Digital Trust
& Safety Partnership

Sistema de Melhores Práticas

A Sistema de Melhores Práticas do DTSP 2025 foi atualizada para se alinhar à publicação da Estrutura Segura como ISO/IEC 25389. Esta Sistema de Melhores Práticas é funcionalmente idêntica às cláusulas quatro e cinco da ISO/IEC 25389:2025.

Digital Trust
& Safety Partnership

Introdução

Os serviços digitais são, cada vez mais, centrais em nossas vidas diárias, facilitando o diálogo social, a atividade econômica e muito mais. Tais serviços fornecem ferramentas poderosas para que usuários em todo o mundo se envolvam em uma ampla gama de valiosas atividades online. Porém, como quaisquer ferramentas, também podem ser utilizadas para facilitar comportamentos e conteúdos prejudiciais. A conscientização e a ação contra essa utilização indevida têm crescido nos últimos anos. Isso tem levado a uma urgência crescente na compreensão, no apoio e na avaliação de formas eficazes de reduzir os danos associados ao conteúdo e ao comportamento online. Ao mesmo tempo, protege-se a capacidade das pessoas de se expressarem, realizarem negócios, acessarem informações, associarem-se, trabalharem, estudarem e participarem de suas comunidades por meio de serviços digitais.

Alcançar esse equilíbrio é um desafio considerável. Para começar, não existe uma abordagem única para lidar com os riscos do conteúdo online e de seus comportamentais associados ou, de forma mais geral, para as operações de confiança e segurança das empresas. Dependendo da natureza dos serviços digitais, cada empresapode enfrentar riscos únicos relacionados aos diferentes produtos ou recursos que fornecem - diferentes ameaças, vulnerabilidades e consequências. Os produtos ou recursos podem ser oferecidos direta ou indiretamente aos usuários finais, bem como a outros serviços ou negócios. O que é uma prática eficaz para um serviço digital pode não ser adequada para o outro, e abordagens altamente prescritivas ou rígidas para definir práticas de confiança e segurança são provavelmente demasiado amplas, demasiado restritas ou terão consequências negativas não intencionais. Além disso, os riscos mudam ao longo do tempo e, por isso, as abordagens para mitigá-los também devem ter espaço para evoluírem.

Dada a diversidade dos serviços digitais, é importante definir um sistema geral e estabelecer um conjunto de objetivos para aquilo que constitui uma abordagem responsável, para os quais os serviços digitais poderão, então, mapear as suas práticas específicas. Esta abordagem flexível foi implementada em outras frentes, como a cibersegurança; mas os sistemas existentes não são suficientemente precisos para serem aplicados de maneira genérica quando se trata de abordar comportamentos e conteúdos nocivos.

A *Digital Trust & Safety Partnership* (DTSP) tem o objetivo de cobrir essa necessidade, por meio da documentação e da facilitação na adoção de compromissos abrangentes e amplamente implementados, estabelecidos neste documento, para promover maior transparência e melhor compreensão da confiança e segurança dentro e fora do setor. Para tal fim, o Sistema de Melhores Práticas DTSP oferece um sistema comum sobre como as empresas abordam os riscos relacionados ao conteúdo e à conduta. Embora os compromissos abrangentes sejam uniformes, o método pelo qual eles são cumpridos – seja pela aplicação das práticas ilustrativas neste documento ou outras alternativas – variará de acordo com o produto ou recurso digital e evoluirá com os desafios enfrentados e os avanços realizados no campo da confiança e segurança.

O DTSP considera que os cinco compromissos abrangentes representam medidas necessárias realizadas pelas Empresas praticantes para identificar e abordar conteúdos e condutas prejudiciais, preservando, ao mesmo tempo, a liberdade de expressão e outros direitos, incluindo padrões de direitos humanos internacionalmente reconhecidos e, também, o valor social e econômico dos serviços digitais.

O Sistema de Melhores Práticas DTSP marca a primeira tentativa de articular os esforços atuais das empresas em abordarem riscos relacionados à conduta e ao conteúdo. O Sistema de Práticas é focado

no desenvolvimento, no gerenciamento, na supervisão, na melhoria e na documentação transparente para serviços e produtos digitais. Com o tempo, este Sistema evoluirá e poderá ser acessado de maneira padronizada, tais como em outras áreas como segurança e privacidade.

Compromisso geral:

considerar riscos relacionados ao conteúdo e à conduta nos domínios de desenvolvimento, gerenciamento, fortalecimento e melhoria do produto, e atribuir responsabilidades e recursos para cada domínio.

As Empresas praticantes levam a sério as questões de confiança e segurança, que se demonstra por meio do investimento e do desenvolvimento em pessoal e tecnologias pertinentes; a adoção de direitos com relação aos princípios de confiança e segurança e a considerações sobre desenvolvimento, gerenciamento, supervisão e melhoria de produtos, e documentação pertinentes para serviços e produtos digitais. Os seguintes compromissos caracterizam a abordagem das Empresas praticantes em cada uma dessas áreas.

Compromisso 1:

identificar, avaliar e ajustar-se aos riscos relacionados ao conteúdo e à conduta no desenvolvimento de produtos.

Meta: garantir que as empresas se envolvam em uma adequada antecipação relacionada aos riscos relacionados ao conteúdo e à conduta, e incorporar, de forma adequada, insights sobre os recursos do produto.

Comentário: a antecipação e a redução de riscos, como parte do desenvolvimento do produto, é uma parte importante da função de confiança e segurança. À medida que as equipes de produto projetarem, se relacionarem e refinarem produtos ou recursos para lançamentos, os produtos serão avaliados desde a perspectiva da abordagem de possíveis riscos relacionados ao conteúdo e à conduta. Para tal fim, existe uma gama de mecanismos por meio dos quais um dado produto ou recurso pode ser configurado para garantir sua adequação às considerações de confiança e segurança. As Empresas praticantes (a) geram processos para avaliar os riscos relacionados ao conteúdo e à conduta, ao desenvolverem produtos a serem lançados para o público; (b) buscam prevenir ou mitigar tais Riscos na etapa de desenvolvimento; e (c) continuam a aperfeiçoar o produto no pós-lançamento, de forma adequada, à luz dos riscos observados.

Exemplos de práticas que integram um compromisso para avaliar e ajustar-se aos riscos relacionados ao conteúdo e à conduta no desenvolvimento de produtos, podem incluir:

- Desenvolver capacidades de insight e análise para compreender padrões de abuso e identificar mitigações preventivas que possam ser integradas aos produtos
- Incluir a equipe de confiança e segurança, ou a parte interessada correspondente, no processo de desenvolvimento do produto nas primeiras etapas, inclusive por meio de comunicações e reuniões, solicitando e incorporando feedback conforme apropriado
- Designar uma equipe ou gerente como responsável pela integração do feedback em confiança e segurança
- Avaliar as considerações sobre confiança e segurança dos recursos do produto, equilibrando a usabilidade e a capacidade de resistência ao abuso
- Utilizar equipes internas ou terceirizadas para realizar avaliações de risco para melhor compreender os riscos potenciais
- Fornecer feedback contínuo de pré-lançamento, relacionado a considerações de confiança e segurança
- Fornecer avaliação pós-lançamento pela equipe responsável pelo gerenciamento de riscos e por aqueles responsáveis pelo gerenciamento do produto ou em resposta a incidentes específicos
- Renovar o produto à luz de considerações de confiança e segurança, inclusive com base no feedback dos usuários ou outros efeitos observados, incluindo a garantia de que as perspectivas das comunidades minoritárias e menos representadas tenham representação
- Adotar medidas técnicas adequadas que auxiliem os usuários a controlarem sua própria experiência com o produto, sempre que adequado (tal como bloquear ou silenciar)

Compromisso 2:

adotar processos que possam ser explicados para o gerenciamento de produtos, incluindo a indicação da equipe responsável pela criação de regras e a explicação sobre como as regras evoluem.

Meta: garantir que as regras e princípios que orientam a conduta e o conteúdo do usuário sejam claras, rigorosas e consistentes.

Comentário: o gerenciamento do produto inclui normas internas, externas e processos pelos quais uma empresa promove determinadas atividades e desestimula outras, em relação aos seus produtos. Esta função existe em complemento ao cumprimento e mitigação de riscos relacionados às leis aplicáveis. Uma das efetivações do gerenciamento do produto são os Termos de Serviço de uma empresa (e para empresas com vários produtos, às vezes vários termos) - o contrato entre os usuários e a empresa que estabeleça expectativas e limites subjacentes. Além disso, algumas empresas podem manter regras adicionais que, de forma mais

direta, abordam uma conduta aceitável, muitas vezes em uma linguagem mais simples e direta. Estas podem ser as denominadas regras e diretrizes da comunidade, políticas de uso aceitável ou políticas de conteúdo. Seu desenvolvimento e evolução podem basear-se em comunidades de usuários ou em uma combinação de terceiros interessados com relações variadas com a empresa.

Os exemplos de práticas, que incorporam um compromisso de adotar processos explicáveis para o gerenciamento do produto, podem incluir:

- Estabelecer uma equipe ou função que desenvolva, conserve e atualize o corpo de conteúdos, o conteúdo e/ou as políticas aceitáveis da empresa
- Instituir processos para analisar as considerações dos usuários ao projetar e atualizar o pertinente gerenciamento do produto
- Desenvolver descrições e explicações de políticas voltadas para o usuário, em linguagem fácil de entender
- Criar mecanismos para incorporar a contribuição da comunidade de usuários e das pesquisas de usuários nas normas da política
- Trabalhar com especialistas e grupos de terceiras partes, reconhecidos pela sociedade civil, para obter informações sobre políticas
- Documentar, para uso interno, a interpretação das regras políticas e sua aplicação, com base em antecedentes ou outras formas de investigação, pesquisa e análise
- Facilitar, quando apropriado, a autorregulação do usuário ou da comunidade; por exemplo, fornecendo fóruns para governança liderada pela comunidade ou ferramentas para moderação da comunidade, e encontrar oportunidades para educar os usuários sobre as políticas, por exemplo, quando estes violarem as regras

Compromisso 3:

conduzir operações de conformidade e cumprimento para implementar o gerenciamento do produto.

Meta: garantir a existência de operações que implementem os objetivos estabelecidos no gerenciamento do produto, para abordar riscos relacionados ao conteúdo e à conduta.

Comentário: as empresas adotam diversas abordagens para a aplicação do gerenciamento do produto, já que cada instância depende da natureza dos serviços digitais fornecidos e reflete as interações entre uma determinada comunidade de usuários. No entanto, existem alguns pontos em comum de alto nível. As empresas devem definir o papel da função de supervisão dentro da empresa (em relação a funções tais como produto, jurídicas, comunicações, negócios, área executiva e políticas públicas) e dentro da equipe (com perfis funcionais tais como operações, políticas e revisores). As empresas, de modo frequente, investem em uma gama de tecnologias e pessoal para a realização de tarefas, incluindo: detectar de forma proativa os conteúdos

que violam as normas, permitir que as pessoas reportem conteúdo ou condutas infratoras, desenvolver sistemas de gerenciamento de informações a partir de relatórios recebidos, estabelecer filas e processos para os trabalhadores tomarem decisões e implementá-las, e sistemas de fiscalização para dissuadir maus atores ou comportamentos infratores adicionais.

Os exemplos de práticas, que incorporam um compromisso de conduzir operações de fiscalização para implementar o gerenciamento do produto, podem incluir:

- Garantir que a empresa possua infraestrutura de pessoal e de tecnologia para gerenciar os riscos relacionados à conduta e ao conteúdo, para o qual a empresa poderá:
  - Estabelecer perfis e/ou equipes dentro da empresa, responsáveis pela criação de políticas, avaliações, implementação e operações
  - Desenvolver e revisar uma infraestrutura operacional que facilite a classificação de relatórios de violações e caminhos de escalonamento para questões mais complexas
  - Determinar como serão fornecidas as ferramentas de tecnologia relacionadas à confiança e segurança (ou seja, elaborar, comprar, adaptar e colaborar)
- Formalizar programas de treinamento e conscientização para manter o ritmo com o conteúdo online e assuntos relacionados, para informar o projeto sobre soluções associadas
- Investir no bem-estar e na resiliência das equipes que tratam materiais confidenciais, tais como ferramentas e processos para reduzir exposição, treinamento de funcionários, rotações em revisão de conteúdos e benefícios tais como o aconselhamento
- Sempre que viável e apropriado, identificar áreas onde a detecção antecipada e, potencialmente, a intervenção, sejam garantidas
- Implementar métodos pelos quais o conteúdo, a conduta ou uma conta de usuário possam ser facilmente denunciadas como violadoras em potencial da política (tais como fluxo de relatórios no produto, formulários ou endereço de e-mail atribuído, de fácil localização)
- Operacionalizar ações de supervisão em escala, quando:
  - Os padrões sejam definidos para resposta oportuna e priorização com base em fatores que incluam o contexto do produto, a natureza, a urgência e a extensão do dano potencial, a provável eficácia da intervenção e a fonte do relatório
  - Estejam disponíveis recursos de decisões ou outro acesso apropriado para soluções
  - As denúncias apropriadas sejam feitas fora da empresa, tais como às autoridades, em casos de ameaça admissível e iminente à vida
- Garantir a existência de processos pertinentes que permitam, aos usuários ou outros, "etiquetar" ou reportar conteúdo, conduta ou conta de usuário como potencial violadora da política, e opções de aplicação nessa base
- Trabalhar com terceiras partes reconhecidas (tais como verificadores qualificados de fatos ou grupos de direitos humanos) para identificar respostas substanciais de aplicação
- Trabalhar com parceiros do setor, e outros, para compartilhar informações úteis sobre Riscos, quando consistente com as obrigações legais e as melhores práticas de segurança

Compromisso 4:

avaliar e melhorar processos associados a riscos relacionados ao conteúdo e à conduta.

Meta: garantir que existam mecanismos dentro da empresa para acompanhar e responder à evolução dos riscos e abordagens relacionados ao conteúdo e à conduta.

Comentário: as Empresas praticantes avaliarão a eficácia do seu trabalho na prevenção e mitigação de Riscos, aplicarão um sistema que possa ser explicado para analisar a informação produzida a partir dessa avaliação e melhorarão os seus processos para acompanhar a evolução destes Riscos e outros desenvolvimentos pertinentes no campo.

As Empresas praticantes adotam métodos para reunir feedback sobre a eficácia de sua abordagem para mitigar riscos relacionados a conteúdo e conduta e, a seguir, desenvolvem sua abordagem para acompanhar as lições a partir da experiência, desenvolvimentos do produto e tendências no campo.

Os exemplos de práticas que integram um compromisso para, de modo regular, avaliar e melhorar processos associados aos riscos relacionados ao conteúdo e à conduta, podem incluir:

- Desenvolver métodos de avaliação para qualificar políticas e operações de precisão, modificar práticas do usuário, danos emergentes, eficácia e melhoria de processos
- Estabelecer processos para garantir políticas e operações alinhadas com tais compromissos
- Utilizar avaliações de risco para determinar a alocação de recursos para Riscos emergentes relacionados ao Conteúdo e à Conduta
- Promover caminhos de comunicação entre a Empresa Praticante, por um lado, e os usuários e outras partes interessadas (tais como a sociedade civil e grupos de direitos humanos), por outro, para a atualização sobre os desenvolvimentos e reunir feedback sobre o impacto social do produto e das áreas a serem melhoradas
- Estabelecer mecanismos de remediação adequados para os usuários que tiverem sido diretamente afetados por decisões de moderação, tais como a remoção de conteúdo, suspensão da conta ou rescisão

Compromisso 5:

garantir que as políticas pertinentes de confiança e segurança sejam divulgadas ao público e informar, de forma periódica, ao público e outras partes interessadas, sobre as ações realizadas.

Meta: garantir que o público e outras partes interessadas obtenham insights sobre as metas, desafios e atividades de confiança e segurança da empresa.

Comentário: a transparência desempenha uma função primordial na informação do público e na educação de várias partes interessadas sobre as Práticas de confiança e segurança de uma Empresa Praticante, ao mesmo tempo que constrói confiança ao longo do tempo para a suficiência do padrão de cuidado de uma indústria.

Os exemplos de práticas, que incorporam um compromisso para publicar e informar sobre políticas pertinentes em confiança e segurança, podem incluir:

- Publicar relatórios periódicos de transparência, incluindo dados sobre riscos destacados e ações de enforcement aplicáveis, que podem cobrir áreas tais como abusos informados, processados e resolvidos, e solicitações de dados processadas e atendidas
- Providenciar aviso a usuários cujo conteúdo ou conduta esteja em disputa em um processo de execução (com exceções pertinentes, tais como proibição legal ou prevenção de danos adicionais)
- Registrar reclamações, decisões e ações de fiscalização recebidas, de acordo com as políticas pertinentes de dados
- Criar processos para apoiar pesquisadores acadêmicos e outros que trabalhem em assuntos pertinentes (na medida permitida por lei pertinente e consistente com os padrões apropriados de segurança e privacidade, bem como com considerações comerciais, tais como os segredos comerciais)
- Quando apropriado, criar indicadores, no produto, sobre as ações de supervisão tomadas, incluindo amplo aviso público (por exemplo, ícone indicando conteúdo removido, fornecendo determinados detalhes) e atualizações para usuários que informaram violação de conteúdo e acesso a soluções

Definições

Para objetivos do Sistema de Melhores Práticas DTSP e o Comentário que o acompanha, aplicam-se as seguintes definições:

Sistema de Melhores Práticas: consulte este documento.

Compromisso: para os objetivos do Sistema de Melhores Práticas, constitui-se nas ações realizadas pela a organizações para identificar e ajustar riscos relacionados ao conteúdo e à conduta.

Empresas praticantes: fornecedores de produtos ou serviços que tenham adotado os compromissos aqui descritos.

Gerenciamento de produto: indica o conjunto de acordos, normas e orientações que mediam a interação do usuário com o serviço digital, e estrutura condutas relacionadas ao produto (os exemplos incluem os termos de serviço, políticas de privacidade, orientações à comunidade, políticas de conteúdo, políticas de uso aceitável, códigos de conduta e quaisquer processos da empresa pelos quais tais declarações de gerenciamento são criadas, adotadas ou iteradas).

Confiança e segurança: relaciona-se ao campo e práticas que gerenciam os desafios relacionados aos Riscos Relacionados à Conduta e ao Conteúdo, incluindo, entre outros, considerações de segurança por design, Gerenciamento de Produto, avaliação, detecção e resposta a riscos e garantia de qualidade e transparência.

Riscos relacionados à conduta e ao conteúdo: à possibilidade de conteúdos ou comportamentos ilegais, perigosos ou, por outro lado, prejudiciais, incluindo riscos aos direitos humanos, riscos esses que são proibidos por políticas pertinentes e termos de serviço (as referências a "riscos" podem ser compreendidas ao consultar-se riscos relacionados ao conteúdo e à conduta).