Updates
February 05, 2024
Online Safety Risk Assessments Have Arrived: Five Steps for Building a Globally Adaptable Process

Safety risk assessments are becoming a preferred regulatory tool around the world.

Online safety laws in Australia, Ireland, the United Kingdom, and the United States will require a range of providers to evaluate the safety and user-generated content risks associated with their online services.

While the specific assessment requirements vary across jurisdictions, the common thread is that providers will need to establish routine processes to assess, document, and mitigate safety risks resulting from user-generated content and product design. This Update offers practical steps for providers looking to develop a consolidated assessment process that can be easily adapted to meet the needs of laws around the world.

Notable New Requirements

Up until recently, safety and content risk assessment requirements were limited to the largest online platforms. The EU Digital Services Act, which came into effect for "very large online platforms" (VLOPs) in 2023, requires VLOPs to assess systemic risks posed by their platforms, including safety risks. Now, with the adoption of online safety laws in Australia, Ireland, and the U.K., as well as children's privacy and safety laws in the United States and abroad, safety risk assessment requirements will extend to a far broader array of online services. The notable requirements are as follows:

Australian Online Safety Codes (2023 - 2024). Industry codes and standards adopted pursuant to the Australian Online Safety Act require, or will soon require, numerous online services to conduct risk assessments. As of December 2023, providers of social media services (as defined by the regulation) must conduct a risk assessment or default to the highest risk profile. The assessed risk profile must then inform the service's compliance measures under the Social Media Services Online Safety Code.

Several other categories of services will soon have risk assessment requirements, including search, dating, and gaming platforms. Search services will be required to conduct a risk review starting in March 2024. Providers of "designated internet services" (DIS)—defined to include the majority of apps and websites that can be accessed in Australia—will very likely be subject to risk assessment requirements under a forthcoming industry standard issued by the Australian eSafety Commissioner. "Relevant electronic services" (RES), which are defined to include gaming and dating services, will probably have similar requirements. Drafts of both the DIS and RES industry standards are available here.

Providers subject to risk assessment requirements will need to conduct an initial assessment as soon as is practical (and no later than six months after) the code or standard applicable to their service(s) comes into effect.

U.K. illegal content risk assessments (2024). Under the U.K.'s Online Safety Act (U.K. OSA), all user-to-user or search services in the U.K. will need to carry out an illegal content risk assessment. The assessment will need to take into account how a service's design, user base, algorithmic systems, and functionalities affect the likelihood and impact of 15 categories of illegal content. On the basis of the assessment, providers will be required to decide on the risk mitigation measures needed to comply with the U.K. OSA's safety duties and implement and record the changes. Thereafter, risk assessments will need to be updated before launching a new service or making a significant change to an existing service.

Ireland Online Safety Code (expected in 2024). Ireland's Online Safety and Media Regulation Act, which was adopted in December 2022, authorizes a newly created media commission and online safety commissioner to issue a binding Online Safety Code. The draft Online Safety Code, which is currently open for public consultation, would require designated Ireland-based video-sharing platform services to integrate safety impact assessments into their product development process. Providers would be required to prepare an assessment methodology that incorporates safety-by-design principles and mitigates child safety risks, hate speech, and violence risks.

Children's Risk Assessments

- **U.S. states (2024).** Several U.S. states will soon require children's risk assessments. To enjoy safe harbor from liability under California's recently adopted AB 1394, a social media platform must conduct at least a biannual audit of its designs, algorithms, practices, affordances, and features to detect whether any could contribute to child sexual exploitation or abuse.

Connecticut and Florida's respective teen privacy and safety laws, which are scheduled to go into effect in 2024, will require controllers to assess the risks of harm to minors arising from a product's features or practices. And California's Age-Appropriate Design Code (ADCA), which is currently enjoined and on appeal in the U.S. Court of Appeals for the Ninth Circuit, would require similar assessments if it goes into effect.

- **U.K. (2025).** In addition to illegal content risk assessments, user-to-user services that are likely to be accessed by children must carry out a children's risk assessment. The assessment will need to consider, among other things, the age of the user base, the level of risk that children using the service will be exposed to harmful content, and whether the service's design and functionalities exacerbate those risks.
- **EU (2025).** The EU Commission has convened a special group to develop a comprehensive EU code of conduct on age-appropriate design. The EU code will assist with the implementation of the Digital Services Act's (DSA) requirements on the protection of minors. Similar to other age-appropriate design codes, the EU's code may include a children's risk assessment requirement.

Practical Steps To Consider

Selecting the right methodology is critical to developing a risk assessment process that can be adapted to meet requirements across jurisdictions. Rather than approaching each risk assessment requirement on an ad hoc basis, providers have an opportunity to develop and institutionalize an assessment process that enables compliance with regulatory requirements around the world and mitigates related litigation risks.

Below are some steps that providers in scope may wish to consider.

Step 1: Develop assessment guidelines that maintain attorney-client privilege. It is important to select assessors with the requisite skill and know-how to conduct a credible assessment that can be defended before regulators. This initial step should also anticipate the possibility that information gathered and risks identified may be subject to legal process in one or more jurisdictions. Recent consumer protection investigations and enforcement actions related to online safety issues have relied heavily on internal documents and risks identified by providers themselves. As a result, assessment guidelines should take privilege rules in high-risk jurisdictions into account.

Step 2: Map the relevant stakeholders. Risk assessment requirements across jurisdictions require evaluation and testing of various safety policies, processes, and features. For most providers, this will implicate numerous cross-functional teams. Identifying relevant stakeholders at the outset enables greater efficiency, adaptability, and repeatability. Regulators have noted that silos within company structures contribute to safety risks.

Step 3: Discovery process. There are opportunities to develop standardized questionnaires and discovery processes to streamline internal information gathering, as informed by the stakeholder mapping. This standardized process can then be tailored as needed to relevant regulatory requirements.

Step 4: Assess and test for safety risks. This step involves the review and testing of relevant systems, processes, and policies to identify current and foreseeable safety risks.

Step 5: Prioritize and document risk mitigation measures. Identifying mitigation measures that respond to identified risks is a common requirement across jurisdictions. A standardized process to align on priority risk mitigation measures, as informed by the risks identified in Step 4, facilitates compliance. Regulators are also expecting to be able to review documented findings and safety controls. The assessment team should agree on a privilege-protective method of documentation that facilitates routine updates and external reporting.

Takeaways

Risk assessments are one of several tools regulators are using to try to generate online safety accountability from within companies. Global online safety regulators are coordinating their efforts in this space, and other markets, such as India, are actively considering safety risk assessment requirements. In light of this trend, providers may wish to develop a consolidated risk assessment process that can be adapted to varying requirements.

In addition to mitigating regulatory risks, risk assessments may also help to defend against product safety and negligence claims in the United States and abroad. In recent months, several safety-related claims against online platforms alleging product design defects and negligence have survived the motion to dismiss stage in U.S. and state courts. Regulators in Australia, the EU, the U.K., and the United States have also increased their scrutiny of online safety practices through requests for information, investigations, and enforcement actions.

A robust and consolidated safety risk assessment process may enable providers to avoid liability and build trust with regulators and consumers around the world.

Authors

Explore more in

Technology Transactions & Privacy Law Privacy & Security Digital Media & Entertainment, Gaming & Sports

Related insights

Update

February Tip of the Month: Federal Court Issues Nationwide Injunction Against Trump Executive Orders on DEI Initiatives

Update

New US Commerce Prohibitions on Chinese and Russian Connected Vehicle Technology