I contributed by proposing the color indicator system, conducting fundamental research on NLP, refining the scoring system and multilateral feedback ecosystem, and presenting the Colour Guard system.

Colour Guard is powered by advanced NLP tools like Detoxify and ToxiCraft. NLP models analyze user content and assign deductions based on the type, severity, and frequency of harmful behavior. Combines NLP deductions, historical penalties, and time-decay factors to calculate a behavior score. The behavior score is mapped to a corresponding color on the color bar. Marks and color codes are visible only to minors to guide their interaction decisions.