	Feature	Description	Technical Considerations	Ethical Considerations
1	Data Sources	Collect users' public posts, comments, earlier performance, and reports from other users to gather behavioral data.	Ensure user consent for data collection, anonymization, and legal compliance with data privacy laws.	Respect users' privacy and only use publicly available information.
2	Behavior Assessment Criteria	Evaluate for bullying language, inappropriate posts, and past flagged behaviors.	Develop clear and unbiased criteria for evaluating harmful behaviors.	Avoid over-penalizing borderline cases and provide clear feedback on assessments.
3	NLP Model Role	Analyze text content to detect harmful language and behaviors using NLP tools.	Utilize advanced NLP models like Detoxify or ToxiCraft to ensure robust detection.	Ensure fairness in NLP model training and prevent biases in detection algorithms.
4	Assessment Result	Provide a score (0-100) reflecting the user's behavior quality, based on NLP analysis and other inputs.	Implement scalable scoring logic and real-time updates for dynamic behavioral changes.	Avoid stigmatization of users with low scores; promote improvement instead.
5	Color Coding System	Assign colors to scores: Red for aggressive/inappropriate behavior, Green for safe and appropriate behavior.	Ensure accurate and non-biased color assignment for behavior ratings.	Communicate transparently about how scores and colors are determined.
6	Visibility and Impact	Marks and color codes are visible only to minors to guide their interaction decisions.	Prevent misuse or overexposure of marks, maintaining user trust and ethical handling.	Protect minors' sensitive data and identities while displaying adult behavior ratings.
7	User Safety Mechanism	Alerts minors about potentially harmful interactions, enabling them to avoid engaging with flagged users.	Develop safeguards against false positives/negatives and provide user appeal mechanisms.	Promote education on cyber safety and respectful communication.