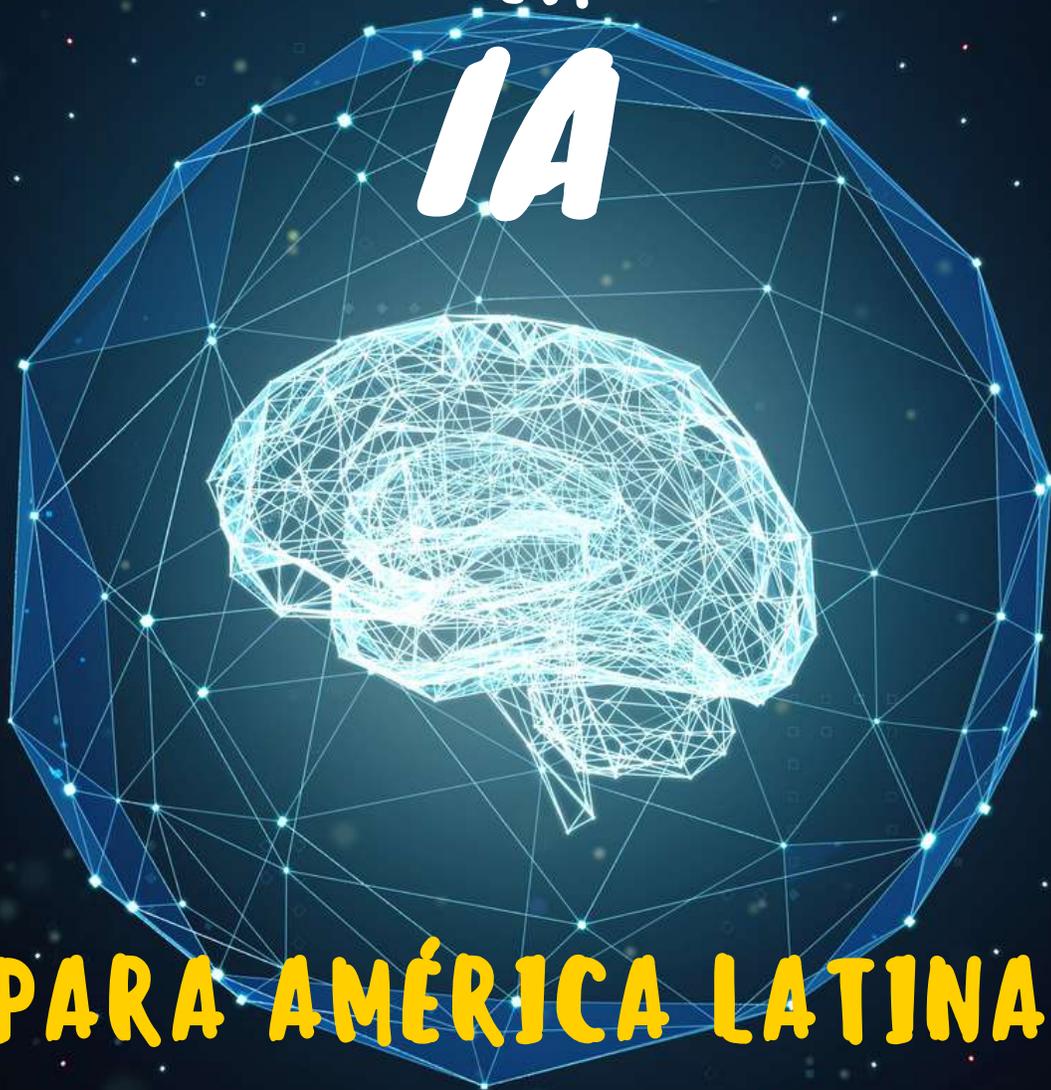


A ÉTICA

EM

IA



PARA AMÉRICA LATINA

Um pequeno guia para start-ups e empresas desenvolvidas por:



Com o apoio e as contribuições de



A ÉTICA EM IA PARA EMPRESAS DA AMÉRICA LATINA

UM PEQUENO GUIA PARA EMPRESAS DE INTELIGÊNCIA ARTIFICIAL NA AMÉRICA LATINA PREPARADO PELA EON RESILIENCE LAB DA C MINDS E PELA META.

Agosto 2022

Autores: Claudia May Del Pozo, Diretora do Eon Resilience Lab da C Minds y Carla Vázquez Wallach, Consultora Jurídica da C Minds.

Principais colaboradores: Norberto de Andrade, Diretor de Política e Governança da Meta, Paula Vargas, Diretora de Privacidade para a América Latina da Meta, Constanza Gómez Mont, Fundadora e Presidente da C Minds, Ana Victoria Martín del Campo Alcocer, ex-Coordenadora de Projetos da C Minds, Daniel Castaño, Socio Fundador em Mokzy.

Colaboradores: Alejandro Cobando da Talov, Alexei Stanislavski da Maat.ai, Andres Felipe Montoya Nieta da Nediari, Antonio Henrique Dianin da Project Company, Cuco Vega da Bexi, Eduardo Farina da Blue Messaging, Esteban Gorupizc da Atexto, Erick Carranza e Miguel Ángeles da IRBin, Erick Estrada da Tooring, Genaro Aldana Chavez da ReMap 4.0, Gimena Olguin da Quick Hit Solutions, Ivan Caballero da Citibeats, José Tomás Arenas da TeleDx, Leticia Ramírez da Drone Domain, Oscar Landivar da Wizdem, Pavel Pichardo da Madison, Pedro Vallejo Castillo da Datlas, Rafael Figueroa da Portal Telemedicina, Rodolfo Rubén Álvarez González da Dyoo, Sebastián García da IDATHA, Sebastián Flores da U-Planner, Valeria Resendez da Xira.



CONTEÚDO

Introdução	3
Contexto	10
Resumo dos princípios éticos da IA	13
Segurança e robustez	17
Equidade, inclusão e não discriminação	23
Privacidade	28
Transparência e explicabilidade	33
Responsabilidade e prestação de contas	38
Anexo	45



INTRODUÇÃO

Segundo a Organização para Cooperação e Desenvolvimento Econômico (OCDE), a inteligência artificial (IA) é descrita como

"um sistema computacional que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões e recomendações ou decisões que influenciam ambientes reais ou virtuais". Os sistemas de IA são projetados para operar com diferentes níveis de autonomia.

Sem dúvida, as soluções tecnológicas baseadas em IA estão transformando a prestação de serviços e têm fomentado a criação de novos produtos no mercado, diversificando as oportunidades de negócios.

Os benefícios do uso da IA são cada vez mais evidentes, impactando tanto a eficiência das empresas quanto a qualidade e confiança dos serviços e produtos que os consumidores recebem. De acordo com a iniciativa fAlr LAC projetada pelo Banco Interamericano de Desenvolvimento e pela C Minds, as primeiras aplicações de IA se concentraram no governo eletrônico e na melhoria dos processos de governança. Todavia, iniciativas baseadas no uso da IA para abordar questões sociais, tais como melhorar o acesso à educação, fornecer melhores serviços de saúde, combater a pobreza ou reduzir a desigualdade, entre outras, começaram a ser desenvolvidas.¹

O estudo "Inteligência Artificial para o Bem Social na América Latina e no Caribe": Visão geral regional e instantâneos de doze países" estimou que a IA poderia trazer até 14% a mais de riqueza para as economias emergentes da América Latina.²

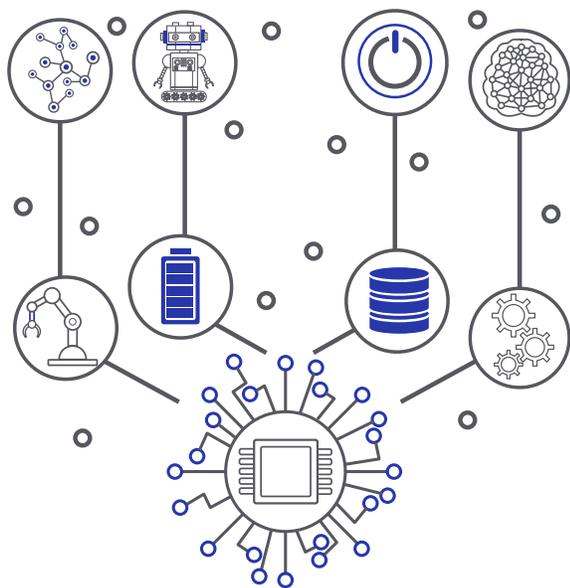
Por sua vez, a OCDE identificou que o processamento massivo de dados através do uso da IA tornou possível enfrentar desafios globais, detectando padrões que fornecem informações estratégicas para os tomadores de decisão, melhorando sua eficiência.³

Entretanto, o uso de IA traz consigo múltiplos desafios, particularmente devido aos riscos potenciais e efeitos adversos para a humanidade decorrentes da opacidade no projeto (ver mais na página 10), da implementação e do uso de sistemas de IA. Em 2019, o Conselho Executivo da Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO) reconheceu que, embora a IA tenha o potencial de transformar o futuro da humanidade para o melhor e de promover o desenvolvimento sustentável, há também uma consciência generalizada dos riscos e desafios que ela traz, especialmente em termos de exacerbar as desigualdades e lacunas existentes, bem como as implicações para os direitos humanos.⁴

Em resposta a essa preocupação, várias organizações e governos têm se preocupado com as considerações éticas da IA, desenvolvendo recomendações, diretrizes, princípios, estudos e relatórios.⁵

Esta preocupação também foi acompanhada pelos setores acadêmico e privado, através do debate e a elaboração de diretrizes e/ou recomendações sobre a implementação de conceitos éticos no uso da IA através de práticas concretas que materializam princípios e valores no ciclo de vida do sistema de IA, tais como a criação de códigos de conduta ou políticas internas sobre a implementação de princípios éticos nos processos de desenvolvimento de produtos e serviços.⁶

Este guia promove uma cultura ética junto a empresários e empresas que projetam, desenvolvem e implementam sistemas de IA na América Latina para: (i) atingir



Uma compreensão e conscientização que lhes permita identificar riscos potenciais e analisar o impacto de suas soluções, (ii) tomar medidas apropriadas e proporcionais de acordo com a magnitude dos riscos identificados, e (iii) quando apropriado, prevenir ou minimizar os riscos. Sob o escopo desta orientação, o uso ético da IA tem um duplo propósito: por um lado, significa colocar esta tecnologia a serviço da humanidade, em benefício da sociedade e para melhorar a vida dos seres humanos⁷ e, por outro lado, reduzir os riscos e impactos negativos não intencionais que o uso da IA pode gerar, seja devido a um projeto, desenvolvimento ou implementação deficiente, seja devido a uma aplicação inadequada.

Ao analisar os impactos dos sistemas de IA, podem surgir tensões entre os diferentes princípios e nem sempre é possível alcançar um equilíbrio rigoroso. As empresas devem, portanto, adotar um processo de reflexão contínua sobre as boas práticas propostas no guia.

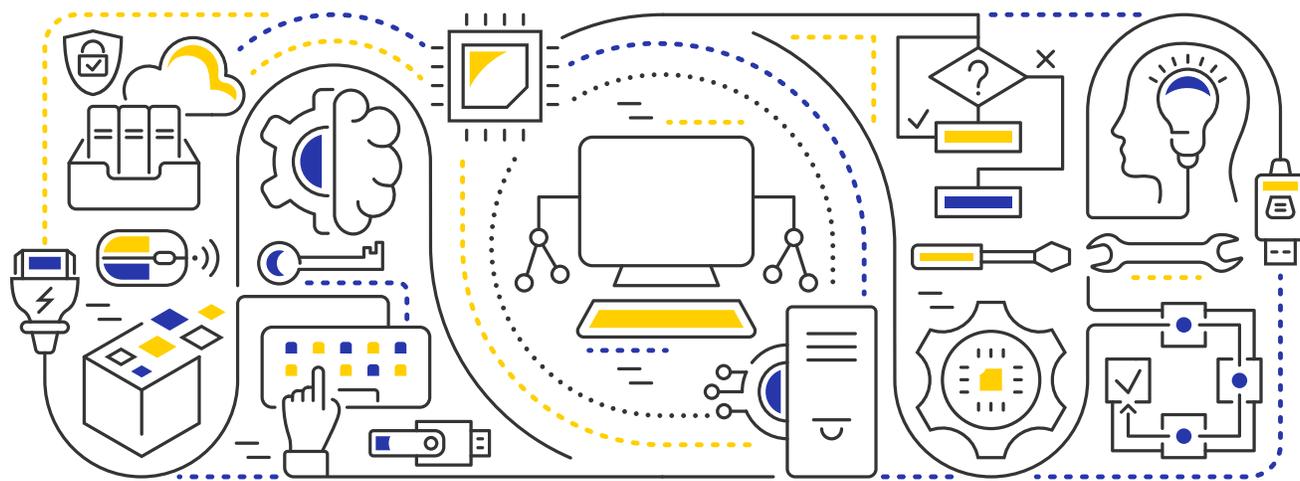
As empresas devem prever e, se necessário, implementar medidas específicas para lidar com possíveis efeitos adversos dos sistemas de IA, por exemplo, projetando avaliações de impacto periódicas, estabelecendo graus de supervisão e controle na tomada de decisões ou reforçando os mecanismos de segurança, tudo isso dependendo do

risco de cada sistema, sua probabilidade de ocorrência e possíveis consequências.

Através de uma abordagem prática, as empresas e os empresários poderão adotar uma cultura ética para usar os sistemas de IA de forma responsável em seu ambiente, particularmente se esses sistemas tiverem, de alguma forma, impacto sobre grupos vulneráveis, tais como setores tradicionalmente desfavorecidos ou discriminados (seja por suas características ou por causa de assimetrias sociais), pessoas com deficiência ou crianças.

A metodologia utilizada para a concepção das boas práticas apresentadas neste guia foi estruturada com base em:

- ➔ A análise de mais de 20 documentos relevantes (ver seção de documentos-chave) sobre ética da IA ao nível internacional;
- ➔ A realização de entrevistas diagnósticas com 23 empresas latino-americanas (ver anexo 1) que utilizam sistemas baseados em IA para aprender sobre suas perspectivas, necessidades, experiências e lições aprendidas.
- ➔ Os princípios da OCDE adotados pelos países membros em 2019⁸, assim como a adoção da Recomendação sobre a Ética da Inteligência Artificial (IA) pelos membros da UNESCO em 22 de novembro de 2021.⁹





A QUEM SE DESTINA ESTE GUIA?

Este guia foi elaborado para empresas e empreendimentos que estão ativamente envolvidos no projeto, desenvolvimento e implementação de um sistema de IA. Inclui todos aqueles que desempenham um papel ativo no ciclo de vida do sistema de IA, tanto indivíduos como pessoas jurídicas.¹⁰

Para maior clareza, os diferentes papéis que podem ocorrer no ciclo de vida de um sistema de IA de acordo com a UNCITRAL¹¹ são descritos abaixo:



Desenvolvedor:

Pessoa responsável pelo projeto teórico de alto nível, programação, treinamento e verificação do sistema de IA, e por sua interface e integração com hardware externo e aplicações externas e fontes de dados antes da implantação;



Fornecedor de dados:

Uma pessoa que fornece dados ou é responsável pelo fornecimento de dados ao sistema (ou seja, os dados necessários para apoiar o treinamento, a implementação ou a operação);



Implementador:

Uma pessoa que implementa o

sistema, integrando-o em suas operações (por exemplo, os bens e serviços que eles fornecem), em particular, configurando, gerenciando, mantendo e apoiando o fornecimento dos dados e infraestrutura necessários para a operação e monitoramento do sistema de IA e sua interação com os dados fornecidos, uma vez implementados;



Operador:

a pessoa que opera o sistema: i) em muitos casos, o operador é a pessoa que implementa o sistema; ii) em alguns casos, o operador pode ser o usuário final dos bens ou serviços com a IA incorporado (por exemplo, se o usuário final tem algum grau de controle sobre a operação desses bens ou serviços); e



Usuário final:

Qualquer outra pessoa afetada pela operação de um sistema de IA, inclusive pela interação com o sistema (por exemplo, ao fornecer dados ao sistema) ou por ser o usuário final de bens ou serviços que incorporam a IA.

O guia será útil para empresas e empreendimentos que atuam como desenvolvedores e implementadores. Cada um desses participantes poderá adotar valores e princípios éticos voluntariamente para criar uma cultura de conscientização e compromisso dentro de sua organização e para a sociedade, com dois objetivos específicos:

- 1 Promover a confiança dos usuários no uso da inteligência artificial, incluindo aqueles que adquirem sistemas de IA para fornecer serviços ou bens a terceiros, e
- 2 Mitigar possíveis riscos e efeitos negativos para a sociedade.



COMO UTILIZAR ESTE GUIA?

Este guia é uma ferramenta para conscientizar as empresas e empreendimentos sobre a importância e a necessidade de implementar princípios éticos no ciclo de vida de um sistema de IA. O grau de adoção das boas práticas contempladas no guia é determinado pela empresa ou empreendimento de acordo com seu compromisso ético com o meio ambiente e com os usuários finais dos bens ou serviços de IA que oferece, bem como com os princípios específicos que se aplicam ao sistema de IA no contexto em que eles são projetados, desenvolvidos e/ou implementados.

Através do guia, as empresas obterão informações claras e simples para compreender o significado dos princípios éticos, bem como boas práticas para criar uma conduta ética no uso de sistemas de IA. Eles também serão capazes de medir seu avanço progressivo, identificando

gradualmente as ações concretas dentro da sua organização e em relação aos seus usuários.

O guia contém 5 princípios éticos que são descritos individualmente em uma seção específica com os seguintes títulos:

O QUE É ISSO?

Esta seção explica de modo geral o Princípio a fim de alcançar uma compreensão clara e simples de sua aplicação em sistemas de IA.



ADOÇÃO GRADUAL

Inclui boas práticas relacionadas com a implementação de cada Princípio, bem como atividades a serem implementadas para que empresas e empreendimentos possam medir seu progresso em relação a suas metas e objetivos.

A empresa ou empreendimento será capaz de implementar as boas práticas contidas neste guia de acordo com seu compromisso com os princípios éticos. São incluídas ações específicas e consecutivas que refletem o maior compromisso da organização com o impacto do sistema de IA em seu ambiente.

Exemplo: :

Equidade, inclusão e não discriminação

Com base neste princípio, o guia recomenda 8 níveis, sendo o 8º nível a ação mais complexa em termos de implementação. As empresas, dependendo de sua capacidade e recursos, decidirão paulatinamente quanto ao progresso na adoção.

- 1** *Revisar a qualidade dos dados (incluindo consistência, diversidade, integralidade, integridade, acessibilidade, precisão e completude).*
- 2** *Conceituar a revisão de dados como sendo um processo transversal contínuo que vai desde o projeto, implementação técnica de bancos de dados, padrões e práticas usadas para armazená-los/modificá-los, bem como protocolos de segurança para acessá-los e compartilhá-los.*
- 3** *Promover a participação ativa de pessoas de diversas origens, independentemente de raça, cor, descendência, idade, sexo, idioma, religião, opinião política, status econômico ou social, na conceituação do projeto do sistema de IA e, quando apropriado, na análise do impacto do sistema de IA em suas comunidades.*
- 4** *Conduzir uma análise de desempenho para diferentes subgrupos, revendo os efeitos causados pelo resultado do processamento de dados, bem como as decisões ou recomendações do sistema de IA.*

VOCÊ QUER SABER MAIS?

Esta seção inclui fontes adicionais de informação para que as empresas e empreendimentos possam aprofundar o conhecimento e aplicação dos princípios éticos, assim como casos de uso e/ou exemplos de aplicação prática desses princípios.



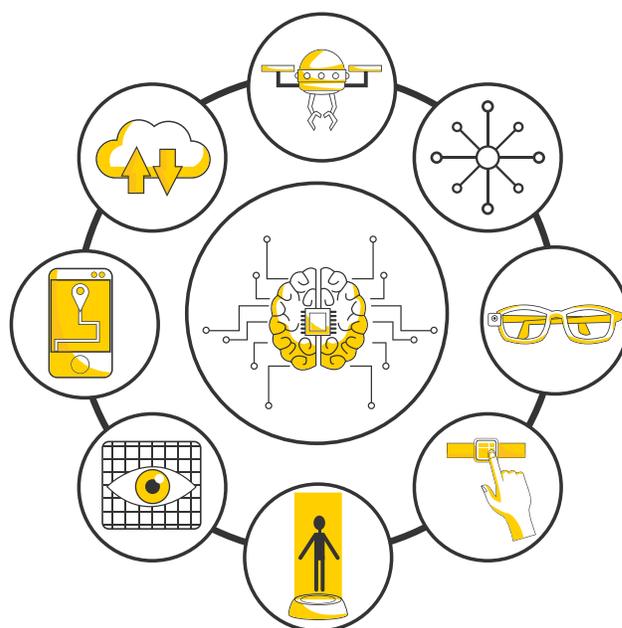
IMPORTANTE

Este guia não fornece conselhos ou recomendações legais sobre o cumprimento do marco legal aplicável aos sistemas de IA. Ele não substitui qualquer regulamentação legal que se aplique ao sistema de IA, à empresa e ao empreendimento, nem cobre requisitos legais que possam ser aplicáveis ao contexto no qual os sistemas de IA serão implementados.

A empresa e seu empreendimento são responsáveis pela análise de viabilidade legal de seu modelo de IA, considerando, por exemplo, o quadro legal aplicável ao caso de uso correspondente, a observância dos direitos humanos, propriedade intelectual, proteção de dados e privacidade, proteção dos direitos do consumidor, requisitos de qualidade ou segurança que possam se aplicar a

produtos, requisitos de validade e legalidade de atos emitidos por instituições públicas, entre outros.

O guia é um documento dinâmico que adotará boas práticas para o projeto, desenvolvimento e à implementação responsáveis de sistemas de IA que sejam identificados no futuro, bem como aqueles que venham a emergir da constante reflexão para assegurar a proteção da dignidade humana, seus direitos e liberdades. Assim, as boas práticas contidas neste documento não devem ser consideradas como recomendações definitivas.





CONTEXTO

O QUE QUEREMOS DIZER COM "INTELIGÊNCIA ARTIFICIAL"?

Atualmente não há consenso sobre o significado de IA. Entretanto, para os fins deste guia, a definição de Inteligência Artificial fornecida pela OCDE é adotada, pois propõe uma definição pragmática baseada nos diferentes estágios do ciclo de vida de um sistema de IA.

A OCDE define IA como:

"uma máquina que pode, de acordo com um conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou tomar decisões que têm influência em ambientes reais ou virtuais. Os sistemas baseados em IA são projetados para operar com diferentes níveis de autonomia".



QUAL É O USO ÉTICO DA IA?

Inovações em diferentes áreas da vida moderna têm facilitado nossa vida diária através do uso da IA, como no transporte, comunicação, medicina, educação, ciência, vida financeira, direito, serviços de entretenimento, entre outros. No entanto, o uso de IA trouxe consigo debates sobre os desafios éticos, desde o desaparecimento dos empregos tradicionais, reproduzindo e reforçando os preconceitos existentes, criando desigualdades mais profundas, preconceitos ou estereótipos, responsabilidade por possíveis danos físicos ou psicológicos aos seres humanos, até a desumanização geral das relações humanas e da sociedade em geral.

As empresas e os empreendimentos que desenvolvem ou implementam sistemas de IA devem ser sensibilizados aos impactos sobre o meio ambiente ou sobre a vida das pessoas. Devem também maximizar os benefícios da IA, eliminando preconceitos ou discriminação que não foram identificados previamente.



QUATRO RAZÕES PARA ADOPTAR PRINCÍPIOS ÉTICOS NO USO DA IA

1

Uma oportunidade de construir a confiança da sociedade na inovação e evolução tecnológica:

Ao adquirir um produto ou receber um serviço, os usuários geralmente têm a expectativa de que seus direitos serão respeitados e que suas liberdades não serão comprometidas.

Entretanto, os sistemas de IA podem, mesmo não intencionalmente, ter impactos adversos sobre os indivíduos ou a sociedade como um todo. Quando empresas e empreendimentos adotam uma cultura ética de IA, os usuários confiam no uso da tecnologia, não apenas devido aos benefícios diretos do produto ou serviço, mas porque percebem o compromisso genuíno da empresa em tomar as precauções necessárias para mitigar os riscos e efeitos negativos sobre eles.

2

Uma maneira de garantir a sustentabilidade da empresa:

Todas as empresas enfrentam diferentes desafios durante o processo de

amadurecimento de seu modelo comercial.

Entretanto, para aqueles que adotaram ou adotarão o uso da IA, sua permanência dependerá, em grande parte, da consciência social que tiverem com relação aos possíveis impactos de seus produtos ou serviços na sociedade, pois os usuários agora não apenas consideram a reputação, mas também avaliam o que uma marca diz, faz e representa.

As empresas que são capazes de se conectar com seus consumidores, criando conexões profundas com seus valores, têm melhor desempenho em termos de rentabilidade.

3

Uma vantagem competitiva e uma porta de entrada para novos mercados:

A adoção estrutural e voluntária de valores e princípios éticos no uso da IA se traduz na criação de vantagens competitivas para as empresas, pois os usuários vão querer interagir com produtos ou serviços que se destacam no mercado por serem responsáveis, confiáveis e seguros em relação àqueles que não o são. No relatório da MIT Technology Review Insights (2018) *Humans + Bots: Tension and Opportunity*, é mencionado que:

"as mudanças de produtividade geradas pela IA e pelo aprendizado de máquinas se

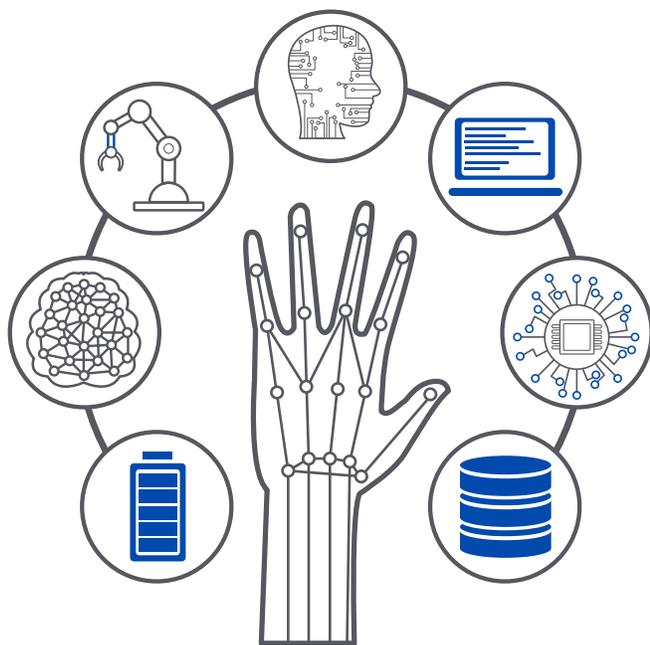
traduzem rapidamente em poder responder melhor às diversas demandas dos clientes e atender consistentemente às suas expectativas em relação à marca.

Mesmo as empresas que adotaram uma abordagem ética, por exemplo, no tratamento de dados, vêem isso não apenas como um requisito para manter uma posição relevante no mercado, mas como uma necessidade para a sociedade como um todo.



Melhora o posicionamento da empresa para a obtenção de financiamento:

As empresas e os empreendimentos frequentemente enfrentam desafios econômicos e financeiros nos estágios iniciais de seu desenvolvimento. É por isso que as políticas públicas, particularmente



na América Latina, visam fomentar a criação, a estabilização e a expansão das empresas.

De forma coordenada, os setores público e privado desenvolveram redes de apoio ao empreendedorismo inovador, tais como grupos de investidores anjos, incubadoras, aceleradores ou mecanismos de financiamento colaborativo.

Particularmente na América Latina, as redes de financiamento têm se concentrado naqueles empreendimentos que buscam resolver problemas locais ou regionais, por exemplo, melhorar a eficiência na prestação de serviços públicos, reduzir a desigualdade social, enfrentar os impactos ambientais, evitar a discriminação contra grupos vulneráveis, promover o acesso ao financiamento, desenvolver capacidades educacionais, entre outros.

Na avaliação de projetos de base tecnológica, particularmente aqueles que utilizam IA, não apenas o potencial econômico é considerado, mas também a adoção de uma cultura empresarial ética para criar valor na indústria ou setor com o qual interage.



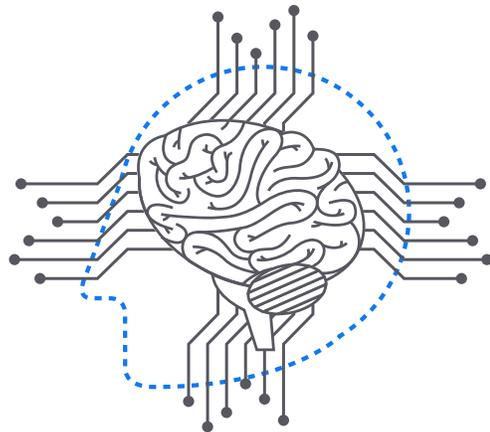
RESUMO DOS PRINCÍPIOS ÉTICOS DA IA

O uso ético da IA baseia-se na vontade das empresas e empreendimentos de adotar uma cultura de lidar com os impactos que tais sistemas têm sobre seu ambiente. Os princípios éticos encorajam as empresas a antecipar e implementar as mudanças necessárias para assegurar que o uso e aplicação da IA sejam feitos de forma responsável e ética.

As boas práticas contidas neste documento propõem plantar uma cultura de responsabilidade através da conscientização dos princípios éticos para os atores envolvidos na concepção, desenvolvimento, implementação, monitoramento, supervisão e avaliação dos sistemas de IA. }

Os princípios éticos, ao contrário das normas obrigatórias, não devem ser interpretados de forma restrita, ou seja, sua adoção é voluntária e deve ser analisada de acordo com os casos específicos de uso, bem como seu contexto. Eles são parte de um processo iterativo para garantir que o sistema de IA funcione de acordo com os resultados esperados e atenda o equilíbrio de interesses envolvidos (por exemplo, o interesse da empresa em desenvolver um sistema com alto potencial de

crescimento que processe dados sobre o comportamento dos usuários em plataformas de entretenimento - vídeo ou música - em contrapartida ao direito dos usuários à privacidade e a expectativa de que o sistema só trate dados indispensáveis para o uso dessas plataformas).



Empresas e empreendimentos, com base em sua capacidade e recursos, bem como em seu nível de comprometimento com os impactos do uso da IA, decidem voluntariamente avançar progressivamente em direção à adoção dos princípios éticos da IA.

O guia inclui cinco princípios baseados nas três mais recentes construções globais de ética da IA, as desenvolvidas pela UNESCO, pela União Europeia, assim como os princípios adotados pela OCDE.

PRINCÍPIO	EXPLICAÇÃO
 <p>SEGURANÇA E ROBUSTEZ</p>	<p>Como qualquer sistema, os sistemas de IA devem assegurar que o sistema seja projetado, desenvolvido e implementado de forma segura para evitar vulnerabilidades.</p> <p>Com base nos resultados da avaliação de risco de um sistema de IA, serão necessários padrões de segurança robustos e avaliações periódicas de desempenho para garantir que esses sistemas se comportem com segurança e funcionem como esperado, mesmo quando sofram um ataque em alguma vulnerabilidade.</p>
 <p>EQUIDADE, INCLUSÃO E NÃO DISCRIMINAÇÃO</p>	<p>Equidade, inclusão e não discriminação significa assegurar uma distribuição justa e igualitária de benefícios e custos, e assegurar que indivíduos e grupos não sofram preconceitos, nem sejam discriminados ou estigmatizados, sempre buscando que o sistema de IA trate todos os usuários de forma justa.</p> <p>Os agentes de IA devem desenvolver e implementar seus sistemas de forma a minimizar os danos ao usuário, evitando reforçar ou perpetuar vieses inadequados durante todo o ciclo de vida dos sistemas de IA.</p> <p>Particularmente nos casos que houver danos substanciais, uma solução eficaz contra a determinação e a discriminação algorítmica injusta deve ser implementada.</p>



PRIVACIDADE

Os atores envolvidos no ciclo de vida de um sistema de IA devem fomentar ou desenvolver uma cultura de gerenciamento de dados, englobando não apenas a observância da proteção dos dados pessoais e privacidade dos indivíduos, mas também um comportamento responsável no uso dos dados.



TRANSPARÊNCIA E EXPLICABILIDADE

A transparência visa fornecer informações adequadas aos usuários finais para permitir a compreensão do sistema de IA, buscando criar confiança no uso desses sistemas, bem como para informar os mecanismos de controle disponíveis sobre as decisões que os impactam. Os indivíduos têm o direito de saber quando uma decisão é tomada com base em algoritmos de IA e, em tais circunstâncias, ter mecanismos claros e simples para solicitar explicações e informações sobre ela. A explicabilidade pretende tornar os resultados dos sistemas de IA inteligíveis e fornecer informações sobre eles.



RESPONSABILIDADE E PRESTAÇÃO DE CONTAS

Embora um sistema de IA seja projetado para operar de forma autônoma, os humanos têm um papel relevante em relação ao desenvolvimento, implementação e uso do sistema. Sob este princípio identifica-se quem são os diferentes atores, dependendo de seu papel, que serão responsáveis pelo bom funcionamento do sistema de IA, particularmente se houver uma violação dos direitos humanos e das liberdades fundamentais dos usuários.

Para atender este princípio, é necessário aplicar procedimentos ou metodologias de avaliação de risco durante o ciclo de vida do sistema de IA, integrar recursos eficazes contra danos, manter uma ampla documentação relacionada ao desenvolvimento, teste e implementação de sistemas de IA e tomar medidas para prevenir ou minimizar danos potenciais.

A glowing blue brain is the central focus, surrounded by a network of white lines and dots. The background is a dark blue gradient with scattered light blue particles. The text is written in a bold, yellow, sans-serif font.

**IMPLEMENTAÇÃO
GRADUAL DE BOAS
PRÁTICAS**

SEGURANÇA E ROBUSTEZ





SEGURANÇA E ROBUSTEZ

O QUE É ISSO?

Como qualquer outro sistema, os sistemas de IA também devem adotar medidas de segurança para lidar com possíveis vulnerabilidades (seja referente aos dados, ao modelo ou à infraestrutura). Dependendo do nível de maturidade do sistema, um conjunto de capacidades-chave será adotado para promover sistemas de IA sustentáveis a longo prazo.

Para atender este princípio, uma abordagem de prevenção de risco deve ser desenvolvida para garantir que o sistema continue funcionando para cumprir o propósito para o qual foi projetado. Dependendo do resultado dessa análise, medidas de segurança adicionais ou mais robustas precisarão ser adotadas para assegurar que esses

sistemas se comportem de forma segura e funcionem como esperado, mesmo se vierem a sofrer um ataque ou a ser identificada vulnerabilidade.

Este princípio deve ser observado durante todo o ciclo de vida do sistema de IA, para monitorar quaisquer mudanças significativas no ambiente com o qual o sistema interage que possam exigir - por exemplo - a inclusão de novos dados no treinamento do modelo. A robustez e segurança de um sistema visam proteger tanto a empresa quanto seus usuários (diretos e indiretos), e seu cuidado evitará que a empresa incorra em custos adicionais para mitigar dívidas técnicas ou reparar danos, evitando, por sua vez, uma possível queda na confiança do usuário no sistema.





ADOÇÃO GRADUAL

BOAS PRÁTICAS

1

Destinar recursos para educar e monitorar a adoção e integração das melhores práticas de segurança informática.

2

Implementar uma política de controle de acesso gradual aos componentes do sistema.

3

Ter as ferramentas certas para o desenvolvimento ordenado e de qualidade, por exemplo, repositórios de código com controle de versão, mecanismos de revisão de código, documentação suficiente sobre seleção de dados, como o modelo foi desenvolvido, assim como os testes que foram executados antes da implementação.

4

Implementar as melhores práticas de gerenciamento de bancos de dados, por exemplo, isolar bancos de dados de qualquer interação com usuários ou terceiros (usando APIs ou camadas de software para impedir a manipulação direta), criptografia no nível do sistema de arquivos (por exemplo, Criptografia Transparente de Dados), bem como higienização de entrada para evitar injeções de comandos.

5

Integrar uma equipe de Garantia de Qualidade (GQ) no processo de desenvolvimento e implantação de produtos/serviços.

6

Construir software seguro desde o projeto, adotando mecanismos de segurança como certificados, criptografia de qualquer transmissão de dados e melhores práticas de segurança para reduzir vulnerabilidades.

7

Incorporar metodologias robustas de desenvolvimento de software (como o Desenvolvimento Orientado a Testes (TDD) ou Programação Robusta).

8

Ter um modelo de OSI (Open Systems Interconnection) para implementar estratégias e mecanismos avançados de segurança.

9

Possuir mecanismos de proteção contra ataques.

10

Monitorar o desempenho do sistema de IA através de métricas e sistemas de alerta automático para alcançar maior eficiência de engenharia (sustentabilidade do modelo).

11

Realizar análises de desempenho cobrindo longos períodos de tempo, a fim de detectar a degradação da estabilidade do algoritmo ou a necessidade de reengenharia do modelo ou de estabelecer mecanismos para aumentar a agilidade do modelo.

12

Definir indicadores de desempenho para avaliar o modelo, por exemplo, resultados sobre o comportamento do algoritmo ou sobre o desempenho do sistema.

13

Implementar mecanismos para alertar a equipe quando esses indicadores se comportam de forma anormal.

14

Criar um protocolo de resposta de emergência baseado em IA para emergências de segurança e/ou mau funcionamento do sistema.

15

Integrar as equipes DevOps e SecOps.

VOCÊ QUER SABER MAIS?

O Banco Interamericano de Desenvolvimento, em conjunto com a Organização dos Estados Americanos, através de seu Observatório de Cibersegurança, publica o relatório sobre Cibersegurança na América Latina:

- Relatório Cibersegurança 2020, "Risks, progress and the way forward in Latin America and the Caribbean", <https://publications.iadb.org/publications/spanish/document/Reporte-Ciberseguridad-2020-riesgos-avances-y-el-camino-a-seguir-en-America-Latina-y-el-Caribe.pdf>

A Comissão Econômica para a América Latina e o Caribe (CEPAL) publicou uma análise da relação entre cibersegurança e governança corporativa no contexto latino-americano:

- CEPAL, "Cybersecurity and the role of the Board of Directors in Latin America and the Caribbean", Héctor J. Lehuedé, setembro de 2020, https://repositorio.cepal.org/bitstream/handle/11362/45988/S2000552_en.pdf?sequence=4&isAllowed=y

A Organização dos Estados Americanos em conjunto com a AWS publicou uma ferramenta que permite gerenciar riscos

cibersegurança de forma flexível e adaptável para qualquer organização, independentemente de seu tamanho ou indústria, em que propõe três estratégias para seu uso: (i) revisão básica das práticas de cibersegurança; (ii) criação ou aperfeiçoamento de um programa de cibersegurança; (iii) comunicação dos requisitos de cibersegurança às partes interessadas.

- A Organização dos Estados Americanos em conjunto com AWS publicou o documento "CYBERSECURITY - NIST FRAMEWORK - A Comprehensive Approach to Cybersecurity", 2019, <https://www.oas.org/es/sms/cicte/docs/OEA-AWS-Marco-NIST-de-Ciberseguridad-ESP.pdf>



O Institute of Electrical and Electronics Engineers (IEEE) também desenvolveu vários padrões sobre sistemas de IA que contribuíram para a aplicação prática dos princípios e estruturas de segurança.

- IEEE, "Artificial Intelligence Systems (AIS) Related Standards", <https://standards.ieee.org/initiatives/artificial-intelligence-systems/standards.html>
- IEEE 7000™-2021 - IEEE Standard Model Process for Addressing Ethical Concerns During System Design, https://engagestandards.ieee.org/ieee-7000-2021-for-systems-design-ethical-concerns.html?utm_source=ieeesa&utm_medium=ae&utm_campaign=ais-2021
- IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems, <https://standards.ieee.org/project/7009.html>.

Da mesma forma, alguns estudos e análises sobre uma IA segura são mencionados abaixo.

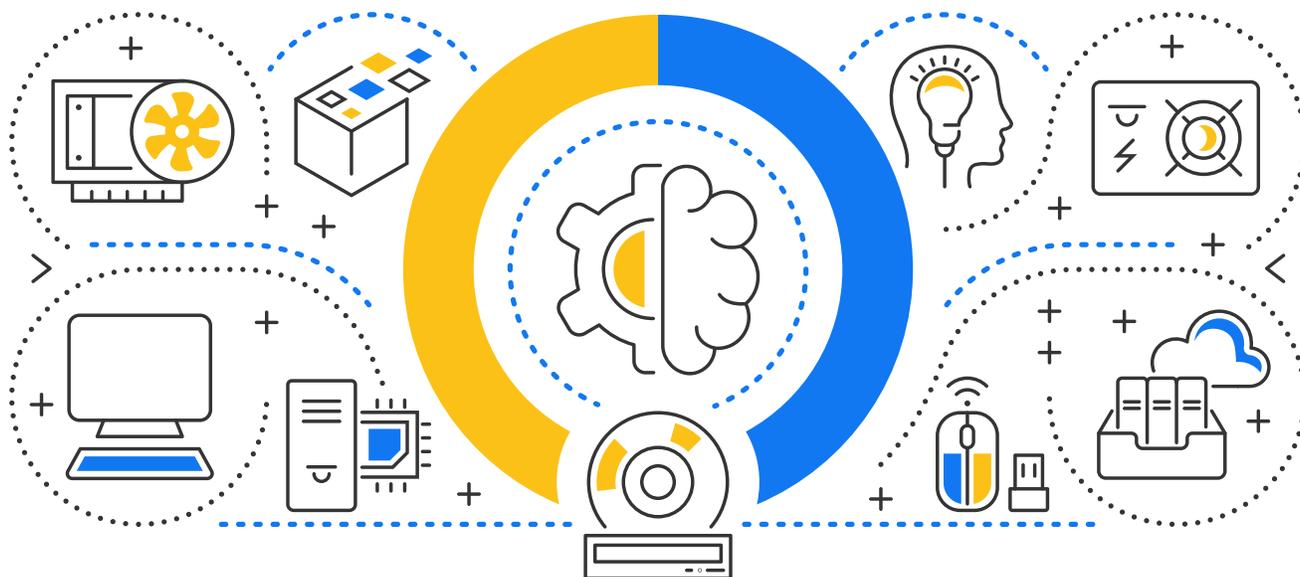
- CSET, "CENTER for SECURITY and EMERGING TECHNOLOGY", Tim G.J. Rudner and Helen Toner, "Key Concepts in AI Safety: Robustness and Adversarial Examples", março 2021,

<https://cset.georgetown.edu/publication/key-concepts-in-ai-safety-robustness-and-adversarial-examples/>

- European AI Alliance, "Refining Technical Robustness and Safety Questions", Shahar Avin, Novembro 2019, <https://futurium.ec.europa.eu/en/european-ai-alliance/best-practices/refining-technical-robustness-and-safety-questions>

QUIDADE, INCLUSÃO E NÃO DISCRIMINAÇÃO





EQUIDADE, INCLUSÃO E NÃO DISCRIMINAÇÃO

O QUE É ISSO?

A equidade é um princípio intimamente ligado à inclusão e à não discriminação, e significa igualdade de tratamento e de oportunidades. Ela procura assegurar que todos que interagem com o sistema de IA sejam tratados de forma justa.

Os usuários do sistema não devem sofrer preconceitos, estigmas ou estereótipos com base na raça, cor, sexo, idioma, religião, opinião política ou outra convicção, origem nacional ou social, propriedade, nascimento ou outro status.

O sistema de IA, através de projetos inclusivos, poderia promover a acessibilidade dos usuários a certos bens

ou serviços, bem como sua participação. O princípio de equidade, inclusão e não-discriminação pode ser adotado em três níveis: (i) na concepção do sistema; (ii) nas políticas organizacionais internas; e (iii) na implementação. A adoção deste princípio facilitará aos programadores de IA identificar se um algoritmo tomará uma decisão que poderá representar um viés discriminatório.



ADOÇÃO GRADUAL

BOAS PRÁTICAS

1

Revisar a qualidade dos dados (incluindo consistência, diversidade, integridade, integridade, acessibilidade, precisão e completude).

2

Conceituar a revisão de dados como um processo transversal contínuo que vai desde o projeto, implementação técnica de bancos de dados, normas e práticas usadas para armazená-los/modificá-los, bem como o projeto de protocolos de segurança.

3

Promover a participação ativa de pessoas de diversas origens, independentemente de raça, cor, descendência, idade, sexo, idioma, religião, opinião política, status econômico ou social, na conceituação do projeto do sistema de IA e, quando apropriado, na análise do impacto do sistema de IA em diferentes comunidades afetadas.

4

Realizar uma análise de desempenho para diferentes subgrupos, revendo os efeitos causados pelo resultado do tratamento de dados.

5

Comparar como as métricas de treinamento e avaliação do sistema se comportam em diferentes subgrupos da população identificados como usuários e usuários em potencial. Por exemplo, analisar a taxa de falsos positivos e falsos negativos por subgrupo pode ajudar a empresa a identificar subgrupos que experimentam um desempenho desproporcionalmente melhor ou pior.

6

Integrar processos de mitigação de vieses inconscientes. As ferramentas de avaliação de vieses inconscientes podem ajudar a identificar riscos potenciais de vieses ou até mesmo mitigar aqueles existentes.

7

Corrigir o viés onde ele ocorre através da reciclagem ou enriquecimento de bancos de dados.

Quando padrões discriminatórios originários de um banco de dados não representativo são detectados, a equipe técnica deve ter a capacidade de requalificar o sistema usando mais e/ou melhores dados.

8

Criar um teste que teste o sistema para situações limite ou cenários onde se prevê que o software possa se comportar de forma inesperada ou fora do projeto.



VOCÊ QUER SABER MAIS?

Há uma série de ferramentas que ajudam a verificar a diversidade de dados, bem como portais de dados abertos de vários países em nível nacional e local. Estas ferramentas também ajudam a comparar os fatores que levam um algoritmo a tomar uma decisão ao invés de outra.

- OECD, "Tools for Trustworthy AI", A framework to compare implementation tools for trustworthy AI Systems", junho 2021, <https://www.oecd-ilibrary.org/docserver/008232ec-en.pdf?expires=1644115818&id=id&accname=guest&checksum=1BDB568CCFF115511CD8B68B2BFA7E09>
- Banco Interamericano de Desarrollo, fAlr LAC, "Ethical Self-Assessment", <https://fairlac.iadb.org/es/emprendimiento>
- Google, "Visually probe the behavior of trained machine learning models, with minimal coding", <https://pair-code.github.io/what-if-tool/>.
- META AI, "How we're using Fairness Flow to help build AI that works better for everyone", 2021, <https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/>

- META AI, "How we're using Fairness Flow to help build AI that works better for everyone", 2021, <https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/>



- META AI, desenvolveu uma série de recursos sobre equidade: <https://ai.facebook.com/blog/what-ai-fairness-in-practice-looks-like-at-facebook/>
- Microsoft, "AI Fairness Checklist", 2020, <https://www.microsoft.com/en-us/research/project/ai-fairness-checklist/>.
- ALTAI, "The Assessment List on Trustworthy Artificial Intelligence", <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>
- IBM, "AI Fairness 360", <https://aif360.mybluemix.net/>.

- IEEE, "IEEE P2863™ - Recommended Practice for Organizational Governance of Artificial Intelligence", 2020, <https://standards.ieee.org/project/2863.html>.
- IEEE, "IEEE P7003™ - Standard for Algorithmic Bias Considerations", <https://standards.ieee.org/project/7003.html>.

CASOS DE USO:

- Center for Applied AI at Chicago Booth, "ALGORITHMIC BIAS PLAYBOOK", https://www.ftc.gov/system/files/documents/public_events/1582978/algorithmic-bias-playbook.pdf

PRIVACIDADE



PRIVACIDADE

O QUE É ISSO?

Este princípio não pretende substituir ou esgotar as obrigações que as empresas e empreendimentos já têm que cumprir sob os marcos legais nacionais, regionais e internacionais, aplicáveis na área de privacidade e proteção de dados.

O princípio refere-se às empresas que integram no projeto, desenvolvimento e implementação do sistema de IA um compromisso ético com o uso de dados e o respeito à privacidade das pessoas que utilizam esses sistemas. O princípio prevê a adequada coleta de dados e sua proteção, e um uso aceitável e apropriado durante todo o ciclo de vida dos dados (coleta, gerenciamento e utilização). A privacidade pode ser reforçada pela sua integração no projeto do sistema através de Tecnologias de Melhoria da Privacidade (Privacy

Enhancing Technologies, em inglês), que são tecnologias que incorporam os princípios fundamentais da proteção de dados, minimizando o uso de dados pessoais, maximizando a segurança dos dados e empoderando as pessoas.

Quando os usuários interagem com um sistema de IA, eles depositam sua confiança na expectativa de que seus dados serão utilizados com integridade e honestidade. Essa expectativa é baseada na explicação que lhes é fornecida sobre como seus dados serão utilizados e para quais fins. As empresas devem, portanto, equilibrar seus interesses comerciais com o tratamento cuidadoso dos dados das pessoas. Por tratamento cuidadoso, entendemos padrões de diligência na busca do bem-estar da pessoa em questão e respeito por sua dignidade.





ADOÇÃO GRADUAL

BOAS PRÁTICAS

1

Reduzir, se aplicável, a coleta de dados ao absolutamente necessário para os objetivos do sistema de IA e armazená-los apenas pelo tempo que for necessário. Especialmente se trabalhar com dados sensíveis, tais como dados biométricos ou psicométricos.

2

Implementar técnicas de anonimização e pseudoanonimização de dados.

3

Assegurar uma compreensão completa da estratégia de dados na capacidade da pessoa encarregada de supervisionar o tratamento e a qualidade dos dados.

4

Implementar uma política onde os usuários tenham o maior conhecimento possível sobre a maneira em que seus dados são usados por um sistema de IA (que dados um sistema de IA pode usar para processar e como os dados devem ser usados).

5

Quando apropriado, integrar ferramentas que permitam ao usuário escolher os dados que deseja fornecer, de acordo com suas preferências. Dependendo do risco e do impacto do sistema de IA, esclarecer que, considerando os dados fornecidos, a qualidade do serviço ou produto pode mudar.

6

Estudar e conhecer o escopo da política de privacidade de terceiros (fornecedores, colaboradores, aliados).

7

Adotar a prática de "privacidade por projeto", o que implica levar em consideração o princípio de privacidade desde o projeto e durante todo o processo de desenvolvimento, engenharia e implantação, integrando a questão da privacidade na própria tecnologia.

8

Realizar avaliações de risco, considerando a avaliação do impacto na privacidade, para implementar sistemas eficazes para gerenciar riscos e controles internos.



VOCÊ QUER SABER MAIS?

De acordo com a Organização dos Estados Americanos (OEA), a maioria dos países membros garante o respeito e a proteção dos dados pessoais como um direito distinto e complementar aos direitos à privacidade, dignidade pessoal e honra familiar, a inviolabilidade das comunicações domésticas e privadas e conceitos correlatos.

A OEA atualizou os princípios sobre privacidade e proteção de dados pessoais, explicando seu escopo não apenas do ponto de vista obrigatório, mas também do ponto de vista da atitude que seria esperada do controlador dos dados.

- Organização dos Estados Americanos, "Updated Principles of the Inter-American Juridical Committee on Privacy and Personal Data Protection, with annotations", 2021

http://www.oas.org/es/sla/cji/docs/CJI-doc_638-21.pdf.

Em 2019, a OEA emitiu o documento "Classificação dos Dados", fornecendo ferramentas para que as organizações pensem em dados, com base na sensibilidade e no impacto comercial, o que ajuda a organização a avaliar os riscos associados a diferentes tipos de dados.

- Organização dos Estados Americanos & AWS, "Data Classification," 2019, <https://www.oas.org/es/sms/cicte/docs/ESP-Clasificacion-de-Datos.pdf>.

A META AI disponibilizou a OPACUS, uma ferramenta de código aberto para treinamento de modelos de aprendizagem automatizada, com privacidade diferencial para ajudar a avançar o estado da arte e melhorar a privacidade da IA:

- META AI, "Introducing Opacus: A high-speed library for training PyTorch models with diferencial privacy", 20201, <https://ai.facebook.com/blog/introducing-opacus-a-high-speed-library-for-training-pytorch-models-with-differential-privacy/>

O IEEE desenvolveu normas relacionadas ao ciclo de vida dos dados nos sistemas de IA:

- IEEE, "IEEE P2807.1™ - Standard for Technical Requirements and Evaluation of Knowledge Graphs", https://standards.ieee.org/project/2807_1.html.
- IEEE, "IEEE P7002™ - Standard for Data Privacy Process", <https://standards.ieee.org/project/7002.html>.
- IEEE, "IEEE P7005™ - Standard for Transparent Employer Data Governance", <https://standards.ieee.org/project/7005.html>.
- IEEE, "IEEE P7006™ - Standard for Personal Data Artificial Intelligence (AI) Agent", <https://standards.ieee.org/project/7006.html>.

A Rede Ibero-Americana de Proteção de Dados emitiu recomendações sob uma

abordagem preventiva para orientar os desenvolvedores sobre as exigências no processamento de dados pessoais desde o projeto do produto:

- Ibero-American Data Protection Network, "General Recommendations for Data Processing in Artificial Intelligence", 2019, <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf>.
- Ibero-American Data Protection Network, 'Specific guidelines for compliance with the principles and rights that govern the protection of personal data in artificial intelligence projects', 2019, <https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf>.

FAIRsFAIR - fornece soluções práticas sobre o uso de dados através da pesquisa do ciclo de vida dos dados:

- FAIRsFAIR, "Fostering Fair Data Practices in Europe", <https://www.fairsfair.eu/tools-software>

TRANSPARÊNCIA E EXPLICABILIDADE





TRANSPARÊNCIA E EXPLICABILIDADE

O QUE É ISSO?

Transparência e explicabilidade (T&E) não se destinam a dar pleno acesso aos códigos e modelos das empresas, pois isso poderia afetar suas vantagens competitivas e não forneceria informações claras sobre como os sistemas funcionam. Em vez disso, a T&E serve, entre outras coisas, para entender melhor o escopo da tomada de decisão automatizada e as razões de uma decisão específica, bem como para melhorar o comportamento futuro dos sistemas.

- **Transparência** se refere à prática de tornar os processos dos sistemas de IA visíveis para as partes interessadas, sem revelar segredos comerciais.

- **Explicabilidade** se refere a ter a mecânica interna de um sistema de decisão automatizado explicada em termos humanos - conhecer as razões pelas quais uma decisão específica é tomada pelo sistema de IA.

Este princípio deve ser adotado considerando o contexto para o qual o sistema de IA foi projetado, desenvolvido e implementado, como um equilíbrio entre a T&E e os princípios de privacidade e segurança dos dados.



ADOÇÃO GRADUAL

BOAS PRÁTICAS

1

Relatar quando os usuários interagem com a IA, gerando a consciência do usuário desde os primeiros contatos com o sistema de IA, se as decisões são baseadas em algoritmos de IA ou se um determinado resultado de um sistema de IA é levado em conta para tomar uma decisão

2

Ter um registro de como o sistema foi projetado, construído e mantido (mantê-lo sempre atualizado).

3

No contexto, elaborar uma comunicação simples, evitando linguagem confusa, de maneira clara e apropriada ao nível de alfabetização digital dos usuários sobre como cada etapa do sistema de IA é implementada para aumentar a conscientização sobre o uso do sistema de IA.

A comunicação do conteúdo poderia responder perguntas importantes, dentre outras, como: Qual é o nível de confiança do uso de IA no sistema de tomada de decisão? Quais dados foram usados para treinar o modelo? Qual é o mecanismo de atualização de dados? Como é garantida a qualidade dos dados usados? Quais são os elementos organizacionais (processos, modelos de negócios, governança, etc.) ligados ao funcionamento do sistema de IA? Quais são os canais de comunicação entre o usuário e a empresa?

4

Fornecer informações sobre como o sistema de IA toma decisões, revelando os fatores que contribuem para a decisão, bem como os mecanismos para que os usuários forneçam feedback.

5

Oferecer mecanismos intuitivos para os usuários entenderem como o sistema de IA funciona, bem como processos para solicitar explicações e informações adicionais sobre como o sistema de IA funciona.

6

Quando a decisão afetar os direitos e liberdades dos usuários, prever um mecanismo para solicitar revisões ou alterações.

7

Divulgar a existência ou não de garantias para os usuários.



VOCÊ QUER SABER MAIS?

Algumas tecnologias de IA já possuem ferramentas ou bibliotecas fornecidas por seus criadores ou por terceiros para compreender melhor como funcionam os algoritmos e modelos. Se não houver, a explicação dos modelos se torna uma tarefa mais avançada, a ser desenvolvida pela empresa.

Várias organizações desenvolveram ferramentas para orientar os responsáveis pela transparência e explicabilidade dos sistemas de IA. Vejam aqui alguns exemplos:

- IBM. "IBM's Principles for Trust and Transparency", https://www.ibm.com/blogs/policy/wp-content/uploads/2018/06/IBM_Principles_SHORT.V4.3.pdf.

- IBM, "AI FactSheets 360", https://aifs360.mybluemix.net/?_ga=2.114415694.891271442.1623897216.1860865525.1623897216&_gac=1.251733499.1623897216.CjwKCAjwwqaGBhBKEiwAMk-FtLnz4b6lnj74bS3CfWK2tIL-RxPxPn_DiqjbXrmgithyEhiQVQ1lTRoCbUoQAvD_BwE.
- IEEE, "IEEE P7001™ - Standards for Transparency of Autonomous Systems", 2020, <https://standards.ieee.org/project/2863.html>.
- Google, "Responsible AI practices", <https://ai.google/responsibilities/responsible-ai-practices/?category=interpretability>.

- META AI, "CAPTUM" Model Interpretability for PyTorch, <https://captum.ai/>
- META AI, "AI Explainability", TTC Labs, <https://www.ttclabs.net/theme/ai-explainability>
- Microsoft, "InterpretML", <https://www.microsoft.com/es-mx/ai/responsible-ai-resources?activetab=pivot1%3aprimar4>

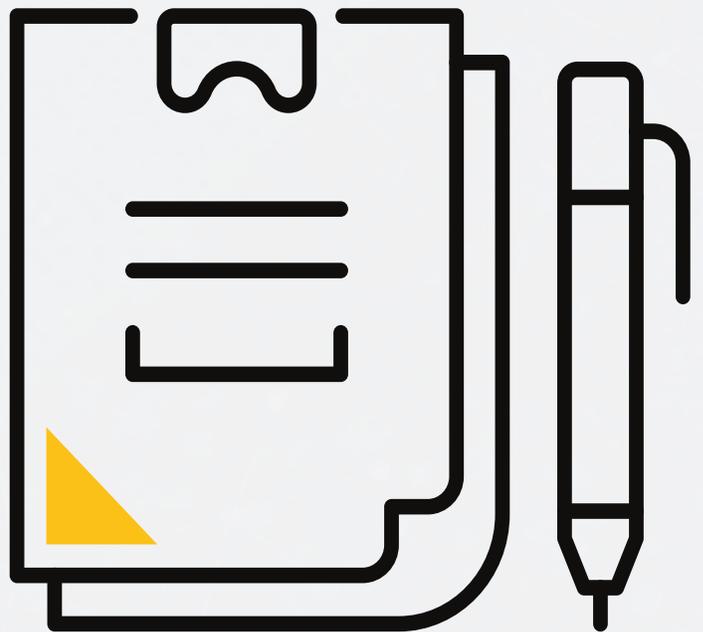


- Microsoft, "InterpretML", <https://www.microsoft.com/es-mx/ai/responsible-ai-resources?activetab=pivot1%3aprimar4>
- Microsoft, "Fairlearn", <https://www.microsoft.com/es-mx/ai/responsible-ai-resources?activetab=pivot1%3aprimar4>
- Microsoft, "Datasheets for Datasets", <https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets/>
- UK Information Commissioner's Office, (2019), "Explaining Decisions with AI. Part 2: Explaining AI in Practice ", <https://ico.org.uk/media/2616433/explaining-ai-decisions-part-2.pdf>
- TRUST-AI, "Transparent, reliable and unbiased smart tool for AI", <https://cordis.europa.eu/project/id/952060>

CASOS DE USO:

- META AI, "CAPTUM" Model Interpretability for PyTorch, <https://captum.ai/>
- META AI, "AI Explainability", TTC Labs, <https://www.ttclabs.net/theme/ai-explainability>
- Google, "Participatory approaches to transparency in dataset documentation", https://pair-code.github.io/datacardsplaybook/?utm_source=pocket_mylist

RESPONSABILIDADE E PRESTAÇÃO DE CONTAS





RESPONSABILIDADE E PRESTAÇÃO DE CONTAS

O QUE É ISSO?

O sistema de IA deve operar de acordo com o propósito previamente definido e servir os propósitos para os quais o sistema de IA foi projetado, desenvolvido, implementado e utilizado. As ferramentas técnicas ajudam na identificação e mitigação dos riscos causados por um sistema de IA; entretanto, somente as pessoas (agindo individualmente ou em grupo) podem ser responsáveis pelos sistemas de IA.

Portanto, é necessário avaliar o sistema à luz do contexto específico de sua aplicação e identificar possíveis riscos para os usuários finais, por exemplo, um

sistema de IA que é usado como uma ferramenta para fornecer informações para que um ser humano tome uma decisão será diferente daquele que é automatizado para tomar uma decisão sem intervenção humana.

Empresas e empreendimentos estabelecem a governança do sistema de IA, definindo políticas internas, bem como a responsabilidade que cada parte interessada tem de orientar o projeto, desenvolvimento, implementação e uso dos sistemas de IA dentro de uma organização.





ADOÇÃO GRADUAL

BOAS PRÁTICAS

1

Escolher o modelo de IA de acordo com o objetivo que o sistema de IA está buscando. Uma vez definido o objetivo legítimo pretendido pelo sistema de IA, deve-se assegurar que o ciclo de vida do sistema de IA responde adequadamente ao propósito determinado (nenhum processamento adicional deve ser executado além do estritamente necessário).

2

Ser claro sobre as responsabilidades de cada pessoa envolvida no desenvolvimento, implementação e uso do sistema de IA.

3

Documentar o desenvolvimento, testes e implementação do modelo do sistema de IA, incluindo pelo menos as métricas utilizadas, os testes executados para alcançar o funcionamento correto do sistema, bem como o registro rastreável e auditável do uso do sistema.

4

Adotar metodologias de identificação, avaliação e mitigação de riscos sobre o impacto do sistema de IA durante todo o ciclo de vida do sistema de IA, classificando o tipo de risco, seu nível de impacto e seu escopo.

5

Projetar, com base na avaliação de risco, uma estrutura de governança para tomar decisões sobre o tratamento e a mitigação dos diferentes tipos de riscos identificados.

6

Incluir o nível de intervenção humana e supervisão na tomada de decisões. A supervisão humana deve ser considerada necessária quando a decisão afeta diretamente os direitos de uma pessoa ou quando é de alto impacto, este último dependendo do resultado do processo de avaliação de risco).

7

Fornecer informações sobre o modelo adotado considerando: o propósito e os valores pretendidos, proporcionar acesso às políticas que regem o comportamento, bem como as entradas e saídas, como o sistema de IA funciona, o comportamento esperado, métodos de treinamento, avaliações de desempenho realizadas, bem como limitações ou riscos identificados, por exemplo, em que casos o sistema de IA deve ou não ser usado, o nível de eficácia do modelo (taxa falsa positiva e negativa) como indicadores da maturidade do sistema e os possíveis efeitos de sua adoção.

8

Ter mecanismos simples e claros que permitam ao usuário solicitar uma revisão de uma decisão apoiada ou tomada por um sistema de IA, quando apropriado, e dependendo da avaliação de risco.

9

Desenvolver um protocolo para resposta e monitoramento de cenários imprevistos ou indesejados na operação de sistemas baseados em IA, incluindo remédios eficazes para danos, considerando a análise e avaliação de risco realizadas.

10

Criar um conselho interno para a governança da IA, onde, por exemplo, poderiam ser definidos papéis e responsabilidades (daqueles envolvidos na concepção, desenvolvimento e implementação do sistema de IA), ações de verificação, monitoramento e aplicação, periodicidade das avaliações de risco, bem como fluxos de trabalho para intensificar a tomada de decisões sobre danos potenciais ou mitigação de risco.



VOCÊ QUER SABER MAIS?

Há algumas ferramentas para orientar o processo de definição das melhores práticas dentro da companhia para cumprir o princípio de responsabilidade e prestação de contas:

- BID, "IA Manual Técnico Responsável – Ciclo de Vida da Inteligência Artificial", 2020, <https://bit.ly/IAResponsableManualTecnicoCiclodevida>

- IBM, "Accountability", <https://www.ibm.com/design/ai/ethics/accountability/>.
- IEEE, "Ethically Aligned Design", https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined.
- IEEE, "IEEE P2863 - Recommended Practice for Organizational Governance of Artificial Intelligence", <https://standards.ieee.org/project/2863.html>.
- META AI, "Facebook's five pillars of Responsible AI", 2021, <https://ai.facebook.com/blog/facebook-fives-five-pillars-of-responsible-ai/>
- Microsoft, "Responsible AI", <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimar6>.
- OCDE, "Accountability (Principle 1.5)", <https://oecd.ai/dashboards/ai-principles/P9>.

Além disso, algumas metodologias para implementar avaliações de risco são listadas de acordo com o context em

que o sistema de IA deve ser utilizado (setor, mercado, usuários):



- EY, "Supervisory expectations and sound model risk management practices for artificial intelligence and machine learning", https://assets.ey.com/content/dam/ey-sites/ey-com/en_us/topics/banking-and-capital-markets/ey-mrm-ai-ml.pdf?download
- EY, "Understand model risk management for AI and machine learning", 2020, https://www.ey.com/en_us/banking-capital-markets/understand-model-risk-management-for-ai-and-machine-learning.
- McKinsey, "Derisking AI by design: How to build risk management into AI development", 2020, <https://mck.co/3yvl73x>

- OPEN LOOP, "AI Impact Assessment: A Policy Prototyping Experiment", 2021, https://d32j3j47emgb6f.cloudfront.net/wp-content/uploads/2021/01/AI_Impact_Assessment_A_Policy_Prototyping_Experiment.pdf
- IEEE, "IEEE 7010-2020™ (Standard Now Available) - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being", 2020, <https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html>
- PWC, "Model risk management of AI and machine learning systems", 2020, <https://www.pwc.co.uk/data-analytics/documents/model-risk-management-of-ai-machine-learning-systems.pdf>
- COUNCIL OF EUROPE, AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), "The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law", 2020, <https://rm.coe.int/cahai-2020-06-fin-c-muller-the-impact-of-ai-on-human-rights-democracy-/16809ed6da>
- COUNCIL OF EUROPE, "Towards Regulation of AI Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law.", 2020, <https://www.coe.int/en/web/artificial-intelligence/-/-toward-regulation-of-ai-systems>

Em conjunto com os modelos de avaliação de risco, recomenda-se que as empresas criem uma cultura de respeito aos direitos humanos em toda a sua organização. Para este exercício, sugere-se consultar o trabalho do Conselho da Europa, que analisou o impacto da IA nos direitos humanos, na democracia e no Estado de direito, bem como as possíveis estratégias que as empresas podem implementar:

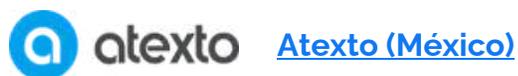
CASOS DE USO:

- Microsoft, Guidelines for Human-AI Interaction, "HAX Toolkit", <https://www.microsoft.com/en-us/haxtoolkit/>



ANEXO A

Sobre as empresas participantes:



Cria, coleta e transcreve arquivos de fala para ajudar as máquinas a entender as pessoas.



Digitaliza a captura de informações para processos no setor financeiro utilizando assistentes virtuais.



Ajuda a compreender as necessidades dos cidadãos para que as instituições e os governos possam tomar melhores decisões.



É uma plataforma que projeta, lança e otimiza campanhas de marketing digital.



Classifica automaticamente os resíduos para reciclagem através do reconhecimento da forma e do som



Desenvolve soluções para transformar dados em decisões.



DRONE DOMAIN



Oferece tecnologia para alcançar melhores rendimentos, melhor qualidade e reduzir o impacto ambiental da agricultura.



DYOO (México)

Desenvolve tecnologia de reconhecimento facial para fornecer análises em tempo real e gerar soluções em segurança, controle de acesso e análises para análise de mercado.



Maat.ai (México)

É uma ferramenta que permite as pessoas criar e compartilhar sua identidade digital.



Nedar (Colômbia)

Fornecer tecnologias de realidade virtual e realidade aumentada, melhorando os processos de treinamento através da experiência, motivação e desempenho no aprendizado.



R1T1 (Brasil)

É uma tecnologia de assistência médica em todas as áreas do hospital.

IDATHA IDATHA (Uruguai)

Permite a detecção de bots e notícias falsas.



Madison (República Dominicana)

É um software de faturação e controle de estoque.



Portal Telemedicina (Brasil)

Se concentra nos cuidados médicos e diagnósticos.



Quick Hit Solutions (México)

Otimiza as informações das empresas.



REMAP4.0 (México)

Gera rotas inteligentes e promove o comércio de proximidade.



TALOV [Talov \(Equador\)](#)

Traduz a linguagem dos sinais para surdos e mudos através de vídeo ou imagem.

Tooring [Tooring \(México\)](#)

Desenvolve sistemas de IA em diferentes disciplinas, incluindo geografia, saúde e outras.



WIZDEM [Wizdem \(México\)](#)

Oferece um sistema de referência de projetos sociais e ferramentas de gestão para ajudar a produzir projetos sociais mais eficientes.

TELEDX.org [TeleDx \(Chile\)](#)

Através de seu sistema DART, oferece diagnósticos médicos através do reconhecimento de imagens e previsão de doenças.



u-planner [U-Planner \(Chile\)](#)

Realiza diagnósticos personalizados de melhoria do estudante e orientação institucional



XIRA [Xira \(México\)](#)

Automatiza os processos de atendimento ao cliente com Chatbots e RPA.



ENDNOTES

- [1] Inter-American Development Bank, "Ethical and Responsible Adoption of Artificial Intelligence in Latin America and the Caribbean", fAIRLAC, janeiro de 2020, <https://publications.iadb.org/es/fair-lac-adopcion-etica-y-responsable-de-la-inteligencia-artificial-en-america-latina-y-el-caribe>.
- [2] Inter-American Development Bank, "Inteligência Artificial para o Bem Social na América Latina e no Caribe": Visão geral regional e instantâneos de doze países", fAIRLAC, maio de 2020, <https://publications.iadb.org/es/la-inteligencia-artificial-al-servicio-del-bien-social-en-america-latina-y-el-caribe-panor%C3%A1mica-regional-e-instant%C3%A1neas-de-doce-paises>.
- [3] Organização para Cooperação e Desenvolvimento Econômico, "Artificial Intelligence in Society", 2019, <https://doi.org/10.1787/eedfee77-en>.
- [4] United Nations Educational, Scientific and Cultural Organization, 'Decisions Adopted by the Executive Board at its 206th session', 17 de maio de 2019, https://unesdoc.unesco.org/ark:/48223/pf0000367821_spa.nameddest=42.
- [5] Alguns exemplos: Grupo Independente de Peritos de Alto Nível em Inteligência Artificial, Diretrizes Éticas para IA de Confiança, <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>; Conselho da Europa <https://www.coe.int/ai>; Instituto de Engenharia Elétrica e Eletrônica (IEEE) <https://standards.ieee.org/initiatives/artificial-intelligence-systems/>; Organização para Cooperação e Desenvolvimento Econômico (OECD) <https://www.oecd.org/going-digital/ai/>; União Internacional de Telecomunicações (UIT) <https://www.itu.int/en/ITU-T/IA/Pages/default.aspx>, entre outros.
- [6] Existe uma ampla gama de diretrizes éticas para o uso de IA, três das quais são fáceis de ler: (i) IALATAM, Guia prática para elaborar uma ética IA, 2 de novembro de 2020, <https://ia-latam.com/2020/11/02/guia-practica-para-desarrollar-una-ia-etica/>, (ii) Google, People + IA Guidebook, 8 de maio de 2019, <https://pair.withgoogle.com/guidebook/>, (iii) Red Iberoamericana de



ENDNOTES

de protección de datos Recomendaciones generales para el tratamiento de datos en inteligencia artificial, fevereiro de 2020, <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf>.

- [7] Comissão Europeia, Grupo Independente de Peritos de Alto Nível em Inteligência Artificial, 'Ethical guidelines for trustworthy AI', junho de 2018, <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.
- [8] Organização para Cooperação e Desenvolvimento Econômico, 'Recommendations of the Council on Artificial Intelligence', 21 de maio de 2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449?_ga=2.66351853.2133044157.1621127249-64840690.1621127249.
- [9] Organização das Nações Unidas para a Educação, Ciência e Cultura, 41ª Conferência Geral - Paris 2021, Relatório da Comissão de Ciências Sociais e Humanas, Recomendação sobre a Ética da Inteligência Artificial (IA), Anexo ao documento 41 C/23, https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa.page=15.
- [10] Organização para Cooperação e Desenvolvimento Econômico, 'Recommendations of the Council on Artificial Intelligence', 21 de maio de 2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449?_ga=2.66351853.2133044157.1621127249-64840690.1621127249.
- [11] Comissão das Nações Unidas para o Direito Comercial Internacional. "Legal Issues in the Digital Economy: Artificial Intelligence". Maio de 2020, <https://undocs.org/es/A/CN.9/1012/Add.1>.



Com o apoio e as contribuições de

