

# **Analyzing Joke Comment Behavior on**

## **PBS Eons YouTube Videos**

**Project Manager's Name:** Sumel Rattan

**Team Members:**

Anubhav Mathur

Sumel Rattan

Siddheshwar Singh Negi

**Mentor Name:** Professor Meaghan Wetherell

**Date:** March 3, 2025

## 1. Revision History Table

Revision History Table		Template Date
		4/01/2025
Version	Summary of Changes	Date
0.1	Section Owners assigned; Draft Introduction added	03/03/2025
0.2	First draft of all sections completed	03/05/2025
0.3	All sections updated during team review	03/10/2025
0.4	Version submitted for signatures	03/12/2025

---

## 2. Summary

*This section provides a brief overview of the project and its goals.*

**Author:** Siddheshwar Singh Negi

The project aims to analyze the behavior of joke comments on PBS Eons YouTube videos. By studying the interaction and frequency of joke-related comments, we aim to understand viewer engagement patterns, emotional reactions, and their impact on video views. This analysis will help in understanding the influence of humor on social media interaction and engagement with educational content.

Our solution will employ sentiment analysis and pattern recognition algorithms to classify comments as joke or not a joke, providing insights into the overall engagement dynamics of the

PBS Eons channel. The project will culminate in a comprehensive report and a presentation to **Professor Meaghan Wetherell** for review.

---

### 3. Project Overview

*This section gives a detailed explanation of the problem, its significance, and the scope of the project.*

**Author:** Sumel Rattan

The purpose of this project is to investigate how viewers interact with educational content through humor by classifying the comments as a Joke or not a Joke, specifically for PBS Eons YouTube videos. We will focus on identifying and analyzing comments that contain jokes or humorous elements to understand their influence on community interaction.

**Objectives:**

- Identify the frequency of joke comments on PBS Eons YouTube videos.
- Analyze viewer sentiment regarding jokes and humor.

**Deliverables:**

- A detailed analysis of comments that are identified as Joke or humor.
- Data visualization demonstrates trends and insights.
- Final report and presentation to the Faculty Advisor to show the percentage of Joke comments from the total number of comments from the top 20 viewed videos.

## 4. Product Specification

*This section defines the product in detail, explaining how it works, how it is used, and how it will be tested.*

**Author:** Anubhav Mathur

Our product will be a sentiment analysis tool that categorizes YouTube comments on PBS Eons videos into humorous and non-humorous comments. We will leverage large language models (LLMs) from Hugging Face and other publicly available models to train classifiers and identify whether a comment is a joke or not. Additionally, we will use Python libraries like NLTK and spaCy for text processing. The system will generate insights and reports of engagement patterns based on the identified joke comments.

### **Product Features:**

- Automatic classification of joke comments.
- Sentiment analysis to classify emotional tones.
- Graphical visualizations of viewer engagement patterns.

### **Testing Methodologies:**

- We will test the system by comparing the algorithm's predictions with manually classified comments to ensure accuracy.
  - We will evaluate engagement metrics before and after humorous comments to measure the influence of jokes on audience behavior.
-

## 5. Team Roles and Responsibilities

*This section outlines the responsibilities of each team member, including the project manager and faculty advisor.*

**Author:** Siddheshwar Singh Negi

### **Project Manager Responsibilities:**

- Lead the team meetings and discussions.
- Manage the project timeline and ensure milestones are met.
- Coordinate with the Faculty Advisor for project updates.

### **Team Member Responsibilities:**

- **Anubhav Mathur:** Lead on sentiment analysis algorithms and Python scripting.
- **Sumel Rattan:** Responsible for the overall project management with a focus on data collection, and pre-processing.
- **Siddheshwar Singh Negi:** Responsible for analyzing viewer engagement, analysis report, and project documentation.

### **Faculty Advisor Responsibilities:**

- Professor Meaghan Wetherell will provide guidance on methodology and feedback on the deliverables.

---

## 6. Ground Rules and Expectations

*This section explains how the team will conduct the business of completing the project, manage dissenting views, and make decisions.*

**Author:** Sumel Rattan

As a team, we agree to the following ground rules:

- Maintain open communication during all stages of the project.
- Meet weekly to discuss progress and address issues.
- Provide constructive feedback and resolve conflicts collaboratively.
- Ensure all team members meet deadlines and deliver high-quality work.

---

Contents

1. Executive Summary.....	7
2. User/Market research.....	8
3. Final Product Deliverable & Features.....	12
Feature 1: .....	13
Feature 2: .....	13
4. Project Timeline & Gannt Chart .....	15
5. Ethics .....	17
6. Approvals.....	20
7. Appendix.....	21
1) Project Team Responsibilities.....	21
2) Faculty Advisor Responsibilities .....	21
A. Ground Rules.....	22

## 1. Executive Summary

*Written by Siddheshwar Singh Negi*

**Our project, "Analyzing Joke Comment Behavior on PBS Eons YouTube Videos," aims to explore the patterns, sentiment, and engagement surrounding joke comments on PBS Eons' YouTube channel. The project will leverage Natural Language Processing (NLP) and machine learning techniques to analyze user-generated comments, identify trends in humor, and assess the impact of joke comments on audience engagement. Key features of the project include sentiment analysis, topic modeling, and engagement metrics tracking. By visualizing our findings through data-driven report, we aim to provide insights into how humor shapes online discourse in educational content.**

Understanding the dynamics of joke comments is crucial, as humor plays a significant role in online discussions, fostering engagement and shaping community interactions. Current research on **social media comment analysis** often focuses on sentiment classification, misinformation detection, or general user engagement, but few studies specifically examine **humor-driven comment behavior in an educational context**. Our analysis will bridge this gap by providing **quantifiable insights** into joke comment trends, helping content creators and educators **optimize viewer interaction strategies**. Unlike existing sentiment analysis tools, our solution will **differentiate between humor and general commentary**, ensuring a **more nuanced understanding of audience engagement**.

Our team—**Sumel Rattan, Anubhav Mathur, and Siddheshwar Singh Negi**—will conduct this project under the guidance of **Professor Meaghan Wetherell**. The development process

involves **data collection from PBS Eons YouTube videos, preprocessing and cleaning the comment data, applying machine learning models for sentiment and humor detection, and visualizing the results through a publishable report.** We will use **Python, LLMs from Huggingface and other publicly available models, NLP libraries (NLTK, spaCy, transformers), and data visualization tools (Plotly, Matplotlib, Tableau).** By the end of the semester, we will deliver a **functional analytical tool**, a detailed research report, and a final presentation showcasing our findings.

**Table 1: Preliminary Feature Responsibilities**

Team Member	Feature responsibility
Sumel Rattan	Data collection, preprocessing, model training (ML/NLP)
Anubhav Mathur	Sentiment analysis, topic modeling, data visualization
Siddheshwar Singh Negi	Analysis Report, engagement metrics tracking, project documentation

## 2. User/Market Research Analysis

**Project Name:** *Analyzing Joke Comment Behavior on PBS Eons YouTube Videos*

**Team Members:** Sumel Rattan, Anubhav Mathur, Siddheshwar Singh Negi

**Mentor:** Professor Meaghan Wetherell

---

### Overall Market Analysis

The product falls within the domain of **social media analytics, natural language processing (NLP), and digital content engagement analysis.** The global **social media analytics market** is projected to grow significantly, with a compound annual growth rate (CAGR) of around **23.6%**



from 2021 to 2028. The demand for **comment analysis, sentiment detection, and user engagement prediction** is increasing due to the rise of online content consumption, particularly on platforms like **YouTube, TikTok, and Instagram**.

Key Market Factors:

- **YouTube as a dominant platform:** With **2.7+ billion monthly active users**, YouTube generates billions of comments, making comment analysis a lucrative area.
- **Rise of AI in digital marketing:** Companies and creators are leveraging **AI-powered NLP models** to improve video reach and audience engagement.
- **Virality and content optimization:** Understanding what drives engagement and virality on educational channels like PBS Eons can help **science communicators and content creators** enhance their marketing strategies.

---

### Existing Competitors

Several companies and tools already provide **comment analysis and engagement tracking** on YouTube:

1. **Vidooly** – Provides YouTube analytics, including comment trends, but lacks deep NLP-based joke classification.
2. **Socialbakers** – Focuses on social media insights but does not specialize in joke/meme detection.
3. **Brandwatch** – A sentiment analysis tool for social media, but primarily for marketing purposes.

4. **YouTube Studio (Native Analytics)** – Offers comment filtering and engagement insights but lacks AI-driven classification for joke formats.

**How Our Product Differs:**

- Focuses on **meme/joke detection in comments**, which is **not the primary focus of current tools**.
- Uses **NLP and machine learning (Transformers, BERT, LLMs)** to classify humor patterns, sentiment, and engagement trends.
- **Identifies patterns in virality** based on repetitive joke comments, providing **actionable insights** for educational content creators.

**User Insights (Empathy Interviews & Pain Points)**

Key stakeholders identified during the meeting:

1. **Educational YouTube content creators** (e.g., PBS Eons)
2. **Social media marketers and analysts**
3. **Researchers in digital engagement and NLP**

**Insights & Pain Points:**

User Pain Points	How Our Product Addresses Them
YouTube video virality is unpredictable.	Our tool identifies how repetitive joke comments correlate with increased engagement and virality.
Meme-based humor is difficult for NLP models to classify.	We analyze grammatical structures and common patterns to improve meme classification.
YouTube comment filtering affects data reliability.	We consider pre-filtered comment behavior and explore access to hidden/removed data.

Commented [SR1]: Can you update the table borders to a single line border ?

User Pain Points	How Our Product Addresses Them
<b>Manual classification of comments is time-consuming.</b>	Automates the classification of jokes, gratitude, and other sentiments using AI models.
<b>No existing tools provide structured insights on joke/meme behavior.</b>	Our approach focuses on humor classification and its impact on audience engagement.

---

### Market Opportunity & Future Scope

- **Potential Expansion:** The model can be extended to **other educational YouTube channels, brands, or platforms like TikTok, Instagram, and Reddit** for meme and joke trend analysis.
- **Monetization Possibilities:** The product can be delivered as a **YouTube plugin or a research API**.
- **Further Research Areas:** Can be enhanced with **temporal trend analysis, multimodal sentiment analysis (text + emojis), and cultural variations in humor detection**.

---

### Next Steps

1. **Scrape and analyze YouTube comments** to establish baseline joke detection accuracy.
2. **Refine classification models** to improve joke/meme differentiation.
3. **Validate findings with content creators** (e.g., PBS Eons team) to ensure practical applicability.
4. **Explore additional variables** like timestamps, replies, and comment length for deeper insights.

This research will **not only aid PBS Eons** in understanding their comment behavior but also provide valuable insights for **science communicators, digital marketers, and AI researchers** working on social media engagement strategies.

### 3. Final Product Deliverable & Features

**Project Name:** *Analyzing Joke Comment Behavior on PBS Eons YouTube Videos*

**Team Members:** Sumel Rattan, Anubhav Mathur, Siddheshwar Singh Negi

**Mentor:** Professor Meaghan Wetherell

---

#### Product Overview

The product is an **AI-powered comment analysis tool** that examines joke-based comments on YouTube videos, particularly PBS Eons. It uses **natural language processing (NLP)** and **machine learning** to classify comments into humor types, measure engagement levels, and analyze their impact on video virality.

This tool helps **content creators, digital marketers, and researchers** understand audience behavior and improve content strategies.

### Minimum Feature List

Feature	Description	Owner	Why the User Wants It?
1. YouTube Comment Scraper	Extracts comments from PBS Eons videos using YouTube API.	Sumel and Anubhav	Automates data collection for analysis without manual intervention.
2. Joke Classification Model	Uses NLP to classify joke-based comments. Identifies types of humor.	Anubhav and Sumel	Helps content creators understand the kind of jokes their audience makes.
3. Engagement Analysis	Measures how joke-based comments influence video views, likes, and shares.	All members	Determines if humor contributes to virality and engagement.
4. Sentiment Analysis	Detects sentiment polarity (positive, neutral, negative) of joke comments.	All members	Understands audience reactions beyond simple classification.
5. Analysis Report	Publishable report analyzing comment trends, humor types, and engagement metrics.	All members	Makes insights accessible and actionable for users.

Commented [SR2]: The table needs to be consistent and should have single line borders

### Stretch Goals (Advanced Features for Resume Boosting)

Feature	Description	Owner	Why the User Wants It?
6. Temporal Analysis	Tracks how joke frequency changes over time on trending videos.	Anubhav and Sumel	Helps creators time their content strategy for maximum impact.
7. Emoji & Hashtag Processing	Enhances joke classification by analyzing emojis and hashtags.	Siddheshwar	Captures humor nuances beyond just text.
8. Predictive Model for Virality	Predicts if a video will go viral based on joke comment patterns.	Sumel and Anubhav	Provides insights into video performance before it peaks.

Commented [SR3]: Same as above, this needs single line borders as well

## Scope & Business Need

- **Why Users Want This?**
  - Content creators need **data-driven insights** to understand what makes their audience engage with humor.
  - Digital marketers can leverage joke-based engagement patterns to **optimize branding and ad strategies**.
  - Researchers gain a **real-world dataset for NLP and AI humor detection studies**.
- **Business Benefit**
  - This tool aligns with the growing demand for **AI-driven content analysis** and **social media analytics**.
  - The project can be expanded into a **full-scale SaaS product** for YouTube creators and brands.

---

## Next Steps & Feature Lock

- **Finalize core functionalities** and freeze scope to prevent feature creep.
- **Distribute implementation tasks** based on expertise.
- **Develop, test, and validate** models with real YouTube data.
- **Iterate based on feedback** to improve accuracy and usability.

## 4. Project Timeline & Gantt Chart

**Project Name:** *Analyzing Joke Comment Behavior on PBS Eons YouTube Videos*

**Team Members:** Sumel Rattan, Anubhav Mathur, Siddheshwar Singh Negi

**Mentor:** Professor Meaghan Wetherell

**Milestone** | **Schedule**

Milestone	Date	Description	Owner
Team Formation	03/10/2025	Define team roles and responsibilities.	All Members
Sub-System Partitioning & Assignments	03/13/2025	Divide projects into components and assign tasks.	All Members
Technology Selection	03/16/2025	Finalize programming languages, NLP models, and API usage.	All Members
YouTube Data Collection Script	03/19/2025	Implement and test YouTube API for comment scraping.	All Members
Joke Classification Model Development	03/23/2025	Train and evaluate machine learning models for humor detection.	All Members
Engagement & Sentiment Analysis	03/25/2025	Develop metrics to analyze comment impact and sentiment.	All Members
Signed Proposal Submission	04/04/2025	Submit finalized proposal with project scope and objectives.	All Members
Integration of NLP & Visualizations	04/10/2025	Combine all models and features into a cohesive report.	All Members
Initial Testing & Debugging	04/15/2025	Unit testing and model performance evaluation.	All Members
Full Function Testing & Modifications	04/22/2025	Run extensive tests and make improvements.	All Members
Poster Demo Preparation	04/26/2025	Prepare slides, poster, and report documentation.	All Members
Final Testing & Refinements	04/29/2025	Conduct final tests and polish report for better usability.	All Members
iShowcase Presentation	05/01/2025	Present final product and report at iShowcase event.	All Members

Commented [SR4]: This needs single line borders as well.

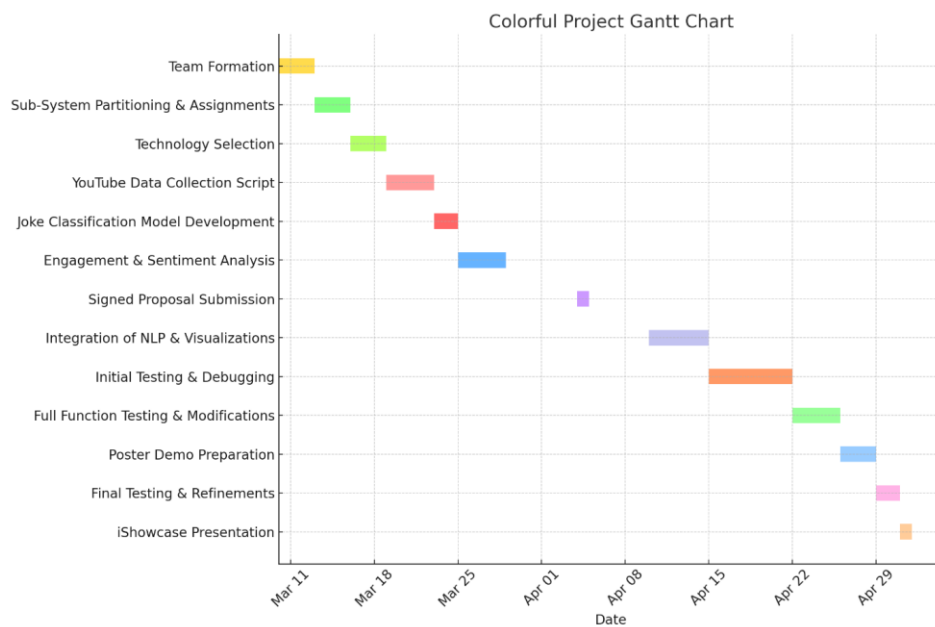
Commented [SR5]: Add the table here

### Gantt Chart Overview

The **Gantt Chart** will visually track each milestone, dependencies, and team responsibilities over time. (To be created using Excel, Trello, or a dedicated Gantt Chart tool like MS Project or ClickUp.)

#### Key Considerations:

- **2-3 weeks of testing** to ensure reliability.
- **Buffer time** for unexpected delays or refinements.
- **Parallel development approaches** where feasible (e.g., sentiment analysis and classification can progress simultaneously).





## 5. Ethics

In this context:

- **Y (Yes)** → The platform allows or could potentially allow this activity.
- **N (No)** → The platform does not allow this activity and has strict measures to prevent it.
- **M (Maybe)** → The platform does not explicitly allow it, but there is a possibility that users could bypass restrictions or exploit loopholes.

#	Question	Generally	Data Breach
1	Could a user sell drugs or other illegal items on your platform?	N	N
2	Could a user of your platform engage in sex trafficking?	N	N
3	Could a user sell class notes or cheat on their homework on your platform?	N	N
4	Could a stalker use your project to find someone?	N	N
5	Could your app be used to spy on or track individuals?	N	N
6	Could your app/software access the camera or microphone and record things without users being aware?	N	N
7	If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way?	M	M
8	Could your algorithm promote material that would traumatize or upset individuals?	M	M
9	Would your users be upset if the data you collect was given to someone else?	M	Y
10	Could a data leak potentially lead to identity theft?	N	N

#	Question	Generally	Data Breach
11	If your site was hacked, would users of that product potentially lose their job, spouse, or family?	N	N
12	Should there be an age limitation on your product?	N	N
13	Could someone use your product to find, contact, and potentially commit elder abuse?	N	N
14	If the data on your platform was breached, could it be used to blackmail the users?	N	N
15	Does the existence of your project imply that a particular racial group, gender, religion, or other protected category is inherently bad, gross, or unwanted?	N	N
16	Could your product be used to commit hate crimes against a specific group?	N	N
17	Does the primary content of your algorithm focus on something considered deeply unethical?	N	N
18	Does your software contain race, gender, or other stereotypes?	M	M
19	Could users of your app scam other individuals?	N	N
20	Is your algorithm biased towards predicting correctly only for one race, gender, or other group?	M	M
21	Are the users of your project or those being surveyed for your data aware of how their data will be used?	Y	Y
22	What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it?	M	M
23	Does the use or purchase of your data potentially contribute to a dangerous group or regime?	N	N
24	Could your virtual reality environment cause injury to the user?	N	N

#	Question	Generally	Data Breach
25	Are your study participants or game players aware that their data will be collected and used?	Y	Y
26	Does your project contain addictive design elements without benefit to the user?	N	N
27	Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment?	N	N
28	Could your research outcomes harm an individual or entity?	M	M

#### Mitigation Strategies:

- **Mental health impact (Q7 & Q8):** The dataset may contain sensitive historical data that could unintentionally trigger certain individuals. Proper disclaimers and ethical use of the data should be ensured.
- **Bias in data or algorithm (Q18 & Q20):** Since data could show performance variations among different groups, it's crucial to acknowledge potential biases and ensure that no unfair conclusions are drawn.
- **Data privacy concerns (Q9 & Q21):** If any personally identifiable information (PII) exists, proper anonymization techniques should be applied before sharing or publishing.
- **Misinterpretation of results (Q22):** Any insights derived from the data should be presented with clear disclaimers to prevent misrepresentation.
- **Potential harm from research outcomes (Q28):** Ensuring that results are used only for ethical and constructive purposes, not for unfair advantages or discrimination.

## 6. Approval Table

Commented [SR6]: This needs to be single-bordered just like above tables

Approver Name	Title	Signature	Date
Sumel Rattan	Team Project Manager	<i>Sumel Rattan</i>	Apr 8, 25
Anubhav Mathur	Team Member	<i>Anubhav Mathur</i>	Apr 8, 25
Siddeshwar Singh Negi	Team Member	<i>Siddeshwar Singh Negi</i>	Apr 8, 25
Professor Meaghan Wetherell	Faculty Advisor		
Professor Greg Chism	Instructor ( <i>signs after grading</i> )		

## Author Contribution Table

Commented [SR7]: Needs single border as well.

Section	Author	Word Count
1. Introduction	Siddheshwar Singh Negi	300
2. Project Scope	Anubhav Mathur	310
3. Deliverables	Sumel Rattan	315
4. Timeline & Milestones	Siddheshwar Singh Negi	350
5. Budget & Resources	Anubhav Mathur	418
6. Risk Management	Sumel Rattan	400
7. Conclusion	Siddheshwar Singh Negi	300
Tables/Diagrams	All Members	200

## 7. Appendix

### 1) Project Team Responsibilities

- The **Project Manager (Sumel Rattan)** will set up and facilitate a weekly call/meeting with the **Faculty Advisor (Professor Meaghan Wetherell)**. The **Project Team** will provide weekly status updates to the Faculty Advisor, including upcoming deliverables, critical issues, and any adjustments to the Project Plan.
- Documents will be provided to the Faculty Advisor with **adequate time for review and signature**. The minimum review time will be **3 days** prior to the document due date.
- Design files will be provided to the Faculty Advisor as requested in a format agreed to with the Advisor.
- Support requirements will be clearly requested from the Faculty Advisor with the dates required and adequate time for fulfilling the request.
- Modification requests to the Project Plan by the Faculty Advisor will be **reviewed and agreed to within 1 week of the request**.

### 2) Faculty Advisor Responsibilities

- The **Faculty Advisor (Professor Meaghan Wetherell)** will provide **knowledge and expertise** to help the group stretch their skills.
- The Faculty Advisor will participate in a **weekly or bi-weekly call/meeting** with the Project Team to review the project status, upcoming deliverables, priorities, issues, and progress according to the agreed Project Plan.

- The Faculty Advisor will provide **document review, feedback, and approval/rejection** with **adequate time** for the Project Team to meet course due dates.
- The Faculty Advisor will provide **feedback and guidance** on design implementation decisions, design files, test plans, test procedures, and test results.
- The Faculty Advisor shall provide **technical advice and guidance**, answering inquiries approximately **1 hour per week**.
- Modifications to the Project Plan by the Project Team will be **resolved and documented within 1 week of the request**.
- The Faculty Advisor will **grade the finalized project using a skill-based rubric**.
- The Faculty Advisor will **attend iShowcase in May**.

#### A. Ground Rules

##### How the team will conduct the business of completing this project:

Each team member must sign the **Approvals** section to acknowledge these ground rules.

As a team and as individual team members, we agree to:

1. **Stay focused on our objectives and goals.**
  - Each meeting will have clearly defined objectives and desired outcomes.
  - Team members will politely remind each other if discussions go off track.
2. **"Sidebar" any issues that are not immediately relevant.**
  - Unrelated but important topics will be noted and discussed later.
3. **Listen when others are speaking.**
  - Team members will listen and consider others' input before responding.

**4. All viewpoints will have an opportunity to be heard.**

- We will encourage quieter team members to share their thoughts.
- No single person will dominate discussions.

**5. Respectfully discuss differences of opinion.**

- Identify areas of agreement before addressing disagreements.
- Encourage open-minded discussions and respect final team decisions.

**6. Look for the good points in new ideas.**

- Every idea will be considered before being assessed.

**7. Focus on the future, not the past.**

- Use past experiences to inform decisions but prioritize future solutions.
- Blaming past performance is counterproductive.

**8. Agree upon specific action items and next steps.**

- At the end of meetings, clear action items will be assigned.

**9. Accountability.**

- Each team member is responsible for their tasks.
- We will honor our responsibilities and **not let the team down**.