Feature 1	**AI-powered comment analysis with real-time moderation.**
Feature 2	**Customizable filters for content-specific needs.**
Feature 3	**Sentiment analysis to identify emotionally charged language.**

---

5. Benefits:

Benefit	Description
Benefit 1	**Reduces toxic comments, improving online safety.**
Benefit 2	**Protects brand image by removing harmful content.**

---

6. Unique Value Proposition (UVP):

- **AI-powered, real-time comment analysis** that detects harmful and toxic comments while promoting a safer and healthier online environment.

---

7. Key Metrics:

- **Comment Classification Accuracy,**
- **Sentiment Detection Efficiency,**
- **False Positive and Negative Rates,**
- **Processing Speed for Real-time Analysis.**

---

8. Feasibility Assessment:

- Availability of large datasets from **social media platforms, online forums, and digital communities**.
- Infrastructure needs, such as **cloud computing resources, machine learning models, and data storage.**
- Ensuring the AI model can process **multi-lingual comments** and minimize bias.

---

9. Next Steps:

- **Develop robust AI models** capable of accurately detecting harmful and toxic comments.
- **Collaborate with social media platforms** to test the model in real-time scenarios.
- **Deploy a pilot project** to analyze user behavior, model performance, and feedback accuracy.
- Continuously **optimize the model** for better accuracy and reduced false positives.
