At its core, Colour Guard is powered by advanced NLP tools like Detoxify and ToxiCraft, which analyze public posts, previous violations, and user engagements to detect harmful behaviors such as hate speech, bullying, or grooming. ***(Fundamental Research)***

The system assigns behavior scores based on severity, frequency, and context—ranging from minor to severe threats—and dynamically updates a color-coded indicator visible only to minors. This indicator transitions from red to green, reflecting a progression from 0% to 100% and a shift from high risk to low risk.

***(Scoring System & Color Coding System)***

To ensure fairness and privacy, we anonymize data and comply with all legal and ethical standards. Historical violations are considered with a time-decay mechanism, allowing scores to recover for improved behavior, encouraging growth instead of stigmatization.

We ensure fairness and privacy by anonymizing data, complying with legal standards, and applying time-decay for score recovery, promoting growth over stigmatization.

**(Fairness & Privacy)**

Colour Guard isn't just a tool, but also a transformative **Multilateral Feedback Ecosystem**. It offers safer interactions for minors with real-time alerts and actionable insights for guardians and educators. For platforms, it drives profitability through regulatory compliance, user trust, and revenue from subscriptions, premium features, and anonymized data. Additionally, platforms can build long-term educational collaborations, enhance brand trust, and unlock diverse advertising opportunities beyond family-focused spaces.

**(Multilateral Feedback Ecosystem)**