

## DTSP Safe Assessments Report

### An Inaugural Evaluation of Trust & Safety Best Practices

The Digital Trust & Safety Partnership (DTSP) brings together technology companies providing a wide range of products and services around a common approach to increase Trust & Safety across the internet. In less than a year and a half since DTSP launched, we have moved rapidly toward creating and executing industry-wide [assessments](#) of how companies are implementing [best practices](#), providing a roadmap to meaningfully increase Trust & Safety online. This inaugural report synthesizes the results of the internal Safe Framework assessments conducted by ten DTSP partners during the first half of 2022. Participating partners for initial assessments were Discord, Google, LinkedIn, Meta Platforms, Inc., Microsoft, Pinterest, Reddit, Shopify, Twitter, and Vimeo.

#### **Successes: many companies reported a mature state of development for core content moderation practices**

Trust & Safety teams and functions across the partners have performed relatively well when it comes to core practices and activities that fall squarely within their domain and can be implemented unilaterally, to some degree. These include constituting the teams responsible for content policies and developing public facing policy descriptions, as well as developing enforcement infrastructures that span people, processes, and technology, and notifying users whose content is subject to an enforcement action for violating policies.

#### **Areas for improvement: many of the least mature practices relate to user feedback and external collaboration**

Three of the practices deemed least mature, according to the self-assessments, related to incorporating user and third-party perspectives into Trust & Safety policy and practices. This illustrates the internal focus of Trust & Safety functions. As a discipline, Trust & Safety has developed with less external engagement outside of companies until recently. The least mature of all assessed practices is the creation of processes to support academic and other researchers working on relevant subject matter.

#### **Areas of ongoing improvement: integrating Trust & Safety into product development**

The majority of assessments indicated companies were in the process of formalizing the relationship between Trust & Safety and product teams to better integrate these perspectives into product development.

This report marks the beginning of our collective effort to evolve and evaluate Trust & Safety practices across the industry. Partner companies are using the results of their self-assessments to enhance their practices, and DTSP will facilitate learning to identify collective opportunities to mature key practices. Looking ahead, we foresee the following lines of effort:

- Evolving the DTSP Best Practices Framework: we will use the insights generated through this process and external input to review and improve the DTSP Best Practices Framework;
  - Moving to third-party assessment: we are working with experts with deep understanding of both Trust & Safety and assessment frameworks to articulate our approach to independent third-party assessment and will share more information on this in the coming months; and
  - Engaging stakeholders globally: as we raise awareness of Trust & Safety best practices, we will put in place specific mechanisms for stakeholder input and engagement and provide opportunities for dialogue.
-