Digital Trust
& Safety Partnership

Sistema de mejores prácticas

El Sistema de Mejores Prácticas del DTSP 2025 se ha actualizado para alinearse con la publicación del sistema de seguridad como ISO/IEC 25389. Este sistema de mejores prácticas es funcionalmente idéntico a las cláusulas cuatro y cinco de ISO/IEC 25389:2025.

Digital Trust
& Safety Partnership

Introducción

Los servicios digitales son cada vez más esenciales en nuestra vida cotidiana, ya que facilitan la interacción y el discurso social, la actividad económica y mucho más. Estos servicios proporcionan herramientas poderosas para que usuarios de todo el mundo participen en una amplia gama de actividades importantes en Internet. Pero, como cualquier herramienta, también pueden utilizarse indebidamente para facilitar comportamientos y contenidos dañinos. La conciencia sobre este uso indebido y las acciones contra ellos se han incrementado en los últimos años, lo cual ha llevado a una mayor urgencia por comprender, apoyar y evaluar formas efectivas de reducir los daños asociados con el contenido y el comportamiento en internet, protegiendo al mismo tiempo la capacidad de las personas para expresarse, hacer negocios, acceder a la información, asociarse, trabajar, estudiar y participar en sus comunidades a través de servicios digitales.

Lograr este equilibrio presenta un desafío considerable. Para empezar, no existe un único enfoque para el manejo del contenido en internet y los riesgos de comportamiento asociados o, más en general, para las operaciones de confianza y seguridad de las empresas. Dependiendo de la naturaleza del servicio digital, cada uno puede enfrentar riesgos únicos en relación con los diversos productos o características que ofrecen: diferentes amenazas, diferentes vulnerabilidades y diferentes consecuencias. Los productos o funciones pueden interactuar con los usuarios finales directa o indirectamente, así como con otros servicios o negocios. Aquello que resulta una práctica eficaz para un servicio digital puede no ser adecuado para otro, y los enfoques altamente prescriptivos o rígidos para definir las prácticas de confianza y seguridad probablemente sean demasiado amplios, demasiado limitados o tengan consecuencias negativas no deseadas. Además, los riesgos cambian con el tiempo y, por lo tanto, los enfoques para reducirlos también deben tener espacio para evolucionar.

Dada la diversidad de servicios digitales, es importante definir un sistema general y un conjunto de objetivos para lo que constituye un enfoque responsable, al cual los servicios digitales puedan luego asignar sus prácticas específicas. Este enfoque flexible se ha implementado en otros ámbitos como la ciberseguridad, pero los sistemas existentes no son lo suficientemente concretos para aplicarse cuando se trata de abordar comportamientos y contenidos dañinos en internet.

La *Digital Trust & Safety Partnership* (DTSP) busca satisfacer esta necesidad documentando y facilitando la adopción de compromisos generales ampliamente implementados, y establecidos en este documento para fomentar una mayor transparencia y una mejor comprensión de la Confianza y la Seguridad tanto dentro como fuera de la industria. Con ese fin, el Sistema de Mejores Prácticas de DTSP ofrece un sistema común sobre cómo las empresas abordan los riesgos relacionados con el contenido y la conducta. Si bien los compromisos generales son uniformes, el método mediante el cual se cumplen (ya sea mediante la aplicación de las prácticas ilustrativas de este documento u otras alternativas) variará según el producto o característica digital, y evolucionará tanto con los desafíos enfrentados como con los avances logrados en el campo de la Confianza y la Seguridad.

DTSP considera que los cinco compromisos generales representan los pasos necesarios que las empresas practicantes adoptaron para identificar y abordar contenidos y conductas nocivas, preservando al mismo tiempo la libertad de expresión y otros derechos, incluidos los estándares de derechos humanos internacionalmente reconocidos, así como también el valor social y económico de los servicios digitales.

El Sistema de Mejores Prácticas de DTSP marca el primer intento de articular los esfuerzos actuales de la industria para abordar los riesgos relacionados con el contenido y la conducta en línea. El Sistema de Prácticas se centra en consideraciones sobre el desarrollo, la gestión, la aplicación, la mejora y la documentación clara de los productos y servicios digitales. Con el tiempo, este sistema evolucionará a medida que crezcan en madurez y podrá evaluarse de formas más estándar, así como en otras disciplinas como la seguridad y la privacidad.

Compromiso general

Tener en cuenta los riesgos relacionados con el contenido y la conducta en los dominios de desarrollo de producto, gestión, aplicación y mejora, y asignar responsabilidades y recursos en cada dominio.

Las empresas practicantes se toman en serio las cuestiones de confianza y seguridad, como lo demuestra la inversión y el desarrollo de personal y tecnología relevantes; la adopción de principios y consideraciones de confianza y seguridad que respeten los derechos en el desarrollo, gestión, aplicación y mejora de productos; y la documentación adecuada de los productos y servicios digitales. Los siguientes compromisos caracterizan el enfoque de las empresas practicantes en cada una de estas áreas.

Compromiso 1

Identificar, evaluar y ajustar los riesgos relacionados con el contenido y la conducta en el desarrollo de productos.

Objetivo: garantizar que las empresas adopten una previsión adecuada en relación con los riesgos relacionados con el contenido y la conducta, e incorporen conocimientos en las características del producto en consecuencia.

Comentario: Anticipar y reducir el riesgo como parte del desarrollo de productos es una parte importante de la función de confianza y seguridad. A medida que los equipos de productos conciben, iteran y perfeccionan productos o características para su lanzamiento, éstos se evalúan desde la perspectiva de abordar los riesgos potenciales relativos al contenido y la conducta. Con este fin, existen distintos mecanismos a través de los cuales se puede configurar un producto o característica determinada para garantizar que aborde consideraciones de confianza y seguridad. Las empresas practicantes (a) crean procesos para evaluar los riesgos relacionados con el contenido y la conducta cuando desarrollan productos para su lanzamiento al

público, (b) buscan prevenir o reducir esos riesgos en la etapa de desarrollo y (c) continúan evolucionando el producto después del lanzamiento de manera adecuada a la luz de los riesgos observados.

Algunos ejemplos de prácticas que incorporan un compromiso de evaluar y ajustar los riesgos relacionados con el contenido y la conducta en el desarrollo de productos pueden incluir:

- Desarrollar capacidades de conocimiento y análisis para comprender patrones de abuso e identificar mitigaciones preventivas que puedan integrarse en los productos.
- Incluir al equipo de confianza y seguridad, o a una parte interesada equivalente, en una etapa temprana del proceso de desarrollo del producto, incluido a través de comunicación y reuniones, solicitando e incorporando comentarios según corresponda.
- Designar un equipo o gerente responsable de integrar la retroalimentación sobre confianza y seguridad.
- Evaluar las consideraciones de confianza y seguridad de las características del producto, equilibrando la usabilidad y la capacidad de resistir abusos.
- Utilizar equipos internos o de terceros para realizar evaluaciones de riesgos y comprender mejor los potenciales riesgos.
- Proporcionar comentarios continuos sobre consideraciones de confianza y seguridad antes del lanzamiento.
- Proporcionar una evaluación posterior al lanzamiento por parte del equipo responsable de la gestión de riesgos y de los responsables de la gestión del producto o en respuesta a incidentes específicos.
- Iterar el producto a la luz de consideraciones de confianza y seguridad, incluidas las opiniones de los usuarios u otros efectos observados, como garantizar que las perspectivas de las comunidades minoritarias y subrepresentadas estén representadas.
- Adoptar medidas técnicas apropiadas que ayuden a los usuarios a controlar su propia experiencia con el producto cuando corresponda (como bloquear o silenciar)

Compromiso 2

Adoptar procesos explicables para la gestión del producto, incluido qué equipo es responsable de crear reglas y cómo evolucionan las reglas.

Objetivo: Garantizar que las normas y principios que rigen el contenido y la conducta de los usuarios sean claros, rigurosos y coherentes.

Comentario: La gestión del producto incluye reglas y procesos externos e internos mediante los cuales una empresa fomenta ciertas actividades y desalienta otras en relación con su(s) producto(s). Esta función existe además del cumplimiento y reducción de riesgos relacionados con la legislación vigente. Una encarnación de

la gestión del producto son los términos de servicio de una empresa (y, para empresas de múltiples productos, a veces términos múltiples): el contrato entre los usuarios y la empresa que establece expectativas y límites subyacentes. Además, algunas empresas pueden mantener reglas adicionales que abordan más directamente la conducta aceptable, a menudo en un lenguaje más sencillo y concreto. Estas pueden denominarse reglas, pautas comunitarias, políticas de uso aceptable o políticas de contenido. Su redacción y evolución pueden recurrir a comunidades de usuarios o a una combinación de partes interesadas con diversas relaciones con la empresa.

Ejemplos de prácticas que incorporan un compromiso de adoptar procesos explicables para la gestión del producto pueden incluir:

- Establecer un equipo o función que desarrolle, mantenga y actualice el *corpus* de contenido, conducta y/o políticas de uso aceptable de la empresa.
- Instituir procesos para tener en cuenta las consideraciones de los usuarios al redactar y actualizar la gobernanza de producto relevante.
- Desarrollar descripciones y explicaciones de políticas orientadas al usuario en un lenguaje fácil de entender.
- Crear mecanismos para incorporar los aportes de la comunidad de usuarios e investigaciones de usuarios a las normas.
- Trabajar con expertos y grupos externos reconocidos de la sociedad civil que ofrezcan aportes sobre las normas.
- Documentar para uso interno la interpretación de las normas y su aplicación con base a precedentes u otras formas de investigación y análisis.
- Facilitar la autorregulación por el usuario o la comunidad cuando corresponda, por ejemplo proporcionando foros para la gestión dirigida por la comunidad o herramientas para su moderación, y encontrar oportunidades para educar a los usuarios sobre las normas, por ejemplo, cuando violen las reglas.

Compromiso 3

Realizar operaciones de cumplimiento para implementar la gestión del producto.

Objetivo: Garantizar que haya operaciones para implementar los objetivos establecidos en la gestión del producto para abordar los riesgos asociados al contenido y la conducta.

Comentario: Las empresas adoptan una variedad de enfoques para hacer cumplir la gestión de productos, porque cada instancia depende de la naturaleza de los servicios digitales proporcionados y refleja las interacciones entre una comunidad de usuarios en particular. Sin embargo, existen algunos puntos en común

de alto nivel. Las empresas deben definir el papel de la función de aplicación dentro de la empresa (en relación con funciones como producto, legales, comunicaciones, negocios, ejecutivo y políticas públicas) y dentro del equipo (con roles funcionales como operaciones, políticas y revisores). Las empresas frecuentemente invierten en una variedad de tecnologías y personal para realizar tareas que incluyen: detectar proactivamente contenido que viola las reglas, permitir que las personas denuncien contenido o conductas violatorias, desarrollar sistemas para administrar la información de los informes recibidos, establecer colas y procesos para que los trabajadores tomen decisiones e implementarlas, y sistemas de aplicación de la ley para disuadir a malos actores o a otras conductas violatorias.

Algunos ejemplos de prácticas que incorporan el compromiso de realizar operaciones de aplicación para implementar la gestión del producto son:

- Asegurar que la empresa cuente con personal e infraestructura tecnológica para gestionar los riesgos de contenido y conducta, para lo cual la empresa podrá:
  - Constituir roles y/o equipos dentro de la empresa que sean responsables de la creación, evaluación, implementación y operación de políticas.
  - Desarrollar y revisar la infraestructura operativa que facilite la clasificación de informes de infracciones y escalaciones para problemas más complejos.
  - Determinar cómo se proporcionarán las herramientas tecnológicas relativas a la confianza y la seguridad (es decir, construir, comprar, adaptar, colaborar).
- Formalizar programas de capacitación y concientización para seguir el ritmo del contenido dinámico en Internet y otros temas relacionados, para nutrir el diseño de soluciones asociadas.
- Invertir en el bienestar y la resiliencia de los equipos que manejan materiales sensibles, como herramientas y procesos para reducir la exposición, capacitación de los empleados, rotaciones de revisión en y de contenido, y beneficios tales como asesoramiento.
- En la medida que sea factible y apropiado, identificar áreas donde se justifique la detección anticipada y potencialmente la intervención.
- Implementar métodos para informar fácilmente que un contenido, una conducta o una cuenta de usuario infringe las normas (como el flujo de informes dentro del producto, formularios que se encuentran fácilmente o una dirección de correo electrónico designada).
- Poner en práctica acciones de aplicación de la ley a escala donde:
  - Se establecen estándares para una respuesta oportuna y una priorización basada en factores como el contexto del producto, la naturaleza, la urgencia y el alcance del daño potencial, la probable eficacia de la intervención y la fuente del informe.
  - Se encuentran disponibles apelaciones a decisiones u otro acceso apropiado a reparación.
  - Se realizan informes adecuados fuera de la empresa, como por ejemplo ante las autoridades policiales en casos de amenaza creíble e inminente a la vida.
- Garantizar que existan procesos relevantes que permitan a los usuarios u otras personas "marcar" o denunciar contenido, conducta o una cuenta de usuario como potencialmente violatoria de las políticas y opciones para aplicar la norma sobre esa base.

- Trabajar con terceros reconocidos (como por ejemplo verificadores de datos calificados u organizaciones de derechos humanos) para identificar respuestas significativas en el cumplimiento de la ley.
- Trabajar con socios de la industria y otros para compartir información útil sobre riesgos, cuando ello sea consistente con las obligaciones legales y las mejores prácticas de seguridad.

Compromiso 4

Evaluar y mejorar los procesos asociados a los riesgos relacionados con contenidos y conductas.

Objetivo: Garantizar que existan mecanismos dentro de la empresa para mantenerse al día y responder a la evolución de los riesgos y enfoques relacionados con el contenido y la conducta.

Comentario: Las empresas practicantes evaluarán la eficacia de su trabajo en la prevención y reducción de riesgos, aplicarán un sistema explicable para analizar la información obtenida a partir de esa evaluación y mejorarán sus procesos para mantenerse al día con la evolución de estos riesgos y otros desarrollos relevantes en ese ámbito.

Las empresas practicantes adoptan métodos para recopilar comentarios sobre cuán eficaz es su enfoque para reducir los riesgos relacionados con el contenido y la conducta, y luego evolucionan su enfoque para mantenerse al día con las lecciones de la experiencia, los desarrollos en el producto y las tendencias en el ámbito.

Algunos ejemplos de prácticas que incorporan un compromiso de evaluar y mejorar periódicamente los procesos asociados con los riesgos relativos al contenido y la conducta son:

- Desarrollar métodos de evaluación para evaluar la precisión de las políticas y operaciones, cambiar las prácticas de los usuarios, los daños emergentes, la eficacia y la mejora de los procesos.
- Establecer procesos para garantizar que las políticas y operaciones estén alineados con estos compromisos.
- Utilizar evaluaciones de riesgos al determinar la asignación de recursos para riesgos emergentes relacionados con el contenido y la conducta.
- Fomentar vías de comunicación entre la empresa practicante, por un lado, y los usuarios y otras partes interesadas (como la sociedad civil y organizaciones de derechos humanos) para actualizar los avances y recopilar comentarios sobre el impacto social del producto y las áreas a mejorar.
- Establecer mecanismos de reparación adecuados para aquellos usuarios que se hayan visto directamente afectados por decisiones de moderación, como la eliminación de contenido, la suspensión o cancelación de cuentas.

Compromiso 5

Garantizar que las políticas de confianza y seguridad adecuada estén disponibles e informar periódicamente al público y a otras partes interesadas sobre las medidas tomadas.

Objetivo: Garantizar que el público y otras partes interesadas tengan conocimiento de los objetivos, desafíos y actividades de confianza y seguridad de la empresa.

Comentario: La transparencia cumple una función clave al informar al público y educar a varias partes interesadas sobre las prácticas de confianza y seguridad de una empresa practicante y, con el tiempo, genera confianza en la capacidad del estándar de cuidado de una industria.

Algunos ejemplos de prácticas que incorporan el compromiso de publicar e informar sobre políticas relevantes de confianza y seguridad pueden incluir:

- Publicar informes periódicos de transparencia que incluyan datos sobre riesgos destacados y prácticas de aplicación relevantes, que pueden cubrir áreas como abusos reportados, procesados y atendidos, y solicitudes de datos procesadas y cumplidas.
- Avisar a los usuarios cuyo contenido o conducta esté en cuestión en una acción de aplicación de la norma (con las excepciones pertinentes, como la prohibición legal o la prevención de daños mayores).
- Registrar quejas recibidas, decisiones y acciones de aplicación de la norma de acuerdo con las políticas de datos relevantes.
- Crear procesos para apoyar a los investigadores académicos y de otro tipo que trabajan en temas relevantes (en la medida en que lo permita la ley vigente y de conformidad con los estándares de seguridad y privacidad pertinentes, así como consideraciones comerciales, tales como los secretos empresariales).
- Cuando corresponda, crear indicadores en el producto que señalen las medidas de aplicación de la ley tomadas, incluido un aviso público amplio (por ejemplo, un ícono que indique el contenido eliminado y brinde ciertos detalles) y actualizaciones para los usuarios que informaron sobre contenido violatorio y acceso a soluciones.

Definiciones

A efectos del Sistema de Mejores Prácticas del DTSP y el Comentario adjunto, se aplican las siguientes definiciones:

Sistema de Mejores Prácticas: hace referencia a este documento.

Compromiso: A los efectos de este Sistema de Mejores Prácticas, las acciones comprometidas por una organización para identificar y abordar el riesgo relacionado con el contenido y la conducta.

Empresas practicantes: Proveedores de productos o servicios que han adoptado los compromisos aquí descritos.

Gestión del producto: el conjunto de acuerdos, reglas y pautas que median en la interacción del usuario con el servicio digital y estructuran la conducta relacionada con el producto (los ejemplos incluyen términos de servicio, política de privacidad, pautas comunitarias, política de contenido, política de uso aceptable, códigos de conducta y cualquier proceso de la empresa mediante el cual se crean, adoptan o repiten estas declaraciones rectoras).

Confianza y seguridad: el campo y las prácticas que gestionan los desafíos relacionados con el riesgo asociado al contenido y la conducta, que incluyen, entre otros, la consideración de la seguridad por diseño, la gestión del producto, la evaluación, detección y respuesta de riesgos, la garantía de calidad, y la transparencia.

Riesgos relacionados con el contenido y la conducta: la posibilidad de ciertos contenidos o comportamientos ilegales, peligrosos o dañinos, incluidos riesgos para los derechos humanos, que están prohibidos por las políticas y términos de servicio pertinentes. (Se entenderá que las referencias a "riesgos" se refieren a riesgos relacionados con el contenido y la conducta).