Safe Online Standards

IMPROVING DIGITAL MENTAL HEALTH

Safe Online Standards (S.O.S.) Ratings

Helping families understand social media safety at a glance

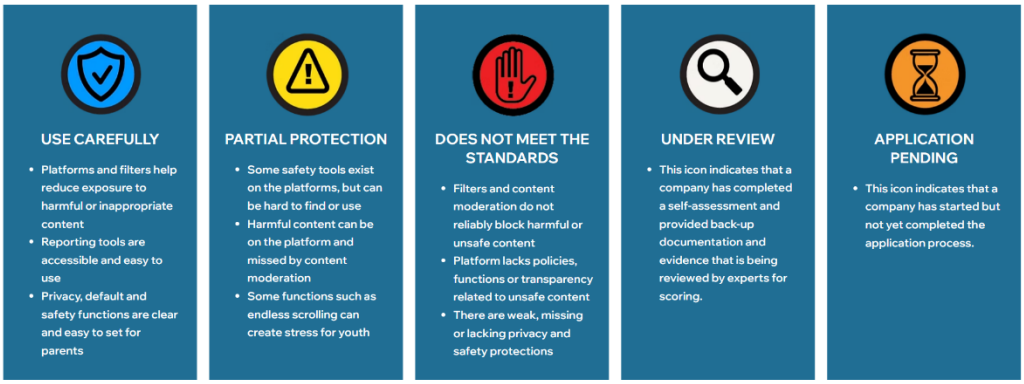
The S.O.S. ratings help parents and teens (ages 13–19) see how social media platforms approach safety, mental health, and well-being. Each platform is reviewed by experts in the field of mental health using expert-developed standards informed by real user experiences. Platforms receive an overall rating that is derived from the ratings given to 3 to 7 standards that are assigned to five categories. The standards have been established to assess strengths and areas of concern. S.O.S. has been designed to inform the user toward safer choices and to encourage the platforms to continually improve their efforts toward a safer platform for its users. It was not designed to provide definitive resolutions to its users or to endorse any platform, and no platform is completely safe.

How to Understand the Ratings

Platforms are reviewed across five categories: Policy, Functionality, Governance and Transparency, Content, and Digital Literacy and Well Being. Icons show how well a platform meets S.O.S. standards established within each of these five categories. Patterns across categories matter more than any single score.

Overall Rating - Icon

The overall rating reflects how consistently a platform meets S.O.S. standards **across all** five categories.



Category Ratings – Color-coded symbol

The category rating shows how a company performed in **each of the** five S.O.S. categories.

SOS Categories

- **Policy:** What safety rules exist and how clearly, they are defined.
- **Functionality:** How platform features, algorithms, and design choices affect teen safety.
- **Governance & Transparency:** How safety decisions are made, enforced, and shared with the public.
- **Content:** How harmful content is identified, moderated, and addressed.
- **Digital Literacy & Well-Being:** How users are supported with education, mental health resources, and tools for healthy use.

Safe Online Standards

IMPROVING DIGITAL MENTAL HEALTH

🔵 **Meets threshold** 🔴 Partially meets threshold 🔴 **Below threshold in the category**

Platform	Content	Digital Literacy	Functionality	Governance & Transparency	Policy
Company A	🔵	🟡	🔵	🔵	🟡
Company B	🔵	🟡	🔴	🔵	🔵
Company C	🔴	🔵	🔴	🟡	🔵

Important Note: All social media platforms carry some level of risk. S.O.S. ratings reflect current evidence and practices and will evolve as research, expert insight, and user feedback grow. Even highly rated platforms benefit from parental involvement, open communication, and ongoing awareness.

In addition, as advertisers significantly fund these platforms, the S.O.S. provides a welcome and transparent vehicle for brands and agencies to publicly align their values with their messaging—supporting digital environments that prioritize youth, mental health and well-being. By incorporating S.O.S. ratings into media planning and brand safety decisions, advertisers can help reinforce industry-wide incentives for safer, healthier online spaces.

Advertisers can play an important role in ensuring technology companies adhere to standards, just as advertisers themselves are expected to follow established standards in the advertising industry. Today, the Brand Safety Institute offered support for the S.O.S. standards.

Risk Categories

This language is provided to help parents and youth more deeply understand where there might be risks based on the scoring of the standards.

Policy: this category has three standards that assess the company's policies on user safety, advertising and research on teens (13-19).

Governance and Transparency: this category has seven standards that are intended to assess the company's Compliance Teams, how transparent they are about harmful content and user reports on their platform, what data is collected and/or publicly shared around research, and exposure to violence and advertising to teens (13-19).

Content: this category has five standards that are intended to assess how harmful content is identified and addressed, if credible sources of reliable information are provided to users, and how the company addresses content moderation.

Functionality: this category has four standards that are intended to assess various features and design choices that effect teen safety on the platform regarding harmful content and how to report it, amount of user time on the platform, and parental tools and controls provided by the platform to protect teens (13-19).

Digital Literacy and Wellness: this category has five standards that are intended to assess how a company supports users with educational content, mental health resources and tools for healthy use.

Risk Details

Policy: generally, this category informs users on the accessibility, development (user input), evidence-based, enforceable actions, company response, of policies around mental health, suicide and self-harm issues, advertising and research on teens (13-19).

Governance and Transparency: generally this detail informs users on the roles and qualifications of personnel at the company, external experts/advisors used by the company, and compliance/regulatory departments within the company, how transparent the company is regarding harmful content and exposure to violence users are, the type and frequency of research and advertising that is conducted and/or participated in by the company both internally and external to the company.

Content: generally, this detail informs users what types of harmful content are prohibited by the company, how the company addresses credible vs. not credible sources of health-related information, specific guidelines developed for allowable content, how the company categorizes and ensures content is developmentally, age and culturally informed, respectful and inclusive content that does not discriminate against users, how the content might impact a user's mental health and details regarding how content is moderated by the company.

Functionality: generally this detail informs users on the specific functions of the platform (for example, ease and access to user reporting, responses from a company to a report, technology and user features that protect users from exposure to harmful content such as privacy settings and filters), time and break functions, as well as parental controls and settings to protect teens (13-19).

Digital Literacy and Wellness: generally this detail helps users understand the educational content that the platform provides to users on crisis support and service, ease of access to educational content and how it was developed, how to delete personal data from the platform (the right to be forgotten), as well as what specific developmentally informed features exist to determine ability to use the platform.