Inteligência Artificial e Proteção de Dados:
Fornecendo responsabilidade sustentável da IA na Prática

Segundo relatório:

Questões difíceis e Soluções Práticas

Fevereiro de 2020

Centre for Information Policy Leadership
HUNTON ANDREWS KURTH

Prefácio

Bojana Bellamy
Presidente do CIPL

O crescimento e rápida expansão da tecnologia de Inteligência Artificial é um dos grandes traços da Quarta Revolução Industrial. Não há precedentes para seu potencial de transformação para nossa sociedade digital e nossa capacidade de extrair benefícios para cidadãos, governos e organizações. Para realizar esse potencial e garantir sua sustentabilidade, precisamos construir a IA com base em confiança e respeito por nossos valores humanos, direitos e leis de privacidade de dados.

Este segundo relatório do Centro de Liderança em Políticas de Informações (CIPL) de nosso projeto sobre **Inteligência Artificial e Proteção de Dados** pretende fornecer insights sobre soluções emergentes para fornecer IA confiável e responsável.

O progresso em velocidade da luz nos domínios tanto da inteligência artificial quanto da lei de proteção de dados criou novos problemas, novas perguntas e, às vezes, até mesmo tensões entre esses campos. Em outubro de 2018, o CIPL destacou muitas dessas tensões principais em seu primeiro relatório neste projeto. Nossa ambição era provocar um diálogo global sobre como podemos enfrentar os desafios que a tecnologia de IA apresenta à proteção de dados ao mesmo tempo em que se permite inovação e avanço nessa área de rápido crescimento.

No ano passado, o CIPL envolveu-se nesses debates com organizações de diferentes setores da indústria, reguladores de proteção de dados na América do Norte, Europa e Ásia, legisladores e governantes, acadêmicos e outras partes interessadas. Através de mesas redondas, oficinas, reuniões paralelas e interações com profissionais, o CIPL identificou uma infinidade de métodos práticos e medidas que as organizações que estão desenvolvendo ou usando tecnologia de IA podem implementar hoje para garantir que não apenas estejam em conformidade com as exigências de proteção de dados, mas que estejam realmente entregando IA sustentável e responsável na prática.

Sou muito grata aos integrantes do CIPL e a outros participantes nos inúmeros eventos do ano passado por suas contribuições e melhores práticas emergentes para uma governança efetiva da IA que foram parte essencial para construir este segundo relatório. Estamos ansiosos para continuar nossa colaboração juntos à medida que avançamos para a próxima fase deste projeto, o qual olhará mais profundamente os mecanismos de governança da IA, as abordagens em camadas para regular a IA e tecnologias específicas de IA, incluindo reconhecimento facial.

Bojana Bellamy
Presidente

Sumário

I. Resumo executivo

"As organizações estão começando a desenvolver melhores práticas e estas estão sendo moldadas na forma de estruturas de IA mais coerentes e abrangentes, incluindo a base da Roda de Responsabilidade do CIPL."

Em outubro de 2018, o CIPL publicou o primeiro relatório de seu Projeto sobre *Inteligência Artificial e Proteção de Dados: Fornecendo responsabilidade sustentável da IA* na Prática. O relatório detalhou o uso generalizado, as capacidades e o potencial notável das aplicações de IA e examinou algumas das tensões existentes entre as tecnologias de IA e os princípios tradicionais de proteção de dados. O relatório concluiu que garantir a proteção de dados pessoais "exigirá práticas inovadoras das empresas e interpretação razoável das leis existentes por parte das autoridades reguladoras caso se deseje que os indivíduos sejam protegidos de forma eficaz e que a sociedade aproveite os benefícios das ferramentas avançadas de IA." [1]

Após a publicação do primeiro relatório, o CIPL se aprofundou mais na análise de alguns dos desafios mais difíceis com relação à IA, incluindo a articulação de mesas redondas e workshops em todo o mundo com autoridades reguladoras, legisladores e governantes, líderes da indústria e acadêmicos para identificar ferramentas e melhores práticas emergentes para abordar essas questões-chave.[2] Essa abordagem não tentou fornecer uma análise abrangente de todas as questões. Em vez disso, focamos em questões particularmente críticas para garantir o uso responsável da IA no contexto das leis de proteção de dados que, muitas vezes, foram promulgadas antes da explosão das tecnologias de IA. Essas questões incluem justiça, transparência, especificação de objetivos e limitação de uso, além da minimização de dados.

Nestas oficinas e outras conversas com especialistas em IA e proteção de dados, e com autoridades reguladoras em todo o mundo, seis mensagens surgiram com consistência notável:

1. A proliferação de ferramentas de IA, seu impacto crescente nos indivíduos e sociedades e sua dependência de grandes volumes de dados granulares e, muitas vezes, pessoais requerem **proteção efetiva dos dados** tanto por parte do setor privado quanto dos governos.

2. Há escopo suficiente nas medidas atuais de proteção de dados para fornecer grande parte dessa proteção. No entanto, alcançar isso requer **criatividade, flexibilidade, agilidade, cooperação e vigilância contínua** das organizações, dos reguladores e dos governantes à medida que evoluem as tecnologias e aplicações de IA, além das percepções do público e de nossa compreensão dos riscos.

3. Em algumas áreas, e incluindo algumas das questões que abordamos abaixo, provavelmente serão necessárias **inovações em governança, abordagens e interpretação regulatórias** para garantir que indivíduos e comunidades possam desfrutar de todo o potencial da IA sem comprometer a proteção de dados pessoais ou outros direitos fundamentais. Felizmente, a IA pode ajudar a facilitar novas abordagens, e é importante garantir que essas abordagens sejam consistentes com as regulamentações existentes, e não uma duplicata destas.

4. O processo de **criar, inovar e colaborar de forma multidisciplinar nas equipes** é fundamental para conseguir o uso responsável da IA e a proteção de dados pessoais, e nos ajudará a estarmos equipados para antecipar outros desafios de proteção de dados no futuro e para responder a eles adequadamente.

5. Precisamos ser **razoáveis em nossas expectativas** de IA, principalmente no início. Como humanos, raramente somos, de forma consistente, racionais, livres de tendências ou capazes de explicar por que chegamos às decisões que tomamos. A longo prazo, é muito mais provável que a IA, e não os humanos, alcance as metas de racionalidade, consistência e justiça, e devemos aspirar a que a IA faça isso. Porém, se insistirmos e exigirmos, desde o início, que a IA alcance padrões que o comportamento humano não consegue alcançar, corremos o risco de restringir o desenvolvimento de novas ferramentas com enorme potencial para os indivíduos e a sociedade.

6. Muitas **organizações e líderes de tecnologia de IA estão começando a trabalhar com os riscos, desafios e tensões de forma proativa** para entregar IA responsável e em conformidade com as leis de privacidade de dados e com as expectativas da sociedade. As organizações estão começando a desenvolver melhores práticas e estas estão sendo moldadas em estruturas de IA mais coerentes e abrangentes, incluindo a base da Roda de Responsabilidade do CIPL (ver Parte III.F. a seguir).

Esse segundo relatório fornece uma visão geral do que aprendemos sobre os desafios de IA no contexto da proteção de dados, de abordagens concretas para mitigá-los e alguns exemplos-chave de abordagens e ferramentas criativas que podem ser implementadas hoje para estimular um melhor futuro no qual possam prosperar a IA centrada no humano, a privacidade e a proteção e uso produtivo de dados pessoais.

Na seção seguinte, vamos debater quatro desafios de proteção de dados significativos apresentados pela IA (justiça, transparência, especificação de propósito / limitação de uso e minimização de dados) e fornecer exemplos de ferramentas para gerenciá-los. Na seção final, descreveremos temas transversais mais amplos que surgiram em nossa pesquisa, assim como melhores práticas, controles e ferramentas que estão ajudando a resolver ou mitigar esses desafios. Também detalhamos como a Roda de Responsabilidade do CIPL pode ser uma estrutura útil para que as organizações e as autoridades reguladoras organizem essas melhores práticas de tal forma que possam entregar IA confiável e responsável na prática.

II. Compreendendo as questões

Avanços consideráveis em IA criaram ou exacerbaram desafios para os princípios existentes de proteção de dados, alguns dos quais estão se mostrando particularmente importantes e difíceis de enfrentar por parte de organizações e autoridades reguladoras. Esta seção analisará quatro das questões mais desafiadoras com as quais estão lutando reguladores, líderes de indústria, engenheiros de IA, estudiosos e sociedade civil no que se refere a IA e a proteção de dados. Ofereceremos também algumas das interpretações e soluções que estão sendo usadas para alcançar um equilíbrio adequado entre a proliferação de aplicações de IA e a proteção de dados pessoais.

A. Justiça

"Uma ferramenta-chave para avaliar e conseguir maior justiça é o uso de um método baseado em risco ou prejuízo para orientar as decisões. Essa norma se baseia no impacto dos usos dos dados e o potencial de prejuízo para os indivíduos, mais que nas expectativas de uma razoável pessoa hipotética."

O processamento justo é um princípio fundamental de proteção de dados e uma exigência do Regulamento Geral de Proteção de Dados da UE (RGPD) e de outras leis de proteção de dados.[3] Porém, apesar de sua importância, o princípio de processamento justo não foi definido de forma consistente ou oficial. A definição de "justiça" tem sido um constante desafio, tanto no contexto da IA quanto em outros âmbitos com relação a privacidade e proteção de dados. O teste duradouro para o que é uma prática de negócios "injusta" empregado pela Comissão Federal do Comércio dos Estados Unidos (FTC) é se a prática causa um prejuízo substancial que não seja remediado por nenhum benefício compensatório para os consumidores ou concorrência que a prática produz, e se causa um prejuízo que os próprios consumidores não poderiam, razoavelmente, ter evitado.[4]

A UE parece considerar a "justiça" como um conceito muito mais amplo. Os 23 idiomas oficiais da União Europeia para os quais os princípios do RGPD foram traduzidos sugerem uma ampla gama de significados, incluindo boa fé, honestidade, propriedade, bondade, retidão, equidade, lealdade, confiabilidade, fidelidade, objetividade, devido processo, *fair play*, integridade, confiabilidade, correção, virtuosidade, imparcialidade, justiça e devoção (ver Apêndice A). O recente projeto de diretrizes do Comitê Europeu para a Proteção de Dados sobre *Proteção de Dados Desde a Concepção (by Design) e por Padrão (by Default)* tentam avançar na interpretação do processamento justo. O Comitê define justiça exigindo que os dados pessoais não sejam processados de forma "prejudicial, discriminatória, inesperada ou enganosa referente ao titular dos dados", e passa a destacar 12 elementos para a avaliação de justiça: autonomia, interação, expectativa, não discriminação, não exploração, escolha do consumidor, equilíbrio de poder, direitos e liberdades, ética, verdadeiro, intervenção humana e algoritmos

justos.[5] Todos esses podem, de fato, ser atributos desejáveis.[6] Porém, como um princípio fundamental para a proteção de dados e um requisito legal sob o RGPD, seriam desejáveis algumas deliberações mais amplas entre as partes interessadas sobre o conceito de processamento justo (ou justiça e injustiça) para autoridades reguladoras e organizações reguladas.

Na prática, a justiça parece ser um conceito amorfo que é **subjetivo, contextual e influenciado por uma série de fatores sociais, culturais e jurídicos.** Os mesmos dados usados em contextos diferentes podem suscitar reações totalmente diferentes a questões de justiça. Por exemplo, se as universidades usarem futuros alunos para treinar um algoritmo que personalize a propaganda para "futuros alunos não tradicionais" tais como estudantes universitários de primeira geração, a avaliação de justiça poderá ser diferente do que se os mesmos dados forem usados para identificar os alunos mais capacitados a pagar pela universidade e direcionar propaganda para as populações com melhores recursos.[7] A natureza contextual da justiça cria desafios significativos para os reguladores encarregados de interpretar e fazer cumprir a lei, para as organizações encarregadas de implementá-la e para os indivíduos cujos direitos deveriam estar protegidos por ela.

A dificuldade e importância de definir e garantir **justiça são apenas ampliados em contextos de IA.** Isso é verdade por causa da escala, da velocidade e do impacto da IA; a complexidade dos algoritmos de IA; a variedade e origem às vezes incerta dos dados de entrada; a imprevisibilidade, ou às vezes resultados inesperados de certos algoritmos de IA; uma falta frequente de interação direta com os indivíduos; e expectativas menos bem-formadas ou definidas no indivíduo médio. Essas características de IA com frequência exacerbam os desafios de a justiça não ter um significado claro e consistente e de que esse significado depende, pelo menos em parte, do contexto e de outros fatores.

Outro desafio em potencial da justiça é a **falta ou invisibilidade de um prejuízo individualizado.** Resultados injustos são, muitas vezes, impactos de amplo espectro na sociedade como um todo, e mesmo quando o prejuízo individual ocorre, não é facilmente reconhecido em um nível individual. Isso cria uma necessidade imperativa de que haja mais ação por parte das organizações e das autoridades reguladoras para resguardar a justiça. Por exemplo, no estado atual de desenvolvimento, as tecnologias de reconhecimento facial tendem a ser mais precisas para indivíduos de pele clara, portanto implantar essas tecnologias para uma população diversa em situações de alto risco poderia criar injustiça. O maior fornecedor de câmeras corporais da polícia nos EUA decidiu parar de usar a tecnologia de reconhecimento facial nos coletes policiais porque acredita que não sejam éticos nem a provável imprecisão, nem o impacto sistematicamente discriminatório em um cenário de alto risco.[8] Da mesma forma, o Information Commissioner's Office do Reino Unido (UK ICO) defendeu que as forças policiais deviam "desacelerar" para considerar os impactos do reconhecimento facial ao vivo e adotar passos para eliminar o viés algorítmico antes da implementação.[9]

Como esses exemplos sugerem, a justiça deve ser abordada em duas dimensões: processo justo (ou seja, processos que levem em consideração o impacto sobre os interesses dos indivíduos) e resultado justo (ou seja, a distribuição apropriada dos benefícios). Ambas as dimensões precisam ser trabalhadas para maximizar o valor dos dados e de suas aplicações para todos os envolvidos. Os princípios e valores

existentes na maioria dos regulamentos de proteção de dados permanecem relevantes, mas agora é necessário que haja mais diálogo e troca de pontos de vista entre as partes interessadas para criar maneiras práticas de atender esses princípios.

Felizmente, tanto as autoridades reguladoras quanto as organizações estão se esforçando para facilitar o **progresso para uma maior justiça no desenvolvimento e implementação da tecnologia de IA.**

Exemplos de Ações Reguladoras

Grupo de Especialistas de Alto Nível da Comissão da UE sobre Diretrizes de IA

O Grupo de Especialistas de Alto Nível (High-Level Expert Group - HLEG) da UE sobre IA publicaram *Diretrizes Éticas para IA Confiável*, incentivando as organizações a "assegurarem uma definição de trabalho adequada de 'justiça'" para aplicar nos projetos de sistemas de IA.[10] As Diretrizes fornecem **perguntas a serem consideradas pelas empresas ao criar políticas para promover justiça**, mas, em última instância, permite que as organizações desenvolvam sua própria definição e abordagem à justiça, assim como os processos e mecanismos para alcançá-la.

Resultados sobre justiça da Autoridade sobre Conduta Financeira do Reino Unido

Outra abordagem das autoridades reguladoras poderia ser **a definição de resultados de justiça ou considerações para a justiça**. Por exemplo, a Autoridade sobre Conduta Financeira do Reino Unido (UK Financial Conduct Authority - UK FCA) requer que todas as empresas sob sua autoridade tratem os clientes com justiça e, para esse fim, definiu seis resultados de justiça que, para serem alcançados, exigem que as organizações criem políticas e procedimentos:

Resultado 1: Os clientes podem estar confiantes de que estão lidando com empresas em que o tratamento justo aos clientes é central na cultura corporativa.

Resultado 2: Produtos e serviços comercializados e vendidos no mercado de varejo são concebidos para atender as necessidades de grupos de consumo identificados e são dirigidos para isso.

Resultado 3: Os consumidores recebem informações claras e são mantidos adequadamente informados antes, durante e depois do ponto de venda.

Resultado 4: Quando os consumidores recebem aconselhamento, o aconselhamento é adequado e leva em consideração suas circunstâncias.

Resultado 5: São fornecidos aos consumidores produtos cujo desempenho é o equivalente ao que as empresas os levaram a esperar, e o serviço associado não apenas é de um padrão aceitável como corresponde ao que foram levados a esperar.

Resultado 6: Os consumidores não enfrentam barreiras pós-vendas irracionais impostas pelas empresas para troca de produto, troca de fornecedor, envio de um pedido ou reclamações.[11]

Esses resultados de justiça essencialmente avaliam e equilibram os riscos para a indústria, exigindo que as organizações desenvolvam processos para alcançá-los.

Essas abordagens reguladoras do HLEG da União Europeia sobre IA e da FCA do Reino Unido permitem que as organizações inovem sobre como definem e garantem justiça ao mesmo tempo em que fornecem orientações sobre o que significa ser justos.

"As organizações precisam calibrar seu programa de privacidade e os controles específicos baseados nos resultados das avaliações de risco que elas conduzem. Quanto maiores os riscos, mais devem fazer através de supervisão, políticas, procedimentos, treinamento, transparência e verificação."

Uma ferramenta-chave para avaliar e conseguir maior justiça é o **uso de um método baseado em risco ou prejuízo para orientar as decisões**. Essa norma se baseia no impacto dos usos dos dados e no potencial de prejuízo para os indivíduos mais do que nas expectativas de uma razoável pessoa hipotética. Certos tipos de decisões trazem diferentes níveis de risco ou potencial de prejuízo, tais como a diferença entre recomendar uma rota de tráfego de logística e tomar uma decisão de atenção à saúde. Para determinar o nível de risco, uma organização pode responder pela população impactada, pelo impacto individual e pela probabilidade ou potencial de prejuízo. À medida que aumentam o potencial de prejuízo ou os impactos discriminatórios aos indivíduos, o mesmo deveria ocorrer com os controles e verificações instaurados por uma organização para limitar as consequências negativas à justiça. Essa também é a própria essência da abordagem de acordo com a estrutura da Roda de Responsabilidade do CIPL (apresentada em detalhe na Parte III.F. abaixo): as organizações precisam calibrar seu programa de privacidade e os controles específicos com base nos resultados das avaliações de risco que elas realizam. Quanto maiores os riscos, mais devem fazer através de supervisão, políticas, procedimentos, treinamento, transparência e verificação.

Qualquer que seja a forma em que uma organização define e avalia a justiça, é importante observar que **a justiça não é absoluta e pode requerer reavaliação contínua e iterativa.** Ao longo de conversas com autoridades reguladoras e líderes da indústria, uma visão comum se concentrou na justiça como algo que existe em um contínuo, mais do que sendo um conceito binário. Os participantes identificaram uma série de medidas, descritas em maiores detalhes abaixo, que poderiam ser usadas para tornar o desenvolvimento e a implementação de IA "mais justos", mas nenhuma dessas medidas são um tiro certeiro para se alcançar

justiça. Mais que isso, a garantia da justiça passa por um processo contínuo. Como será o processo pode depender do contexto, da cultura ou da organização, mas o desenvolvimento de processos e o monitoramento e resultados ajudará as organizações a se moverem ao longo do espectro, ou continuum, em direção à justiça sem importar qual compreensão específica e qual definição de justiça estejam sendo aplicadas.

Exemplos de Ações Organizacionais

As organizações estão começando a desenvolver uma série de ferramentas técnicas, procedimentos e estruturas para ajudar a garantir a justiça nas aplicações de IA.

Ferramentas

Em particular, as organizações estão desenvolvendo **ferramentas** para identificar e enfrentar o risco do viés algorítmico, já que percebem corretamente o viés como sendo um dos indicadores-chave de injustiça. Um exemplo é o teste contrafactual de justiça, que verifica a justiça nos cenários determinando se o mesmo resultado é alcançado quando muda uma variável específica, tal como raça ou gênero.[12] De fato, conforme identificado no primeiro relatório de IA do CIPL[13] e em inúmeros debates com engenheiros de IA, é absolutamente necessário que as organizações processem e retenham categorias de dados sensíveis, tais como etnicidade ou gênero, para evitar viés no modelo, ou para poder testar o modelo, monitorá-lo e afiná-lo em um estágio posterior e, assim, garantir justiça.[14] A Google, por exemplo, desenvolveu técnicas de justiça algorítmica para "fazer aflorar o viés, analisar conjuntos de dados, testar e compreender modelos complexos para ajudar a tornar os sistemas de IA mais justos", incluindo Facets, a ferramenta What-If, modelos e cartões de dados e treinamento com restrições de justiça algorítmica.

Essas e outras ferramentas são descritas em maiores detalhes no relatório de 2019 da Google, *Perspectivas em Questões de Governança de IA*.[15] A Accenture desenvolveu uma ferramenta de justiça para "identificar e remover qualquer influência coordenada que possa levar a um resultado injusto,"[16] que é usada tanto internamente, com relação a seus próprios projetos de IA, quanto externamente, em projetos de clientes que envolvem a implementação de aplicações de IA para ajudar os clientes a buscarem o padrão de justiça. A IBM também criou várias ferramentas para abordar questões de ética e justiça em IA, incluindo o AI Fairness 360, "um kit de ferramentas abrangente de código aberto com métricas para verificar vieses indesejáveis nos conjuntos de dados e modelos de aprendizagem de máquina, e algoritmos inovadores para mitigar esses vieses,"[17] assim como o IBM Watson OpenScale, uma ferramenta para rastrear e medir resultados de IA para ajudar a detectar e corrigir de forma inteligente os vieses e explicar as decisões de IA.[18]

Mecanismos de procedimentos e de responsabilização

Tão importante quanto essas ferramentas técnicas são a variedade de mecanismos de procedimentos e de responsabilização para garantir a justiça. As organizações podem criar estruturas internas de governança e estruturas de prestação de contas e, a seguir, utilizar ferramentas como avaliações de impacto de proteção de dados de IA (Data Protection Impact Assessments - AI DPIAs) ou Comitês de Revisões de Dados (Data Review Boards - DRBs) para implementar a responsabilidade da IA (discutida na Parte III deste relatório). Esses mecanismos são particularmente úteis na fase de desenvolvimento de aplicações de IA, mas também nas fases de revisão e monitoramento. Obviamente, proporcionar transparência e mecanismos para reparação será essencial para garantir a equidade ao longo da implementação de tecnologias de IA. Tudo isso exemplifica o argumento de que **a justiça deve ser assegurada durante todo o ciclo de vida** de um aplicativo de IA - desde a avaliação do caso de uso de AI e dados de entrada, modelagem algorítmica, desenvolvimento e treinamento até implementação, monitoramento, verificação e supervisão contínuos.

Transparência, explicabilidade e reparação

Além disso, muitos consideram que transparência, explicabilidade e reparação/compensação estão intrinsecamente vinculados à avaliação da justiça em um aplicativo de IA. Em outras palavras, proporcionar transparência significativa e foco no usuário, trazer explicabilidade sobre o processo de tomada de decisões de IA e permitir as reparações aos indivíduos provavelmente aumentarão as chances de que um processamento de dados específico em IA seja justo.

Consideração de questões que se referem a justiça além de proteção de dados

Um dos principais desafios ao realizar uma avaliação de justiça é se as organizações ou autoridades reguladoras podem ou devem **considerar questões além da proteção de dados**. Será que as organizações ou autoridades reguladoras devem olhar para além das questões de proteção de dados para avaliar potenciais impactos mais abrangentes de IA, por exemplo, no futuro do trabalho ou na competitividade das empresas? Ou elas deveriam considerar os impactos sobre outros direitos humanos, além da proteção de dados, ao analisar a justiça da aplicação de IA?

Com frequência, há outros órgãos legais responsáveis por avaliar essas preocupações, e as organizações individuais muitas vezes não estão bem equipadas para lidar com os impactos mais amplos nas sociedades de forma geral, que dirá os escritórios de proteção de dados dentro das organizações. Afinal, até mesmo os ministérios de comércio e outros legisladores governamentais, com suas amplas missões e recursos, muitas vezes têm dificuldades em prever o impacto social e econômico de novas tecnologias.

"Ao mesmo tempo em que definir e implementar justiça é um desafio, é também uma oportunidade. A IA, em última instância, pode ajudar a facilitar as metas de justiça, seja ajudando a trazer luz e mitigar vieses históricos ou fornecendo tomadas de decisões mais consistentes e racionais."

Porém, o RGPD parece sugerir que a avaliação de justiça requer considerar questões além da proteção de dados. Por exemplo, as avaliações de impacto de proteção de dados (DPIAs), discutidas abaixo, são necessárias para avaliar o processamento "que provavelmente resultará em um elevado risco aos direitos e liberdades das pessoas naturais,"[19] e isso não se limita a avaliar a proteção de dados ou os direitos de privacidade. Além disso, a 41st International Conference of Data Protection & Privacy Commissioners (ICDPPC) chamou a todas as organizações para que "avaliem o risco à privacidade, igualdade, justiça e liberdade antes de usar a inteligência artificial."[20]

Justiça como uma oportunidade

Dadas essas considerações, torna-se claro de que uma organização que busca avaliar holisticamente uma aplicação de IA ou o uso de dados novos pode querer expandir a abrangência de sua lente para incluir essas questões mais amplas. Isso também pode ser esperado como parte da responsabilidade social corporativa mais ampla das organizações, da qual a responsabilidade digital é um subconjunto. Um pequeno número de organizações estão até mesmo trabalhando para desenvolver uma avaliação mais ampla dos impactos nos direitos humanos que será conduzida ao desenvolver e implementar tecnologia de IA. Essas ferramentas ainda estão nos primeiros estágios de desenvolvimento, mas indicam a direção da estrada para algumas organizações responsáveis.

Finalmente, ao mesmo tempo em que definir e implementar justiça é um desafio, é também uma **oportunidade**. Como a Google recentemente observou, "[s]e bem implementada, uma abordagem algorítmica [à justiça] pode ajudar a impulsionar a consistência da tomada de decisões, especialmente comparando com a alternativa de indivíduos que julgam a partir de suas próprias definições internas (e, portanto, provavelmente variáveis) de justiça."[21] A IA pode, em última instância, ajudar a facilitar as metas de justiça, seja ajudando a trazer luz e mitigar vieses históricos ou fornecendo tomadas de decisões mais consistentes e racionais. Alcançar essa meta vai exigir que as organizações definam justiça e desenvolvam ferramentas e procedimentos para manter e assegurar sua definição ao longo do processo de desenvolvimento e implementação de IA.

"Transparência também significa a capacidade de articular benefícios de uma tecnologia de IA em particular e benefícios tangíveis para os indivíduos, assim como para a sociedade mais ampla."

B. Transparência

A transparência, como a justiça, é uma preocupação exacerbada pela IA, mas é também uma solução em potencial para muitos dos medos que pairam sobre as tecnologias de IA. Transparência em relação à IA exige que "as organizações forneçam aos indivíduos as especificidades do processamento de dados, incluindo a lógica por trás de qualquer tomada de decisão automatizada que tenha efeito legal ou um impacto igualmente significativo nos indivíduos."[22] Os objetivos da transparência são informar os indivíduos sobre como seus dados são usados para tomar decisões, responsabilizar as organizações por suas políticas e procedimentos relativos à IA, ajudar a detectar e corrigir o viés e, em geral, fomentar a confiança no uso e na proliferação da IA. As ferramentas com que as autoridades reguladoras e as organizações contam para facilitar a transparência devem ser desenvolvidas para atender a esses objetivos.

A transparência tem sido um desafio difícil na IA, pois muitas vezes não fica claro o que significa transparência. Como ponto de partida, pode ser útil considerar a alternativa humana de tomada de decisões. Com frequência, os humanos são incapazes de explicar consistentemente suas preferências de uma opção sobre a outra, e há inúmeras situações em que as decisões não são completadas de forma transparente, tais como aprovações de empréstimo ou crédito, ou decisões de contratação. Embora possamos pedir posteriormente uma explicação, essa explicação, na melhor das hipóteses, será lógica e quase certamente não será técnica ou matemática. Considerar abordagens à transparência em um mundo offline pode ser ilustrativo de que nível e tipo de transparência buscar ao construir sistemas de IA.

O desafio da transparência na IA se torna mais difícil devido à complexidade e natureza mutável dos algoritmos de IA. Um dos pontos fortes da IA é detectar padrões complexos que haviam sido esquecidos anteriormente, mas essa complexidade, por natureza, é inerentemente difícil de explicar em termos que sejam facilmente compreendidos pelos seres humanos. Os avanços na pesquisa levaram a ferramentas que podem ajudar os desenvolvedores a entender melhor como seus modelos de IA funcionam, mas isso exige investir tempo e energia para interrogar modelos, o que talvez nem sempre seja possível. Além disso, os sistemas de IA podem ser atualizados e retreinados usando entradas adicionais, para que as decisões não sejam facilmente repetíveis. Como esses sistemas são complexos e geralmente mudam, fornecer informações sobre o algoritmo pode não servir aos objetivos de transparência. Não apenas é improvável que a divulgação do código seja particularmente útil para fornecer clareza sobre a decisão, como a transparência algorítmica poderia ter os efeitos potencialmente prejudiciais de divulgar segredos comerciais ou ajudar as pessoas a burlar o sistema.

Entretanto, a transparência é uma obrigação legal de acordo com o RGPD e outras leis de proteção de dados[23] e uma ferramenta útil para desenvolver a confiança na IA. Por isso, é essencial criar consenso sobre o que significa transparência em qualquer situação e encontrar maneiras de fornecer transparência eficaz e significativa que atinja os objetivos mencionados acima.

Considerações para uma transparência efetiva

Alguns já argumentaram pela **divulgação obrigatória do fato de que está sendo usada IA** quando um indivíduo está interagindo com uma máquina e para uma explicação de por que a IA está sendo implantada e o que se espera de seu uso. Isso pode ser útil em alguns casos, mas muitas vezes será óbvio, excessivamente oneroso ou de outra forma ineficaz para construir confiança. Em última análise, não é a tecnologia que importa, mas o fato de que uma decisão não humana está tendo consequências para um indivíduo de uma forma que ele talvez não esteja esperando.

A transparência pode diferir dependendo do público a que se destina - o indivíduo ou categoria de indivíduos afetados pela decisão, a autoridade reguladora, um parceiro de negócios ou mesmo para fins de transparência interna a um comitê de supervisão líderes seniores. Todos esses públicos diferentes implicam diferentes tipos e requerimentos de transparência que devem ser cumpridos adequadamente. Por exemplo, uma autoridade reguladora pode precisar saber de mais detalhes sobre um caso de uso de IA no contexto de uma investigação ou auditoria - o modelo, os conjuntos de dados, entradas e saídas, medidas para garantir justiça e ausência de preconceito, etc. Por outro lado, para

um indivíduo, esse tipo de informação pode ser demais e será "não ver o todo da floresta por estar vendo as árvores". Da mesma forma, uma organização que desenvolve tecnologia de IA para ser usada por outra organização pode não ser capaz de fornecer transparência diretamente aos titulares dos dados, mas pode ser necessário fornecer transparência adicional sobre as medidas técnicas para garantir o funcionamento adequado do modelo, prevenção de viés, precisão, documentação com relação a trocas, etc. Portanto, pode ser difícil ser categórico sobre a lista precisa de elementos de transparência, pois dependerá muito de quem é o público e do objetivo específico de transparência em um determinado contexto.

"A revisão humana é uma exigência para certas decisões automatizadas de impacto previstas no RGPD. O desenvolvimento de vias eficientes e visíveis para essa revisão - seja antes ou depois de uma decisão - vai constituir uma parte importante da transparência nos contextos de IA."

O nível e o método de transparência devem, em última análise, estar vinculados ao contexto e à finalidade dos aplicativos de IA. Por exemplo, o recente Project Expl*AI*n da UK ICO, um projeto colaborativo entre a ICO e o Alan Turing Institute para criar orientação prática para ajudar as organizações a explicar decisões de IA aos indivíduos, fez uma pesquisa com júris de cidadãos e demonstrou que os indivíduos que enfrentam cenários de atenção à saúde por IA se importavam mais com a precisão do que com a transparência, ao passo que as expectativas com transparência aumentaram no uso de IA no recrutamento de empregos e em cenários de justiça criminal.[24] Isso sugere que a transparência, e as ferramentas usadas para alcançá-la, pode diferir com base em qual o uso da aplicação de IA, quais são as consequências e quais direitos os indivíduos têm.

Para ilustrar essas diferentes considerações sobre transparência, considere o uso de tecnologias de reconhecimento facial das companhias aéreas para verificar os cartões de embarque, ou de funcionários aduaneiros para permitir que indivíduos entrem em um país. A decisão tomada pela IA nesses casos é muito significativa, mas a transparência em relação ao fato de a IA estar sendo usada ou sobre o próprio código provavelmente não será motivo de preocupação para o indivíduo impactado. Em vez disso, a preocupação será como contestar ou alterar a decisão, portanto facilitar os objetivos de transparência exigirão maior ênfase na rapidez e em possibilidades eficazes de reparação. Embora a revisão humana seja uma exigência para certas decisões automatizadas de impacto previstas no RGPD,[25] o desenvolvimento de vias eficientes e visíveis para essa revisão - seja antes ou depois de uma decisão - vai constituir uma parte importante da transparência nos contextos de IA.

O nível de transparência e a quantidade de intervenção humana necessária pode variar **dependendo do risco apresentado ao indivíduo por uma decisão ou da visibilidade do processamento.** Para conceitualizar esse ponto, pode ser útil considerar um quadrante de quatro partes sobre o impacto e a visibilidade (Figura 1). Algumas decisões requerem maior transparência devido ao risco de prejuízos, enquanto outras exigirão maior transparência devido à sua invisibilidade. Alguns podem exigir pouca ou nenhuma transparência adicional, como recomendações para restaurantes ou filmes.
Também está claro que a **transparência é um conceito mais amplo no contexto de IA**, pois inclui explicabilidade e compreensibilidade, assim como a transparência com relação a opções de reparação e a capacidade de contestar uma decisão de IA. O HLEG da UE sobre IA considerou que a transparência incluía elementos de "rastreabilidade, explicabilidade e comunicação."[26] A rastreabilidade requer entradas de dados documentados e outros "conjuntos de dados e processos que levam à decisão do sistema de IA."[27] A criação de códigos de conduta ou melhores

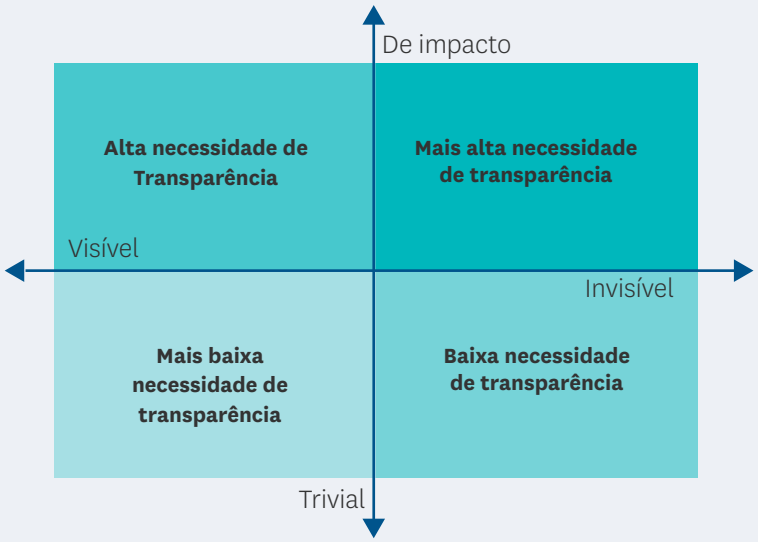

***Figura 1**. A figura acima demonstra a necessidade de transparência com base na visibilidade de processamento ou no uso da tecnologia e quão impactante ou quão grande é o risco de prejuízo para um indivíduo. Naturalmente, os indivíduos e autoridades reguladoras são os mais preocupados com processos que apresentam um maior risco de prejuízo, tais como a consideração de um empréstimo ou seguro. Entretanto, a transparência também é crítica quando não há evidência de processamento, como nos anúncios de marketing político. Quando há pouco ou nenhum risco de prejuízo (ou se o impacto desse prejuízo for trivial), haverá menos foco na transparência.*

A transparência é mais ampla que a explicabilidade

práticas com relação à coleta, implementação e uso de dados pode melhorar a rastreabilidade. Isso pode melhorar a garantia de justiça, realizando-se auditorias e estimulando a explicabilidade. Conforme observado na Estrutura de IA Modelo criada pelo PDPC de Cingapura, "[D]iz-se que um algoritmo implantado em uma solução de IA é explicável se for possível explicar como funciona e como chega em uma previsão em particular."[28] Ferramentas tais como o uso de contrafactuais, fichas de dados (que fornecem informações ou características sobre os serviços de IA),[29] ou Cartões Modelo (documentos curtos que acompanham modelos de IA que podem descrever o contexto nos quais os modelos devem ser usados, procedimentos de avaliação e outras informações relevantes)[30] podem ser úteis para explicar decisões. De forma geral, a rastreabilidade e a explicabilidade são formas de fornecer transparência sobre os dados de IA ou os processos de dados sem fornecer transparência sobre o próprio algoritmo. Esses conceitos promovem as metas de transparência aumentando a confiança e a responsabilização nas decisões.

Finalmente, transparência também significa a capacidade de **articular benefícios de uma tecnologia de IA em particular** e benefícios tangíveis para os indivíduos, assim como para a sociedade mais ampla. Dessa forma, as organizações podem fornecer valor educacional aos indivíduos e mobilizar uma maior confiança e aceitação das novas aplicações.

"Diz-se que um algoritmo implantado em uma solução de IA é explicável se for possível explicar como funciona e como chega em uma previsão em particular."

- PDPC de Cingapura
Estrutura de Governança de IA Modelo

Embora as organizações possam caminhar para oferecer uma transparência mais eficaz com relação a entradas e expectativas, pode haver circunstâncias em que as decisões não possam ser explicadas ao ponto necessário para que as autoridades reguladoras e os indivíduos estejam confiantes na tomada de decisões de IA. Nesse caso, as organizações devem usar outros métodos para estimular essa confiança e fiabilidade na IA. A ferramenta mais importante para ajudar com isso será fornecer vias visíveis de reparação através de responsabilização organizacional, a qual será discutida com mais detalhes na Parte III deste relatório.

A transparência é mais ampla que a explicabilidade:

Compreensibilidade	Rastreabilidade	Explicabilidade	Articulação de benefícios	Comunicação, direitos e vias de reparação
Uma compreensão de como funciona um sistema de IA e o que pretende alcançar	Documentando conjuntos ou processos de dados que resultam na decisão do sistema de IA de capacitar identificação de por que uma decisão de IA estava errada	Capacidade de explicar por que o sistema de IA alcançou uma certa decisão ou resultado	Informações sobre os benefícios tangíveis de uma tecnologia de IA em particular para os indivíduos e para a sociedade	Comunicado aos indivíduos de que eles estão interagindo com um sistema de IA, informações sobre seus direitos e mecanismos de reparação

"As organizações e as autoridades reguladoras devem buscar equilíbrio fornecendo especificações de propósito significativas e usar a limitação ao mesmo tempo em que fornecem flexibilidade para reagir com novas inferências provenientes de conjuntos de dados antigos ou diferentes."

C. Especificação de propósito e limitação de uso

Os princípios de especificação de propósito e limitação de uso exigem que as empresas, respectivamente, especifiquem a finalidade para a qual estão processando dados e usem os dados apenas para esse propósito, ou para algum propósito compatível. Esses princípios já foram desafiados pela prevalência do big data, mas recentemente foram também questionados pela IA. Tanto as empresas quanto as autoridades reguladoras devem avaliar como aplicar de forma significativa a especificação de propósito e usar a limitação de uma maneira que atenda aos propósitos desses princípios, ao mesmo tempo em que permitem que a sociedade se beneficie das capacidades das novas tecnologias.

É importante observar que os princípios de especificação de propósito e limitação de uso não são absolutos. Por exemplo, o princípio de limitação de propósito do RGPD requer que os dados pessoais sejam "coletados para propósitos especificados, explícitos e legítimos, *e que não sejam processados de outra forma que seja incompatível com esses propósitos.*"[31] As Diretrizes de Privacidade da OECD, que sustentam a maioria das leis modernas de proteção a dados, contêm linguagem similar.[32] Esses princípios foram projetados para limitar processamento

de dados não previstos ou invisíveis, então a permissão de processamentos compatíveis serve ao sentido do princípio ao mesmo tempo em que permite alguma flexibilidade.

O desafio apresentado pela IA surge de sua capacidade de algumas vezes descobrir correlações inesperadas ou extrair inferências imprevistas dos conjuntos de dados. Isso pode expor novos usos ou propósitos para dados antigos. Por exemplo, novas tecnologias visuais computadorizadas podem usar exames e prontuários médicos antigos para desenvolver correlações e descobrir valores novos em dados antigos. O uso da limitação e especificação de propósito, se interpretado de forma estrita, poderia sufocar pesquisas futuras e impedir que indivíduos e sociedade reconheçam alguns dos benefícios em potencial da IA.

Balanceando a especificação de propósito e a limitação de uso com a capacidade da IA de descobrir propósitos novos e não previstos

"Em última análise, processamento posterior baseado em 'compatibilidade' deveria ser permitido para usos futuros que sejam consistentes com o propósito original, que possam coexistir com ele e que não o abalem ou neguem. Esses usos devem ser sustentados por fortes garantias baseadas em responsabilização, incluindo avaliações de riscos e benefícios, para assegurar que os novos usos não exponham o indivíduo a riscos ou impactos adversos aumentados e injustificados."

As organizações e as autoridades reguladoras devem promover equilíbrio fornecendo especificações de propósito significativas e usar a limitação ao mesmo tempo em que fornecem flexibilidade para reagir com novas inferências provenientes de conjuntos de dados antigos ou diferentes. O propósito amplo ou o uso de declarações de especificação oferecem pouco significado para os indivíduos e podem, em última análise, degradar a eficácia desses princípios. O espírito da especificação de propósito requer que a notificação seja precisa, já que "uso para IA" por si só não seria nem específico nem suficientemente preciso para fornecer informações significativas para o indivíduo.

Em vez de permitir que os propósitos se tornem tão amplos a ponto de perderem o sentido, as autoridades de proteção de dados têm interpretado os propósitos de forma estrita, o que reforça a necessidade de fornecer flexibilidade para permitir processamento posterior. O RGPD, como as Diretrizes de Privacidade da OECD, permitem explicitamente processamento posterior para novos propósitos "não incompatíveis".[33] Os critérios do RGPD do que é "não compatível" são úteis para permitir usos futuros.[34] Em última análise, processamento posterior baseado em "compatibilidade" deveria ser permitido para usos futuros que sejam consistentes com o propósito original, que possam coexistir com ele e que não o abalem ou neguem. Esses usos devem ser sustentados por fortes garantias baseadas em responsabilização, incluindo avaliações de riscos e benefícios, para assegurar que os novos usos não exponham o indivíduo a riscos ou impactos adversos aumentados e injustificados.
De fato, o próprio RGPD lista "as possíveis consequências do processamento adicional pretendido para os titulares dos dados"[35] como uma consideração da avaliação de compatibilidade. Isto é, efetivamente, uma abordagem baseada em riscos para determinar o que é "não compatível". Quanto maiores os riscos de consequências adversas, menos compatível será o processamento futuro, e vice versa.

Uma distinção útil permitiria **treinar um algoritmo para servir como propósito separado e distinto**.[36] O conceito de uma fase de treinamento é nova para a IA, e muitas vezes são necessários dados em maiores quantidades durante a fase de treinamento do que durante a implementação. Na fase de treinamento, onde não ocorre nenhuma tomada de decisão individual, o risco de prejuízo aos indivíduos

por alterar o propósito dos dados é diminuído ou totalmente eliminado. Assim, o processamento posterior nesta fase deveria ser estimado como compatível com o propósito original. Além disso, o processamento de dados pessoais para propósito de AI/treinamento do modelo poderia ser um bom exemplo de processamento baseado no teste de balanceamento de legítimo interesse sob o RGPD e outras leis que têm disposições semelhantes.

Finalmente, o nível de observação continuada e os requisitos necessários para processamento futuro de dados antigos podem ser compreendidos como uma função do risco de prejuízo apresentado por esse processamento. "Dados usados em um contexto para um propósito ou sujeito a um conjunto de proteções podem ser tanto benéficos quanto desejáveis, ao passo que os mesmos dados usados em um contexto diferente, para outro propósito ou sem proteções adequadas podem ser tanto perigosos quanto indesejáveis."[37] Portanto, a especificação de propósito e a limitação de uso pode ser mais eficaz se esses princípios se basearem menos na avaliação de adequação de um uso pretendido nos termos originais e focarem mais no risco e no impacto do novo uso. Esse foco nos impactos dos dados será mais explorado na Parte III deste relatório.

D. Minimização de dados

*"Para a IA, particularmente nos estágios de desenvolvimento e treinamento, o que é **necessário** é uma quantidade considerável de dados, e ter quantidade insuficiente de dados pode impedir o desenvolvimento de um algoritmo. Por exemplo, pode ser necessário coletar e reter quantias significativas de dados, incluindo dados sensíveis, para atenuar os riscos e garantir justiça em certas aplicações de IA."*

A minimização de dados apresenta um paradoxo semelhante: um princípio para limitar a coleta e a retenção de dados pessoais de indivíduos tem o potencial de impedir avanços em IA que poderiam, em última instância, ser benéficos para a sociedade. Como mencionado acima, a IA tem a capacidade de encontrar usos novos e benéficos para dados antigos, por isso pode ser impraticável minimizar a coleta ou retenção de dados.

Embora a intenção e os objetivos do princípio de minimização de dados ainda sejam possíveis em nosso cenário tecnológico, alcançar esses objetivos exigirá mais soluções criativas e interpretações flexíveis. O RGPD, por exemplo, exige que: "Os dados pessoais devem ser adequados, relevantes e limitados ao que é necessário em relação aos propósitos para os quais são processados."[38] Para a IA, particularmente nos estágios de desenvolvimento e treinamento, o que é *necessário* é uma quantidade significativa de dados, e ter quantidade insuficiente de dados pode impedir o desenvolvimento de um algoritmo. Por exemplo, pode ser necessário coletar e reter quantias significativas de dados, incluindo dados sensíveis, para atenuar os riscos e garantir justiça em certas aplicações de IA. Essa é uma troca contextual que as organizações precisarão avaliar cuidadosamente a fim de encontrar um equilíbrio adequado entre os requisitos concorrentes. Por exemplo, pode ser completamente necessário coletar e reter informações sobre raça e gênero para balancear uma ferramenta de triagem de emprego que está contratando apenas candidatos brancos do sexo masculino devido ao viés inerente ao conjunto de dados de treinamento original. Embora isso pareça contraditório com a compreensão tradicional da minimização de dados, na realidade, é necessário ter mais dados, em alguns casos, para reduzir o risco. No exemplo acima, coletar e manter informações sensíveis é algo que está de acordo com o princípio de minimização de dados, porque, sem esses dados, a ferramenta de triagem de emprego produziria potencialmente decisões tendenciosas (ou seja, injustas). Portanto, a minimização de dados não deve necessariamente significar menos dados, e sim que os dados coletados e retidos são necessários com relação aos propósitos do processamento.

"Distinguir entre a etapa de treinamento e de implementação para fins de minimização de dados poderia ajudar a equilibrar a inovação com o estímulo a uma melhor proteção de dados para os indivíduos."

Reutilizar dados antigos com novos propósitos

Durante as mesas-redondas do CIPL que examinaram essa questão, os participantes observaram que os dados às vezes podem se tornar menos valiosos à medida que envelhecem, seja perdendo sua precisão ou relevância. Porém, em certos setores ou para determinados fins, dados antigos são inestimáveis, seja para identificar tendências ou treinar algoritmos. No setor financeiro, por exemplo, dados antigos podem revelar padrões e identificar tendências que eram desconhecidas no momento da coleta, e isso pode ser particularmente útil para a prevenção de fraudes. No setor de saúde, os dados antigos podem ser úteis no treinamento de algoritmos para ler exames ou prontuários médicos para detectar padrões e aprender mais sobre a prevenção de doenças. Isso também pode ajudar a prever como novos medicamentos em potencial, incluindo medicamentos projetados por IA, se comportarão no corpo humano. [39] Analisar dados antigos usando novas ferramentas de IA pode criar inúmeras oportunidades e benefícios e, em muitos desses casos, o uso futuro dos dados são agnósticos em relação à identidade do indivíduo.

Encontrar o equilíbrio adequado de interesses entre a proteção de dados individuais e as vias para reutilizar dados antigos é um desafio para as autoridades reguladoras e para as organizações. Esse desafio exigirá **flexibilidade e interpretações razoáveis do princípio de minimização de dados**, muitas vezes distinguindo-os entre os vários contextos e propósitos para armazenamento de dados.

Diferenciando dados para as etapas de treinamento e implementação

A diferenciação entre dados usados para treinar IA e dados para implantar IA é muito útil neste contexto. Embora outras ferramentas de responsabilização serão necessárias para governar a fase de treinamento, distinguir entre a etapa de treinamento e de implementação para fins de minimização de dados poderia ajudar a equilibrar a inovação com o estímulo a uma melhor proteção de dados para os indivíduos. Ao limitar o uso de dados na fase de implementação, mas fornecendo mais flexibilidade no uso de dados na etapa de treinamento, as organizações estão gerenciando o prejuízo potencial aos indivíduos e, assim, mantendo a intenção original do princípio de minimização de dados. Isso não sugere que os dados sejam desnecessários na etapa de implementação; na verdade, podem ser fundamentais para otimizar o desempenho algorítmico e monitorar os resultados. Entretanto, o risco de dano aos indivíduos é diminuído durante a fase de treinamento, portanto os padrões de uso de dados durante essa fase podem ser examinados de forma menos estrita. Outra forma potencial de limitar o uso de dados pessoais é usar dados sintéticos (isto é, um repositório de dados gerado por programação), quando estiverem disponíveis e se forem razoáveis economicamente, para treinar o modelo de IA.

Demonstrando a relevância dos dados

Outra interpretação razoável desse princípio avaliaria como permissível a demonstração da minimização de dados com relação a um sistema de IA articulando e documentando de forma proativa a necessidade de coletar e processar dados (sejam dados antigos ou dados que não aparentem ser estritamente necessários ao propósito do processamento), e o que se espera aprender ou conseguir através do processamento dos dados. Isso seria particularmente útil para a fase de treinamento, embora possa ser útil tanto para essa fase quanto para a de implementação. A determinação do que é adequado, relevante e necessário dependerá do contexto, mas essa avaliação proativa e contínua servirá para demonstrar que os dados a serem coletados são relevantes e não excessivos com relação ao propósito de processamento.

Minimizando riscos através de ferramentas tecnológicas

Através do ciclo de vida de uma aplicação de IA, as organizações podem considerar implantar uma variedade de ferramentas para minimizar o risco aos indivíduos. Ferramentas tecnológicas para ajudar na minimização de dados ainda estão em estágio inicial de desenvolvimento, e muitas vezes são dispendiosas para serem implantadas pelas organizações menores, mas sua exploração contínua deve ser incentivada.[40] Por exemplo, em alguns casos, uma aprendizagem unificada poderia capacitar os algoritmos de IA a aprender sem que os dados saíssem jamais do dispositivo, e sem a necessidade de centralizar grandes quantidades de dados em um único local virtual. As organizações também podem considerar a possibilidade de anonimizar ou pseudonimizar os conjuntos de dados, embora isso possa trazer outros desafios próprios.[41] Ao mesmo tempo, embora mais esforços de pesquisa e desenvolvimento futuros sejam necessários para assegurar uma desidentificação adequada, uma interpretação flexível de noções de anonimato ou uso de pseudônimos contribuiria muito para permitir o uso de dados para treinamento de IA e para reduzir os riscos de conformidade para as organizações.

III. Soluções à nossa frente

Este segundo relatório pesquisou algumas das tensões mais difíceis entre IA e proteção de dados e explorou formas de mitigar essas tensões. Através das conversas e mesas-redondas que examinaram essas questões, surgiram seis temas primordiais. Eles atravessam transversalmente as questões já discutidas e oferecem guias úteis para desenvolver, implementar e avaliar soluções práticas para o uso responsável de tecnologia de IA.

"A regulamentação específica para IA pode obstruir a inovação e a criação de IA e de melhores práticas, a menos que permita que organizações de IA responsáveis experimentem, aprendam e cresçam. Onde for inevitável ter regulamentações para IA, esta deve ser cuidadosamente desenvolvida e com tempo suficiente para permitir que uma série de partes interessadas identifiquem, articulem e implementem princípios-chave e melhores práticas."

A. A necessidade de soluções neutras em tecnologia

A maioria dos desafios de proteção de dados identificados no contexto da IA é anterior à IA e é apresentada por outras tecnologias que não a IA. Resumindo, esses desafios são maiores e mais amplos do que a IA, portanto é importante que as soluções também o sejam. Soluções específicas de IA podem não apenas ser limitadas demais, como podem também abordar um sintoma sem resolver o problema subjacente. Por exemplo, o desconforto que pode resultar de tomadas de decisão automatizadas usando IA provavelmente não será resultado da própria tecnologia, mas do fato de que uma máquina está tomando uma decisão significativa que poderia impactar negativamente um indivíduo, ou mesmo criar efeitos legais para os indivíduos. Embora a IA possa agravar essas questões, por exemplo, dando poder às máquinas para que tomem mais decisões que afetam indivíduos, o problema a ser abordado não é a IA, e sim o papel da tomada de decisões não humana, especialmente quando se trata de decisões significativas. O tipo de tecnologia é praticamente irrelevante: a fonte do desconforto ou desconfiança é o impacto da decisão tomada por essa tecnologia. Portanto, a solução deve se focar no problema, e não na tecnologia.

Uma abordagem por camadas para a regulamentação da IA

À medida que os países e regiões considerem a regulamentação relacionada a IA,[42] é importante compreender que as estruturas ou regulamentações legais específicas para IA poderiam falhar em resolver a questão subjacente, enquanto ao mesmo tempo negam potencialmente à sociedade os benefícios de uma IA adequadamente implementada. Isso também potencialmente negaria às sociedades os benefícios da IA que não envolvessem tomadas de decisões automatizadas. Além disso, qualquer tipo de regulamentação que não seja neutra em tecnologia pode se sobrepor a regulamentações (horizontais) já existentes ou duplicá-las, o que seria prejudicial para a segurança jurídica. A regulamentação específica para IA pode obstruir a inovação e a criação de IA e de melhores práticas, a menos que permita que organizações de IA responsáveis

experimentem, aprendam e cresçam. Onde for inevitável ter regulamentações para IA, esta deve ser cuidadosamente desenvolvida e com tempo suficiente para permitir que uma série de partes interessadas identifiquem, articulem e implementem princípios-chave e melhores práticas.[43]

Na medida em que a regulamentação de IA for finalmente introduzida, o CIPL acredita que os legisladores devem abordar a legislação sobre IA com relação a dois princípios essenciais:

- **Construindo com base nas estruturas existentes**—incluindo planos horizontais e leis setoriais específicas — que já fornecem as estruturas de parâmetro, os requisitos, as ferramentas e soluções para governança e uso responsáveis da IA.
- **Adotando uma abordagem reguladora baseada em princípios e em resultados** que seja capaz de se adaptar à variedade e à natureza de rápida evolução de todas as tecnologias relacionadas à IA e aos desafios singulares de indústrias específicas, que evite regras excessivamente rígidas e prescritivas e que permita às organizações operacionalizar esses princípios desenvolvendo práticas responsáveis e baseadas em risco que alcancem os resultados identificados.
- **Fazendo um "teste de balanceamento riscos/benefícios" e uma avaliação de impacto contextual** ferramentas-chave para sustentar o uso benéfico da IA, evitar a reticência a riscos e capacitar uma mitigação de riscos adequada.

Além disso, o CIPL apoia uma abordagem regulatória em camadas para a IA, o que, no contexto de proteção de dados, significa:

- Construir a partir de leis existentes de proteção de dados e tornar essas leis um facilitador de IA através de pensamento prospectivo e interpretação progressiva das exigências feitas por autoridades de proteção de dados;
- Alavancar e incentivar práticas de IA responsáveis nas organizações;
- Fomentar abordagens inovadoras de fiscalização regulamentar (isto é, sandboxes e hubs regulatórios em que reguladores de diferentes disciplinas com interesse em IA possam trocar pontos de vista, resolver conflitos por questões legais, etc.)

Ferramentas e soluções neutras em tecnologia

Onde a regulamentação da IA não for adotada, e no prazo imediato, cumprir as metas de melhorar a proteção de dados exigirá ferramentas e soluções neutras em tecnologia que possam ser aplicadas em uma variedade de situações e contextos.

As soluções neutras em tecnologia ajudarão a atender às metas de proteção de dados de uma maneira mais holística. Como recentemente apontado pela Plataforma para o Relatório da Sociedade da Informação: "[A] maior parte da IA não funciona de forma independente: faz parte de um produto ou serviço."[44] As ferramentas neutras em tecnologia servirão para melhorar o produto, processo ou serviço como um todo, em vez de um segmento de um contexto mais amplo. Os Resultados de Justiça da FCA do Reino Unido, discutidos acima, são um exemplo de como as ferramentas neutras em tecnologia podem ajudar a facilitar o comportamento responsável geral das organizações. As organizações não podem ser dispensadas de sua responsabilidade ao mudar suas tecnologias; elas devem alcançar os resultados da justiça - e outros princípios de proteção de dados - independentemente da tecnologia ou processo utilizado.

B. A importância do processo

"A meta não é determinar se uma aplicação de IA em particular está em conformidade e é justa em um determinado momento, e sim saber se todas as aplicações estão sendo examinadas e monitoradas de forma regular. Portanto, um foco chave tanto das organizações quanto das autoridades reguladoras deveria ser o desenvolvimento, avaliação e melhoramento dos processos para isso."

As ferramentas de IA estão sendo aplicadas amplamente e aproveitam as tecnologias que estão sendo desenvolvidas em ritmo acelerado. Portanto, abordagens no sentido de resolver os desafios de proteção de dados que podem surgir não apenas precisam ser neutras em tecnologia, mas também mais focadas nos processos de tomada de decisões e reparação. Quais são os processos que uma organização ou um regulador podem empregar para garantir que o processamento de dados, qualquer que seja a tecnologia usada, seja responsável e que quando ocorrerem erros, pois inevitavelmente surgirão, sejam detectados e remediados rapidamente? Essa questão é especialmente importante dada a escala na qual esses processos vão precisar operar. A meta não é determinar se uma aplicação de IA em particular está em conformidade e é justa em um determinado momento, e sim saber se todas as aplicações estão sendo examinadas e monitoradas de forma regular com o objetivo fundamental de melhoria contínua e mitigação de risco. Portanto, um foco-chave tanto das organizações quanto das autoridades reguladoras deveria ser o desenvolvimento, avaliação e melhoramento dos processos para isso.

Os processos são úteis e necessários nas etapas de projeto, desenvolvimento e implementação de IA. Por exemplo, a decisão da Axon[45] de não usar reconhecimento facial nas câmeras dos coletes da polícia foi o resultado de um processo de revisão de ética implementado na etapa de pesquisa do desenvolvimento de produto. A tecnologia não foi implantada devido ao descobrimento de preocupações éticas com relação a viés e imprecisão que não podiam ser mitigadas satisfatoriamente. Esse é um exemplo potente do valor dos Comitês de Revisão de Dados, discutidos mais detalhadamente a seguir, para ajudar a garantir não apenas conformidade legal, mas que o uso de uma ferramenta de IA seja continuamente responsável, adequada e consistente com os valores de uma instituição.

Em casos em que a tecnologia é implantada, serão necessários processos para remediar decisões equivocadas, fornecer transparência e garantir justiça. Quanto mais crítico for o impacto da decisão, maior a necessidade de processos imediatos ou instantâneos para repará-la.

Uma forma em que os reguladores de proteção de dados podem melhorar os processos pode ser através do envolvimento com líderes e engenheiros de IA na indústria e com os governos para desenvolver conjuntamente orientações baseadas em resultados e usar os casos que podem, a seguir, ser incorporados nos processos organizacionais. Da mesma forma que o Grupo de Especialistas de Alto Nível da UE sobre as Orientações de IA[46], ou a Estrutura de Governança de IA Modelo de Cingapura,[47] autoridades reguladoras que fornecem orientações para os processos organizacionais podem fomentar a implementação de IA responsável e responsabilizável ao mesmo tempo em que permite inovação tanto na tecnologia quando nos processos usados para conseguir a proteção de dados. Outro exemplo notável vem do UK ICO, que usou uma metodologia inovadora e envolvente para desenvolver sua Estrutura de Auditoria de IA publicando uma série de blogs, incentivando comentários de grupos multifuncionais de especialistas e envolvendo especialistas em tecnologia de IA para trabalhar nas soluções.[48]

"Os processos serão necessários para remediar decisões equivocadas, fornecer transparência e garantir justiça. Quanto mais crítico for o impacto da decisão, maior a necessidade de processos imediatos ou instantâneos para repará-la."

Para praticamente todas as questões enfocadas neste relatório há ferramentas tecnológicas e procedimentais que podem mitigar tensões entre princípios de proteção de dados e tecnologia de IA, e processos robustos para implementar essas ferramentas são fundamentais. O desenvolvimento dos processos adequados ao longo do ciclo de vida dos produtos de uma forma multifuncional e que abranja toda a organização ajuda a promover o desenvolvimento de IA centrada em humanos e construir confiança em sistemas de IA, além de geralmente ajudar as organizações a se tornarem melhores administradoras de dados.

O importante aqui é que as abordagens mais bem-sucedidas para lidar com os desafios de proteção de dados apresentados pela IA até o momento focaram-se não nas determinações sobre tecnologias ou aplicações específicas, mas em processos contínuos para identificar, prevenir e mitigar impactos prejudicais. Esses processos servem como salvaguardas e são cada vez mais necessárias ao longo do ciclo de vida de produtos ou serviços para garantir justiça, transparência e outras metas da proteção de dados.

C. Uma abordagem baseada em riscos para a IA

Ao acessar os desafios de proteção de dados apresentados pelas aplicações de IA, é útil e, em verdade, consistente com as expectativas da maioria dos indivíduos considerar o impacto potencial e quaisquer riscos de prejuízos do processamento proposto para os indivíduos, assim como o risco de não usar informações. Essa abordagem baseada em riscos foi sugerida pela Estrutura de IA Modelo de Cingapura,[49] pelo RGPD[50] e, mais recentemente, pelo Guia para Regulamentação de Aplicações de IA do Escritório de Administração e Orçamento (OMB) dos EUA.[51] Os usos de IA que apresentem pouco risco de prejuízo aos indivíduos, seja porque as decisões tomadas não são relevantes ou porque a probabilidade de um resultado prejudicial é remota, podem, justificadamente, garantir um menor escrutínio. O uso de IA para recomendar canções ou filmes, por exemplo, muito provavelmente demanda menos atenção do que aplicações de IA usadas em carros para evitar bater em pedestres ou outros veículos.

Benefícios-chave da abordagem baseada em riscos

O foco sobre impactos e riscos para os indivíduos não diminui a obrigação de estar em conformidade total com a lei de proteção de dados, mas pode ajudar a determinar a alocação dos escassos recursos das organizações e autoridades reguladoras:

- Pode ajudar a garantir que seja dada atenção adequada a aqueles usos dedados que apresentam maiores riscos;
- Pode ajudar a justificar o uso de processos de mitigação mais penosos ou dispendiosos quando os resultados prejudiciais em potencial o demandam; e
- Pode ajudar a determinar as medidas de precaução ou reparação que deveriam estar ativas.

Enquanto os princípios tradicionais de proteção de dados servem à meta de limitar os impactos dos dados para os indivíduos, novas tecnologias fazem com que seja cada vez mais crítico considerar cada caso de uso específico e avaliar os impactos do processamento de dados. "A natureza da aplicação de IA e o contexto em que é usado define em grande medida quais trocas devem ser feitas em um caso específico... as aplicações de IA no setor médico em parte levarão a diferentes questões e áreas de preocupação comparado com aplicações de IA para operações logísticas."[52] Essas trocas podem variar por setor, mas variam com mais precisão devido a seu impacto sobre os indivíduos. Por exemplo, o impacto prospectivo do uso de dados para treinar IA é inferior ao impacto do uso de dados para tomar uma decisão, e os processos implantados para proteger os indivíduos precisam refletir essa diferença.

Exemplo de como analisar riscos e impactos de um uso de dados proposto

Embora haja diversas formas de conceber uma avaliação de riscos e impactos, a Figura 2 abaixo apresenta uma forma de analisar um uso de dados proposto. Os dois fatores primários são 1) a sensibilidade dos dados e 2) o nível de impacto sobre os indivíduos a partir do uso desses dados. Esses fatores podem ajudar as organizações a determinar o nível de processo necessário com base em um contexto em particular. Embora a avaliação ilustrada pela Figura 2 possa se provar útil para determinar o nível de processo necessário, ela deve caminhar junto a uma avaliação de risco mais holística, já que o risco pode depender de muitos outros fatores não associados com a natureza dos dados.

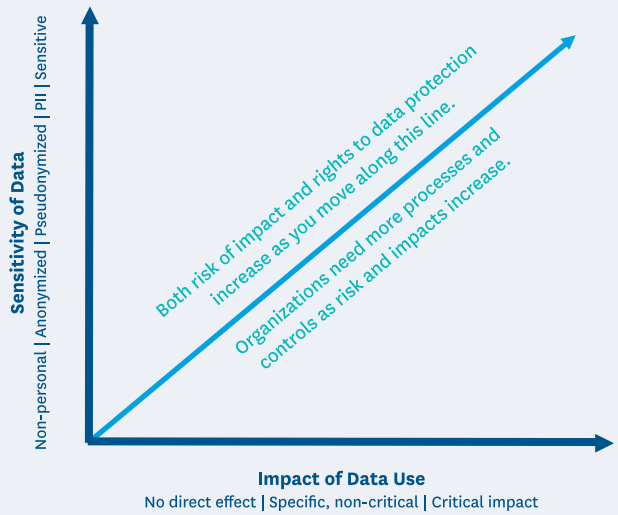

Figura 2. *O gráfico acima demonstra uma conceitualização de como analisar riscos e impactos de um uso de dados proposto. Essa é uma representação de como implementar uma solução que seja neutra em tecnologia, focada em impacto e orientada a processos.*

Questões relevantes para considerar os impactos incluem:

- *Qual é o objetivo de usar IA para esta aplicação?*
- *Os dados são usados para treinar um algoritmo ou para implementá-lo*
- *O algoritmo está tomando uma decisão ou fazendo uma recomendação?*

À medida que aumenta a sensibilidade dos dados e o impacto da tomada de decisões, as organizações deveriam estar desenvolvendo processos e controles adicionais para limitar impactos prejudiciais.

"Há uma necessidade de seguir desenvolvendo uma melhor compreensão dos prejuízos, particularmente os prejuízos não materiais em potencial que podem ocorrer ao coletar e processar dados. Analisar o risco de implementar novos modelos significa compreender como avaliar e medir o prejuízo e sua probabilidade de se materializar nesses novos contextos, desde prejuízos monetários a prejuízos não físicos, como privacidade, segurança e impactos discriminatórios, entre outros."

Nossos debates revelaram que muitas organizações estão desenvolvendo ferramentas para compreender e mitigar os impactos e prejuízos aos indivíduos que partem de aplicações específicas de IA. Elas vão desde as DPIAs, como exigido pelo RGPD, até ferramentas específicas de Avaliação de Impacto de IA. Algumas organizações relatam que elas conseguiam alavancar uma DPIA existente como ferramenta e processo útil sobre a qual trabalhar e com a qual realizar uma avaliação mais ampla de impacto de IA.

Uma importante faceta de enfatizar os impactos de dados e conduzir DPIAs ou Avaliações de Impacto de IA é desenvolver uma estrutura para avaliar de forma mais precisa e consistente o impacto ou prejuízo de um uso de dados em particular. Há uma necessidade de seguir desenvolvendo uma melhor compreensão dos prejuízos, particularmente os prejuízos não materiais em potencial que podem ocorrer ao coletar e processar dados. Analisar o risco de implantar novos modelos requer compreender como avaliar e medir o prejuízo e sua probabilidade de se materializar nesses novos contextos, desde prejuízos monetários até prejuízos não físicos, como privacidade, segurança e impactos discriminatórios, entre outros.".[52] O CIPL escreveu detalhadamente sobre essa questão como parte do seu trabalho sobre mitigação de risco em proteção de dados.[54]

Benefícios do uso de dados e riscos de reticência

Embora um dos principais objetivos de uma avaliação de impacto de IA seja avaliar o risco ou prejuízo de um uso de dados específico, qualquer avaliação de impacto deve também incluir um processo para balancear os riscos, comparando com os benefícios concretos do processamento de dados proposto. Poderia haver grandes riscos relacionados a um sistema específico de IA que poderiam ser superados por imensos benefícios aos indivíduos e sociedade em geral. Por exemplo, a IA fornece imensos benefícios quando usada para monitorar conteúdo nas plataformas online para evitar o terrorismo, o abuso de crianças ou outros comportamentos criminais, benefícios que poderiam pesar mais que os riscos associados com o processamento dos dados pessoais relevantes.

Além disso, ao avaliar adequadamente o impacto da IA e balancear os benefícios e riscos, os assim chamados "riscos de reticência" (isto é, as consequências para os indivíduos e para a sociedade de não prosseguir com um projeto específico relacionado a IA devido aos riscos potenciais) deveriam também fazer parte da avaliação para garantir que todos os fatores relevantes sejam considerados e que informem a decisão final.

D. A necessidade de administração de dados e de responsabilização organizacional

"O desenvolvimento de uma abordagem de proteção de dados orientada a impacto e processos necessariamente vai exigir que as organizações se tornem melhores administradoras de dados."

O rápido e abrangente desenvolvimento de novas tecnologias, incluindo de IA, criou uma necessidade renovada de uma maior responsabilização organizacional e maior administração de dados. O desenvolvimento de uma abordagem de proteção de dados orientada a impacto e processos necessariamente vai exigir que as organizações se tornem melhores administradoras de dados. Isso incluirá a necessidade de gestão de riscos organizacionais, processos melhorados e mais transparência. A administração de dados (*data stewardship*) pode ser alcançada através de um número de práticas e ferramentas, mas geralmente vai ajudar a instaurar práticas responsáveis em desenvolvimento e implementação de IA.

Um aspecto importante da administração de dados envolverá que as organizações desenvolvam princípios e valores em torno da IA. Uma administração de dados responsável começa com a liderança superior de uma organização. Conforme observado pela McKinsey Analytics, "Os CEOs devem esclarecer exatamente quais são as metas e valores da empresa em vários contextos, solicitar às equipes que articulem valores no contexto de IA e incentivar um processo colaborativo para a escolha das métricas."[55] Quando a liderança articula as metas e valores que a organização está procurando manter, ela permite que a administração de dados faça parte da cultura corporativa.

Um foco melhorado na administração de dados e na responsabilização organizacional é particularmente necessário no contexto de IA. Isso ocorre por causa dos desafios de se fornecer aos indivíduos divulgações significativas a respeito de ferramentas e algoritmos de IA que são difíceis de entender até mesmo pelos especialistas. Embora um foco na administração não elimine a necessidade de divulgação e transparência, ele ajuda a reconhecer que as organizações têm uma obrigação de tomar decisões mais cuidadosas e de assumir maior

"Uma administração de dados responsável começa com a liderança superior em uma organização."

"A reparação provavelmente passe a adotar uma nova importância na governança efetiva de IA... devemos nos esforçar para garantir que, particularmente no contexto de tomadas de decisão automatizadas com impacto legal ou similarmente significativo, os indivíduos tenham vias efetivas e eficientes para contestar resultados e apelar de decisões."

responsabilidade pelas consequências de produtos, serviços e tecnologias que estão desenvolvendo em situações nas quais os indivíduos são menos capazes de tomar decisões informadas por si mesmos.

E. Focando em papeis significativos para os seres humanos

Algumas leis de proteção de dados parecem contemplar seres humanos como um freio ou restrição à automação, ao menos quando a automação é usada para tomar decisões que afetam os indivíduos legalmente ou de forma igualmente significativa. No caso da IA, isso faz com que se corra o risco de nunca ir além da capacidade que as mentes humanas têm de serem justas, consistentes e racionais. Deveríamos ser mais ambiciosos e aspirar a que a IA alcance mais do que os cérebros humanos conseguem, mas isso vai exigir considerar papeis mais amplos para os seres humanos dentro de todo o ciclo de vida de uma aplicação de IA, ou seja, do seu desenvolvimento até sua implementação. Esse envolvimento humano deve ser significativo e provavelmente deva incluir supervisão humana do processo de projeto, de desenvolvimento e de implementação, sem importar se está se construindo algoritmos, avaliando a qualidade dos dados ou testando resultados, [56] assim como deve incluir supervisão humana sobre o processo de reparação. Acima de tudo, precisamos garantir que esse papel de envolvimento humano vá além de simplesmente verificar itens de conformidade.

A reparação provavelmente assumirá uma nova importância na governança efetiva da IA, e portanto deverá exigir uma atenção renovada. Até mesmo com os controles adequados e as restrições a algoritmos, nunca poderemos alcançar todo o potencial da IA ao mesmo tempo em que evitamos todos os maus resultados, ou até mesmo todos os prejuízos. Mais do que ver o risco potencial como um motivo para escapar dessas novas tecnologias, devemos nos esforçar para garantir que, particularmente no contexto de tomadas de decisão automatizadas com impacto legal ou similarmente significativo, os indivíduos tenham vias efetivas e eficientes para contestar resultados e apelar de decisões.[57] Isso não apenas preservará a proteção de dados, como também outros aspectos da dignidade humana.

F. Ampla gama de ferramentas disponíveis

Há uma ampla gama de ferramentas disponíveis para organizações que estão tentando melhorar os processos relativos a desenvolvimento e implementação de IA. Essas ferramentas podem ajudar as organizações a facilitarem a responsabilidade e responsabilização em sua abordagem de novas tecnologias e de novos usos de dados. As organizações continuam inovando e melhorando os processos para criar novos métodos de manter a proteção de dados, sendo que as ferramentas listadas abaixo refletem algumas das melhores práticas atuais para usuários responsáveis de dados.

"A Roda de Responsabilidade do CIPL vem sendo usada para promover responsabilização organizacional no contexto da construção, implementação e demonstração de programas de privacidade abrangentes. Essa estrutura também pode ser usada para ajudar as organizações a desenvolverem, implantarem e organizarem medidas robustas e abrangentes de proteção de dados no contexto de IA e também demonstrar a responsabilização em IA."

1. A Roda de Responsabilidade do CIPL:

A Roda de Responsabilidade do CIPL vem sendo usada para promover responsabilização organizacional no contexto da construção, implementação e demonstração de programas de privacidade abrangentes. Essa estrutura também pode ser usada para ajudar as organizações a desenvolverem, implantarem e organizarem medidas robustas e abrangentes de proteção de dados no contexto de IA e também a demonstrar a responsabilização em IA. A Roda de Responsabilidade fornece uma arquitetura uniforme com sete elementos para que as organizações desenvolvam e demonstrem sua responsabilidade: Liderança e supervisão; Avaliação de riscos; Políticas e procedimentos; Transparência; Treinamento e consciência; Monitoramento e verificação; e Resposta e aplicação das normas.[58]

Os esforços organizacionais para promover a confiabilidade dentro da IA podem mapear esta roda para garantir uma abordagem holística, posto que cada elemento fornece proteções importantes para os indivíduos. A tabela do Apêndice B detalha exemplos de algumas das práticas existentes que os membros do CPIL estão desenvolvendo e implementando para promover responsabilização organizacional à medida que desenvolvem e implantam tecnologias de IA. Embora a lista de medidas no Apêndice B não seja de forma alguma um conjunto obrigatório de exigências, ou uma lista totalmente exaustiva de exemplos, ela serve como um ponto de partida útil para que as organizações desenvolvam seus programas de conformidade relativo à privacidade e novas políticas para desenvolvimento, implementação ou uso de IA. Também é uma ferramenta útil de avaliação para que organizações já estabelecidas verifiquem se suas práticas atuais são abrangentes e eficientes.

2. Avaliações de Impacto de Proteção a Dados de IA (IA DPIAs):

Uma das formas mais comuns de avaliar o impacto do uso de dados proposto é através de uma DPIA, que é necessária segundo o RGPD para uma decisão automatizada que produza efeitos legais ou outros igualmente significativos. Muitas organizações hoje usam DPIAs para se ajustar à proteção de dados e demonstrar sua conformidade. Alguns decidiram usar DPIAs em um contexto ainda mais amplo do que aquele requerido por lei, parcialmente para fomentar a privacidade desde a concepção e mitigação de riscos, e parcialmente para estabelecer um léxico e metodologia comuns para avaliar usos de dados em diferentes departamentos e geografias. Essas avaliações podem ter valor adicional no contexto da IA, e algumas organizações estão desenvolvendo DPIAs específicas para IA, seja como um suplemento das avaliações requeridas segundo o RGPD ou como uma avaliação inteiramente separada.

As DPIAs de IA (também chamadas de Avaliações de Impacto de IA ou Avaliações de Impacto Algorítmico) podem fornecer uma abordagem estruturada para que as organizações avaliem questões de justiça, direitos humanos ou outras considerações nessas novas tecnologias. As orientações recentemente lançadas pelo UK ICO sobre DPIAs no contexto de IA devem ser um documento vivo e parte de um processo organizacional contínuo. A Estrutura Modelo de Cingapura, embora não fale especificamente das DPIAs, também enfatiza o benefício das avaliações de risco contínuas para "desenvolver clareza e confiança no uso de soluções de IA", assim como para ajudar a "responder a desafios potenciais por parte dos indivíduos, de outras organizações ou empresas e de autoridades reguladoras".

Essas avaliações podem ajudar as organizações a construir valores corporativos em seus processos, e eventualmente criarão uma estrutura de cenários "pré-aprovados" que podem orientar futuras avaliações. Embora esses tipos de avaliações estejam em seu desenvolvimento inicial, eles têm o potencial de promover justiça, impulsionar a responsabilização e criar consistência. Adicionalmente, as DPIAs de IA podem ajudar as organizações a desenvolverem a documentação necessária para fornecer transparência eficaz para indivíduos e autoridades reguladoras.

3. Comitês de Revisão de Dados (CRDs):

Os Comitês de Revisão de Dados são outra ferramenta potencial para que as organizações estruturem a forma de conduzir o balanceamento de interesses entre o impacto dos usos de dados e as novas aplicações de IA. Eles também entram na categoria de "Liderança e supervisão" na Roda de Responsabilidade acima. Da mesma forma que com as DPIAs de IA, os Comitês de Revisão de Dados (CRDs) podem ajudar as organizações a responderem a novas tecnologias e desenvolver precedente para as tecnologias futuras.[63] Diversas organizações já têm ou estão considerando CRDs, éticas internas ou externas similares ou comitês de IA como forma de garantir uma abordagem centrada em humanos para tomadas de decisões relacionadas a aplicações de IA e novos usos de dados. Os CRDs podem ajudar a impulsionar a responsabilização organizacional, estimular tomadas de decisões conscientes e garantir que os novos usos de dados carreguem consigo os valores corporativos e da sociedade.

Apesar de que a estrutura, os procedimentos e a função de cada CRD serão diferentes, há algumas melhores práticas que podem ajudar as organizações a garantirem a efetividade de sua operação. Por exemplo, garantir que os indivíduos sejam incluídos na composição do comitê, independente do projeto de IA que está sendo examinado e com uma gama de perspectivas externas. Esses indivíduos, idealmente, seriam externos à organização também. Isso garante que, ao examinar um projeto proposto de IA, os especialistas externos estejam desapegados de interesses comerciais e possam fornecer uma análise significativa das questões. Outras coisas a considerar incluem assegurar-se de que a gestão esteja fornecendo suporte adequado ao CRD, dando a ele um papel estrutural no processo de avaliação e criando ferramentas para garantir que ele siga um processo estruturado de uma forma eficiente, transparente e relativamente rápida. É claro que deve haver procedimentos adequados para proteger e se responsabilizar por segredos confidenciais e comerciais das aplicações de IA examinados por qualquer especialista externo.

Criar um CRD exigirá que uma organização considere e documente valores organizacionais que o CRD ficará encarregado de manter. Isso propiciará uma melhor tomada de decisões e responsabilização sobre muitas das tensões importantes entre IA e proteção de dados. Por exemplo, os CRDs podem ser úteis para avaliar e assegurar justiça assim como para avaliar os riscos e impactos de novos usos ou propósitos dos dados. As DPIAs de IA podem ser uma ferramenta usada pelos CRDs durante sua avaliação de novos usos de dados, ou podem ser o que dispara uma questão a ser enviada para o CRD para maiores considerações. Eles também podem ser consultados como parte do próprio processo de DPIA e podem fornecer suas visões sobre o risco de um processamento em particular (ver, por exemplo, o Artigo 35(9) do RGPD solicitando ao controlador para procurar as visões dos titulares dos dados ou de seus representantes quando adequado; da mesma forma, o CRD pode fornecer uma perspectiva diferente e externa).

4. Vias de reparação

Muitas das preocupações com relação à justiça para os consumidores podem ser abordadas, em grande medida, fornecendo-se uma reparação rápida e eficaz através da responsabilização organizacional. A reparação permite que os indivíduos contestem e modifiquem um resultado que eles acreditem ser impreciso, injusto ou de alguma outra forma inadequado.[64] Dependendo das circunstâncias e do impacto da decisão, a velocidade e a natureza da reparação irão diferir. Por exemplo, se uma máquina estiver verificando cartões de embarque para permitir que os passageiros acessem um voo e impede que um indivíduo embarque, será necessária uma reparação eficaz quase instantânea. Decisões mais triviais podem não precisar ser explicadas instantaneamente, e a reparação poderia ser tão simples quanto usar um email ou fazer login em uma plataforma para lançar a revisão da decisão. Em qualquer um dos casos, o caminho de reparação deve ser visível e estar acessível aos indivíduos afetados pela decisão.

A reparação muitas vezes ocorre na forma de revisão humana, mas há outras formas de reparação que podem ser úteis. Por exemplo, muitos smartphones e laptops estão usando dados biométricos para reconhecer os usuários autorizados. Se essa tecnologia não estiver funcionando adequadamente, os consumidores geralmente têm outra forma de escapar dela - normalmente fornecendo uma senha que o usuário já tinha programado anteriormente.

Quando as organizações estão desenvolvendo novas tecnologias e considerando seu impacto, é improvável que possam prever e limitar todos os impactos negativos. Em alguns casos, uma organização pode determinar que os riscos são elevados demais para implementar a tecnologia. Porém, as compensações em outros contextos podem garantir que a tecnologia seja implementada, mas que tenha vias visíveis e efetivas para corrigir situações em que ocorrem tomadas de decisões incorretas ou com viés. As organizações devem garantir que a reparação seja significativa, e que não apenas se torne um carimbo em uma decisão automatizada. Se forem descobertas injustiças ou imprecisões, as organizações devem ter processos instaurados para limitar situações semelhantes no futuro. Considerar e desenvolver essas remediações e processos será parte essencial da implementação de IA, e as autoridades reguladoras que avaliarem o uso da IA e o impacto na proteção de dados devem buscar essas vias visíveis de reparação como uma forma de demonstrar uma implementação responsável de tecnologias de IA.

IV. Conclusão

As tecnologias de IA, o volume e a variedade de ferramentas de IA e a velocidade com a qual estão evoluindo e sendo implantadas apresentam muitos desafios para a proteção de dados. A IA pode incluir decisões automatizadas, mas também pode incluir aumentar a inteligência humana para produzir melhores resultados. Os inúmeros benefícios produzidos pela proliferação de tecnologias de IA não ocorrem sem desafios. Entretanto, um ano de mesas-redondas, debates e pesquisas claramente mostraram que há tanto suficiente flexibilidade na maioria de leis de proteção de dados quanto suficiente criatividade entre as organizações e autoridades reguladoras para estar em conformidade com essas leis e garantir que a IA seja desenvolvida e implementada de formas que não sejam meramente legais, mas também benéficas e responsáveis.

Se você deseja ampliar o debate sobre este trabalho, ou solicitar informações adicionais, entre em contato com Bojana Bellamy, bbellamy@HuntonAK.com; Markus Heyder, mheyder@HuntonAK.com; Nathalie Laneret, nlaneret@HuntonAK.com; Sam Grogan, sgrogan@HuntonAK.com; Matthew Starr, mstarr@HuntonAK.com ou Giovanna Carloni, gcarloni@HuntonAK.com.

Apêndice A. *Traduções de Justiça* – A tabela a seguir fornece a tradução de *fairness* ("justiça", no termo original em inglês) conforme declarado no Artigo 5, em cada um dos 23 idiomas com uma tradução oficial do RGPD.

Idioma	Termo usado no Artigo 5	Tradução do Google
Búlgaro	добросъвестност	Boa fé
Croata	poštenosti	Honestidade
Tcheco	korektnost	Correção, propriedade
Dinamarquês	rimelighed	Razoavelmente
Holandês	behoorlijkheid	Bondade
Inglês	fairness	Justiça
Estoniano	õiglus	Justiça, equidade, retidão
Finlandês	kohtuullisuus	Justiça, equidade, moderação
Francês	loyauté	Lealdade, confiabilidade, fidelidade
Gaélico	cothroime	Justiça
Alemão	Verarbeitung nach Treu und Glauben	Processamento com boa fé
Grego	αντικειμενικότητα	Objetividade
Húngaro	tisztességes eljárás	Procedimento justo, jogo limpo
Italiano	correttezza	Correção, exatidão, justiça, propriedade, honestidade
Letão	godprātība	Boa fé, integridade, honestidade
Lituano	sąžiningumo	Honestidade, integridade, justiça, boa fé
Maltês	ġustizzja	Justiça
Polonês	rzetelność	Confiança, confiabilidade, honestidade, retidão, rigidamente convencional
Romeno	echitate	Equidade, justiça, integridade
Eslovaco	spravodlivosť	Justiça, equidade, retidão, integridade, via estreita, virtuosidade
Esloveno	pravičnost	Justiça
Espanhol	Lealtad	Lealdade, aliança, devoção
Sueco	korrekthet	Correção, propriedade

Apêndice B. *Mapeando melhores práticas em governança de IA para a Roda de Responsabilidade do CIPL*
Através das mesas-redondas, as organizações ofereceram algumas das práticas que utilizam para garantir implementação responsável e responsabilizável de IA. A tabela abaixo faz um levantamento dessas práticas.

Elementos de responsabilização	Práticas relacionadas
Liderança e supervisão	· Compromisso público e exemplo vindo de cima para o respeito de ética, valores e princípios específicos em desenvolvimento de IA · Processos e tomada de decisões de IA institucionalizados · Regras do Código de Ética Interno · IA/ética/comitês de supervisão, conselhos, comitês (internos e externos) · Indicar membro do comitê para supervisão de IA · Indicar líder/oficial responsável de IA · Engenheiros e campeões de privacidade/IA · Criação de um conselho interdisciplinar interno (advogados, equipes técnicas, pesquisa, unidades de negócios) · Indicar administradores de privacidade para coordenar outras pessoas · Engajar-se com reguladores em sandboxes regulatórias
Avaliação de riscos	· Compreender o propósito da IA e utilizar o caso nos negócios e processos - seja paratomar decisões, para entrar nas decisões ou outro · Compreender o impacto nos indivíduos · Compreender e articular os benefícios da aplicação proposta de IA e a reticência a risco · Ferramentas de avaliação de justiça · Avaliação de Impacto Algorítmico · Avaliação de Impacto Ético · Avaliação mais Ampla do Impacto sobre Direitos Humanos · DPIA para processamento de alto risco · Considerar técnicas para anonimato · Documentar trocas/compensações (p.ex. precisão x minimização de dados, segurança x transparência, impacto sobre poucos x benefício para a sociedade)
Políticas e Procedimentos	· Princípios de alto nível para IA - como projetar, usar, vender, etc. · Questões e procedimentos de avaliação · Medidas de responsabilização para duas etapas: treinamento e tomada de decisões · Listas brancas, pretas e cinzas de uso de IA · Avaliar os dados comparado com o propósito: qualidade, proveniência, pessoal ou não, sintético, fontes internas ou externas · Verificação da entrada e saída de dados · Viés algorítmico - ferramentas para identificar, monitorar e testar; incluir dados sensíveis em conjuntos de dados para evitar vieses · Fazer teste-piloto em modelos de IA antes do lançamento · Testar robustez das técnicas de desidentificação · Uso de dados criptografados ou dados sintéticos em alguns modelos de IA/ML e para o treinamento de modelos · Uso de conjuntos de dados de alta qualidade mas menores · Modelos de aprendizagem de IA unificados (dados não saem do dispositivo) · Considerações especiais para empresas que criam e vendem modelos de AI, software, aplicações · Listas de verificação de due diligence para parceiros de negócios usando ferramentas e tecnologia de IA

Elementos de responsabilização	Práticas relacionadas
Transparência	• Necessidades diferentes de transparência para indivíduos, autoridades reguladoras, negócios e parceiros de dados, e internamente para engenheiros e liderança • Explicabilidade faz parte de transparência e justiça • Caminho de transparência: explicabilidade de decisão e trabalho mais amplo de algoritmos + mais sobre o processo que sobre a tecnologia + quais fatores + quais testes para ser justos + responsabilização pelo impacto das decisões sobre a vida de uma pessoa + qual extensão de supervisão humana • Explicar que é uma decisão de IA/ML, se houver possibilidade de confusão (teste de Touring) • Fornecer informações contrafactuais • Transparência diferenciada e flexível, vinculada a contexto, público/usuários, propósito de explicabilidade e risco, gravidade do prejuízo; listas prescritivas de elementos de transparência não são úteis • Cartões modelos e fichas de dados • Compreender as expectativas dos clientes e implementar com base em sua prontidão para acolher IA - transparência em camadas/níveis • Da caixa preta à caixa de vidro: observando os dados assim como o algoritmo/modelo; aspirar à explicabilidade ajuda a compreender a caixa preta e desenvolve confiança
Treinamento e consciência	• Treinamento de cientistas de dados, incluindo como evitar e lidar com viés • Treinamento multifuncional - profissionais de privacidade e engenheiros • Treinamento funcional e ad hoc • Treinamento em justiça • Treinamento em ética • Usar casos em que a implementação problemática de IA tenha sido impedida • Papel dos “Tradutores” nas organizações, explicando o impacto e os desenvolvimentos de IA
Monitoramento e verificação	• O propósito de IA determina quanta intervenção humana é requerida • Humanos no ciclo - no design, na supervisão, na reparação • Compreensão humana dos negócios e processos usando IA • Desenvolvimento humano de software e processos • Auditoria humana de entrada e saída • Revisão humana de decisões individuais • Monitoramento, validações e verificações contínuas • Comitês de supervisão mesmo no estágio de design • Solicitações de reparação para um humano, não para um bot • Monitorando o ecossistema a partir da entrada de fluxo de dados, processo de dados e saída de dados • Confiança em diferentes técnicas de auditoria • Controle de versão e desvio de modelo, rastreamento de caixa preta, algoritmos por engenheiros • Modelos RACI para humanos e interação de IA
Resposta e aplicação das normas	• Lidar com reclamações • Mecanismos de reparação para que os indivíduos reparem decisões de IA • Canal de feedback • Supervisão interna de implementação de IA

Referências

[1] "First Report: Artificial Intelligence and Data Protection in Tension," CIPL (10 de outubro de 2018), disponível em https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ai_first_report_-_artificial_intelligence_and_data_protection_in_te....pdf, na pág. 19.

[2] Os eventos incluíram a uma Sessão Interativa de Trabalho Conjunto da Comissão de Proteção de Dados Pessoais (PDPC) Cingapura/CIPL sobre "IA Responsável e Responsabilizável" (16 de novembro de 2018, Cingapura); Mesa-Redonda com Participantes da Indústria e Reguladores Europeus sobre "Questões Difíceis em IA" (12 de março de 2019, Londres); Mesa-Redonda com O Grupo de Especialistas de Alto Nível da Comissão da UE sobre "Diretrizes Éticas para IA Confiável" (27 de junho de 2019, Bruxelas); Mesa-Redonda com Participantes da Indústria e Reguladores Asiáticos sobre "Desafios e Soluções em Proteção de Dados Pessoais em IA" (18 de julho de 2019, Cingapura); Mesa-Redonda com o UK Information Commissioner's Office (ICO) sobre "Estrutura de Auditoria de IA" (12 de setembro de 2019, Londres); CIPL/TTC Labs Design Jam sobre "Explicabilidade de IA" (3 de dezembro de 2019, Cebu); e Workshop de Indústria do CIPL sobre uma Abordagem Europeia Regulatória de IA (14 de janeiro de 2020, Bruxelas).

[3] Por exemplo, RGPD, Artigo 5 "Os dados pessoais são: (a) Objeto de um tratamento lícito, leal e transparente em relação ao titular dos dados ('licitude, lealdade e transparência')"; Lei de Privacidade da Nova Zelândia, Seção 6, Princípio 4(b)(i) "As informações pessoais não devem ser coletadas por uma agência (b) por meios que, nas circunstâncias do caso (i), sejam injustos;" Projeto de Lei de Proteção de Dados Pessoais da Índia, Seção 5 (a) "Todas as pessoas que processam dados pessoais devem processar esses dados pessoais (a) de maneira justa e razoável e garantir a privacidade do princípio de dados."

[4] Ver "FTC Policy Statement on Unfairness" (17 de dezembro de 1980), disponível em https://www.ftc.gov/public-statements/1980/12/ftc-policy-statement-unfairness.

[5] Comitê Europeu de Proteção de Dados, Diretrizes 4/2019 no Artigo 25: Data Protection by Design and by Default (adotado em 13 de novembro de 2019), disponível em https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201904_dataprotection_by_design_and_by_default.pdf.

[6] Tem havido outras tentativas de descrever ou definir justiça. A UK ICO, por exemplo, descreveu justiça (*fairness*) como "você só deve manipular dados pessoais de maneiras que as pessoas razoavelmente esperariam e não usá-los de maneiras que tenham efeitos adversos injustificados sobre eles". "Principle (a): Lawfulness, Fairness, and Transparency," UK ICO, disponível em https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/. O RGPD também fornece algumas orientações sobre justiça ou como obter justiça, por exemplo, no Considerando 71, que declara: "Para garantir um processamento justo e transparente em relação ao titular dos dados (...) o controlador deve usar procedimentos matemáticos ou estatísticos apropriados para criar perfis, implementar medidas técnicas e organizacionais adequadas para garantir, em particular, que os fatores que resultam em imprecisões nos dados pessoais sejam corrigidos, que o risco de erros seja minimizado, que os dados pessoais sejam protegidos de uma maneira que leve em consideração os riscos potenciais envolvidos para os interesses e direitos do titular dos dados e para impedir, inter alia, efeitos discriminatórios sobre pessoas físicas com base em origem racial ou étnica, opinião política, religião ou crenças, associação a sindicatos, status genético ou de saúde, orientação sexual ou processamento resultante em medidas que tenham esse efeito ".

[7] Para informações básicas, ver Douglas MacMillan e Nick Anderson, "Student Tracking, Secret Scores: How College Admissions Offices Rank Prospects Before They Apply," Washington Post (14 de outubro de 2019), disponível emhttps://www.washingtonpost.com/business/2019/10/14/colleges-quietly-rank-prospective-students-based-their-personal-data/?fbclid=IwAR24p1HKEaHfNoK7kH4H5XBeDw4qgRib_v-048afJ5bF5z1odIe9vCtiVac.

[8] "Em um movimento raramente visto entre as corporações de tecnologia, [a Axon] reuniu o comitê independente ano passado para avaliar as possíveis consequências e custos éticos da inteligência artificial e do software de reconhecimento facial. O primeiro relatório do comitê {...} concluiu que 'a tecnologia de reconhecimento facial não é atualmente confiável o suficiente para justificar eticamente seu uso' - orientação que a Axon planeja seguir." Deanna Paul, "A Maker of Police Body Cameras Won't Use Facial Recognition Yet, for Two Reasons: Bias and Inaccuracy" (28 de junho de 2019), disponível em https://www.washingtonpost.com/nation/2019/06/29/police-body-cam-maker-wont-use-facial-recognition-yet-two-reasons-bias-inaccuracy/.

[9] Elizabeth Denham, "Blog: Live Facial Recognition Technology—Police Forces Need to Slow Down and Justify Its Use," UK ICO, disponível emhttps://ico.org.uk/about-the-ico/news-and-events/blog-live-facial-recognition-technology-police-forces-need-to-slow-down-and-justify-its-use/.

[10] "Ethics Guidelines for Trustworthy AI," Grupo Europeu de Especialistas de Alto Nível (HLEG) sobre Inteligência Artificial (8 de abril de 2019), disponível em https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, na pág. 30.

[11] "Treating Customers Fairly—Towards Fair Outcomes for Consumers," Autoridade em Serviços Financeiros do Reio Unido (julho de 2006), disponível emhttps://www.fca.org.uk/publication/archive/fsa-tcf-towards.pdf, nas págs. 11-13.

[12] Comissão de Proteção de Dados Pessoais de Cingapura, "A Proposed Model Artificial Intelligence Governance Framework," (janeiro de 2019), disponível emhttps://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf, na pág. 15. "Uma decisão é justa para um indivíduo se for a mesma no mundo atual e em um mundo contrafactual em que o indivíduo pertencesse a um grupo demográfico diferente."

[13] *Ver* nota 1, na pág. 18.

[14] Certos conceitos de proteção de dados estão em tensão com as técnicas usadas para identificar o risco de viés. Por exemplo, embora o Artigo 9 do RGPD proíba o processamento de determinadas categorias especiais de dados pessoais a menos que se apliquem certas exceções, é provável que a coleta de dados de origem racial ou étnica seja necessária para identificar se um algoritmo está discriminando por etnia ou raça. Embora os princípios de proteção de dados esperem proteger os indivíduos limitando a coleta dessas informações, esses dados são fundamentais para atingir o objetivo de mitigar e detectar vieses.

[15] Google, Perspectives on Issues in AI Governance (2019), disponível em https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf, na pág. 15.

[16] Rumman Chowdhury, "Tackling the Challenge of Ethics in AI," Accenture (6 de junho de 2018), disponível em https://www.accenture.com/gb-en/blogs/blogs-cogx-tackling-challenge-ethics-ai.

[17] Kush R. Varshney, "Introducing AI Fairness 360," Blog de Pesquisa da IBM (19 de setembro de 2018), disponível emhttps://www.ibm.com/blogs/research/2018/09/ai-fairness-360/.

[18] "Manage AI, with Trust and Confidence in Business Outcomes," IBM, disponível em https://www.ibm.com/downloads/cas/RYXBG8OZ.

[19] RGPD, Artigo 35(1).

[20] "International Resolution on Privacy as a Fundamental Human Right and Precondition for Exercising Other Fundamental Rights," 41st International Conference of Data Protection & Privacy Commissioners (ICDPPC) (21-24 de outubro de 2019), disponível em https://privacyconference2019.info/wp-content/uploads/2019/10/Resolution-on-privacy-as-a-fundamental-human-right-2019-FINAL-EN.pdf.

[21] *Ver* nota 15, na pág. 13.

[22] *Ver* note 1, na pág. 15; Ver também RGPD, Artigo 12.

[23] Ver, por exemplo, RGPD, Artigos 5 e 12; LGPD do Brasil, Artigos 14(2) e (6), Artigo 18 (2), (7) e (8); e Projeto de Lei de Proteção de Dados Pessoais na Índia, Seção 23.

[24] "Project Expl*AI*n: Interim Report," UK Information Commissioner's Office (3 de junho de 2019), disponível em https://ico.org.uk/media/2615039/project-explain-20190603.pdf, na pág. 15.

[25] RGPD, Artigo 22.

[26] *Ver* nota 10, na pág. 20.

[27] *Id.*

[28] *Ver* nota 12, na pág. 13. As Diretrizes para IA Confiável do HLEG classificam as considerações de transparência de maneira semelhante à Estrutura Modelo de Cingapura, com foco na rastreabilidade, explicabilidade e comunicação.

[29] "Como rótulos nutricionais para alimentos ou fichas de informações para aparelhos, as fichas técnicas para serviços de IA forneceriaminformações sobre as características importantes do produto." Aleksandra Mojsilovic, "Factsheets for AI Services," Blog de Pesquisa da IBM (22 de agosto de 2018), disponível em https://www.ibm.com/blogs/research/2018/08/factsheets-ai/.

[30] Uma prática recomendada poderia ser o uso de Cartões de Modelo, que são "documentos curtos que acompanham modelos treinados de aprendizado de máquina, os quais fornecem avaliação comparada em uma variedade de condições (...) relevantes para os domínios das aplicações pretendidas." Margaret Mitchell et al., "Model Cards for Model Reporting" (14 de janeiro de 2019), disponível em https://arxiv.org/pdf/1810.03993.pdf. Ver também "The Value of a Shared Understanding of AI Models," Google, disponível em https://modelcards.withgoogle.com/about.
[31] RGPD, Artigo 5(1)(b) [grifo nosso].
[32] "Os propósitos para os quais os dados pessoais são coletados devem ser especificados no máximo no momento da coleta de dados, e o uso subsequente deve ser limitado ao cumprimento desses propósitos ou de outros que não sejam incompatíveis com esses propósitos e que sejam especificados em cada ocasião da mudança de propósito." Diretrizes Revisadas da OECD sobre a Proteção de Privacidade e os Fluxos Transfronteiriços dos Dados Pessoais (2013), disponível em http://oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.
[33] Ver, por exemplo, RGPD, Artigo 6(4). Observe, no entanto, que o RGPD proíbe o uso de dados pessoais para uma finalidade diferente daquela para a qual foram originalmente coletados, a menos que o novo objetivo seja "não incompatível" com o original.
[34] RGPD, Artigo 6(4). Os critérios para processamento compatível adicional sob o RGPD incluem (a) qualquer ligação entre os propósitos para os quais os dados pessoais foram coletados e os propósitos do processamento adicional pretendido; (b) o contexto em que os dados pessoais foram coletados, em particular no que se refere à relação entre os titulares de dados e o responsável pelo tratamento; (c) a natureza dos dados pessoais, em particular se são processadas categorias especiais de dados pessoais, nos termos do Artigo 9, ou se são tratados dados pessoais relacionados a condenações e infrações penais, nos termos do Artigo 10; (d) as possíveis consequências do processamento adicional pretendido para os titulares dos dados; e (e) a existência de salvaguardas apropriadas, que podem incluir criptografia ou pseudonimização.
[35] RGPD, Artigo 6(4).
[36] Na UE, isto poderia ser potencialmente considerado como "propósitos de pesquisa." RGPD, Artigo 5 (1) (b) (permitindo processamento adicional para "fins de arquivamento de interesse público, fins de pesquisa científica ou histórica ou fins estatísticos").
[37] Fred H. Cate and Rachel D. Dockery, "Artificial Intelligence and Data Protection: Observations on a Growing Conflict," *Seoul National University Journal of Law & Economic Regulation*, Vol. 11. No. 2 (2018), na pág. 123.
[38] RGPD, Artigo 5(1)(c).
[39] Os medicamentos projetados por IA também podem aumentar a velocidade dos ensaios clínicos, com um composto projetado por AI atingindo o estágio de teste clínico dentro de 12 meses comparado com um período de quatro anos e meio segundo as abordagens tradicionais para o desenvolvimento de medicamentos. Ver Madhumita Murgia "AI-designed drug to enter human clinical trial for first time," Financial Times, (29 de janeiro de 2020), disponível em https://www.ft.com/content/fe55190e-42bf-11ea-a43a-c4b328d9061c.
[40] Alternativas potenciais ao uso de dados pessoais incluem o uso de conjuntos de dados sintéticos, privacidade diferencial ou outras técnicas de aprimoramento da privacidade. Para considerações relacionadas com a minimização de dados e IA, ver o blog do UK Information Commissioner's Office blog sobre Estrutura de Auditoria de IA, Reuben Binns e Valeria Gallo, "Data Minimisation and Privacy-Preserving Techniques in AI Systems," UK ICO (21 de agosto de 2019), disponível em https://ai-auditingframework.blogspot.com/2019/08/data-minimisation-and-privacy_21.html.
[41] Ver, por exemplo, a nota 1 ("A IA, e a variedade de conjuntos de dados dos quais muitas vezes depende, apenas exacerba o desafio de determinar quando se aplicam as leis de proteção de dados expandindo a capacidade de vincular dados ou reconhecer padrões de dados que podem tornar os dados não pessoais em algo identificável (...). Dito de forma simples, quanto mais dados disponíveis, mais difícil é desidentificá-los efetivamente.").
[42] Ver, por exemplo, Ursula von der Leyen, "A Union that Strives for More: My Agenda for Europe, Political Guidelines for the Next European Commission," disponível em https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf, na pág. 13 ("Nos meus 100 primeiros dias de mandato, apresentarei legislação para uma abordagem europeia coordenada sobre as implicações éticas e humanas da Inteligência Artificial."). No entanto, essa iniciativa pode, em última análise, assumir a forma de um documento de orientação que definirá diferentes opções para uma estrutura legal sobre IA que pode levar a propostas formais no final do ano.

[43] "A tecnologia de IA precisa continuar a se desenvolver e amadurecer antes de que possam ser estruturadas regras para governá-la. A seguir, é preciso se chegar a um consenso sobre princípios e valores da sociedade para governar a implementação e uso de IA, seguidos de melhores práticas para estar à mesma altura. Então, provavelmente estaremos em uma melhor posição para que os governos criem regras legais e regulatórias para que todos sigam." The Future Computed, Microsoft (2018), disponível em https://blogs.microsoft.com/wp-content/uploads/2018/02/The-Future-Computed_2.8.18.pdf, na pág. 9.

[44] "Artificial Intelligence Impact Assessment," Platform for the Information Society (2018), disponível em https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf, at page 21.

[45] *Ver* nota 8.

[46] *Ver* nota 10.

[47] *Ver* nota 12.

[48] AI Auditing Framework, UK ICO, disponível em https://ico.org.uk/about-the-ico/news-and-events/ai-auditing-framework/.

[49] *Ver* nota 12, na pág. 6-7.

[50] Isso é aparente na exigência da DPIA no Artigo 35, entre outros. Para uma visão geral das disposições do RGPD, ver Gabriel Maldoff, "The Risk-Based Approach in the GDPR: Interpretation and Implications," Associação Internacional de Profissionais de Privacidade, disponível em https://iapp.org/media/pdf/resource_center/GDPR_Study_Maldoff.pdf.

[51] "[A] abordagem baseada em risco deve ser usada para determinar quais riscos são aceitáveis e quais apresentam a possibilidade de prejuízo inaceitável, ou que tenha custos esperados maiores que os benefícios esperados. As agências devem ser transparentes sobre suas avaliações de risco e reavaliar suas suposições e conclusões em intervalos adequados, a fim de promover a responsabilização." Russel Vought, "Draft Memorandum for the Heads of Executive Departments and Agencies: Guidance for the Regulation of Artificial Intelligence Applications," US Office of Management and Budget (7 de janeiro de 2019), disponível em https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf.

[52] *Ver* nota 44, na pág. 8.

[53] Para um exemplo de um esforço para avaliar esses prejuízos, pode ser útil consultar o White Harms Online Paper, do Reino Unido, que faz parte de uma consulta em andamento para analisar prejuízos online e implementar medidas de segurança. "Online Harms White Paper," Departamento Digital, de Cultura, Mídia e Esportes (abril de 2019), disponível em https://www.gov.uk/government/consultations/online-harms-white-paper.

[54] "Paper 1: A Risk-based Approach to Privacy: Improving Effectiveness in Practice," CIPL, (19 de junho de 2014), disponível em https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_1-a_risk_based_approach_to_privacy_improving_effectiveness_in_practice.pdf; "Paper 2: The Role of Risk Management in Data Protection," CIPL (23 de novembro de 2014), disponível em https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_2-the_role_of_risk_management_in_data_protection-c.pdf; "Protecting Privacy In a World of Big Data: Paper 2: The Role of Risk Management," CIPL (16 fevereiro de 2016), disponível em https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/protecting_privacy_in_a_world_of_big_data_paper_2_the_role_of_risk_management_16_february_2016.pdf; "Risk, High Risk, Risk Assessments and Data Protection Impact Assessments under the GDPR," CIPL (21 de dezembro de 2016), disponível em https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_gdpr_project_risk_white_paper_21_december_2016.pdf.

[55] Roger Burkhardt, Nicolas Hohn e Chris Wigley, "Leading Your Organization to Responsible AI," McKinsey Analytics (maio de 2019), disponível em https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Leading%20your%20organization%20to%20responsible%20AI/Leading-your-organization-to-responsible-AI.ashx, na pág. 3-4.

[56] Alguns descreveram o nível de envolvimento humano com os termos *human-in-the-loop*, *human-out-of-the-loop* e *human-over-the-loop*. *Human-in-the-loop* refere-se a situações em que "a supervisão humana está ativa e envolvida, com o ser humano mantendo o controle total e a IA apenas fornecendo recomendações ou sugestões;" *human-out-of-the-loop* refere-se a situações em que "não há supervisão humana sobre a execução de decisões" e "a IA tem controle total sem a opção de substituição humana"; e *human-over-the-loop* "permite que humanos ajustem parâmetros durante a execução do algoritmo". Ver nota 12, na pág. 8.

[57] Este tipo de revisão humana já existe na lei dos EUA em certos contextos em que o processamento automatizado informa para tomadas de decisões. Tanto a Lei de Relatórios de Crédito Justo quanto a Lei de Igualdade de Oportunidades de Crédito fornecem aos consumidores algum direito a explicação e contestação de decisões, enquanto o Título VII da Lei dos Direitos Civis, bem como a Lei da Habitação Justa, fornecem aos indivíduos o direito de contestar decisões.

[58] "The Case for Accountability: How It Enables Effective Data Protection and Trust in the Digital Society," CIPL (23 de julho de 2018), disponível em https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_paper_1_-_the_case_for_accountability_-_how_it_enables_effective_data_protection_and_trust_in_the_digital_society.pdf.

[59] RGPD, Artigo 22; ver também RGPD, Artigo 35(1) ("Quando um tipo de processamento em particular, utilizando novas tecnologias e levando em conta a natureza, o escopo, o contexto e os propósitos do processamento, puder resultar em alto risco para os direitos e liberdades das pessoas físicas, o responsável pelo tratamento deverá, antes do tratamento, proceder a uma avaliação do impacto das operações de tratamento pretendidas sobre a proteção de dados pessoais.").

[60] Um guia potencial para DPIAs de IA é "1. Mapear os benefícios (públicos) de uma aplicação de IA. 2. Analisar a confiabilidade, segurança e transparência de aplicações de IA. 3. Identificar valores e interesses que estão envolvidos na implementação de IA. 4. Identificar e limitar riscos da implementação de IA. 5. Responsabilizar-se pelas escolhas que foram feitas na avaliação de valores e interesses." Ver nota 44, na pág. 25.

[61] As orientações da ICO também sugeriram cinco elementos para as DPIAs de IA, incluindo 1) uma descrição sistêmica do processamento, 2) avaliação de necessidade e proporcionalidade, 3) identificação de riscos a direitos e liberdades, 4) medidas para abordar os riscos e 5) um documento "vivo". Simon Reader, "Data Protection Impact Assessments and AI," UK ICO AI Auditing Framework (23 de outubro de 2019), disponível em https://ai-auditingframework.blogspot.com/2019/10/data-protection-impact-assessments-and.html.

[62] *Ver* nota 12, na pág. 8 (embora estes estejam sendo chamados de "avaliações de risco" nesta estrutura).

[63] A Accenture recentemente publicou um artigo técnico detalhando os benefícios de comitês de ética em IA e dados, explorando diferentes cenários e melhores práticas para eles. O artigo explica que, "para ser bem-sucedido, deve ser projetado de forma bem pensada, receber os recursos adequados, ter um responsável claro, ser suficientemente empoderado e adequadamente localizado dentro da organização." John Basl and Ronald Sandler, "Building Data and AI Ethics Committees," Accenture & Northeastern University Ethics Institute (2019), disponível em https://www.accenture.com/_acnmedia/pdf-107/accenture-ai-data-ethics-committee-report-executive-summary.pdf, na pág. 2.

[64] *Ver* nota 10, na pág. 20 (Quando impactos adversos injustos ocorrem, devem ser previstos mecanismos acessíveis que assegurem reparação adequada. Saber que a reparação é possível quando as coisas dão errado é fundamental para garantir confiança); ver também nota 12, na pág. 17 (discutindo a necessidade de canais de feedback para que os indivíduos possam revisar e corrigir seus dados, assim como canais de revisão de decisões para contestar decisões adversas).

Sobre o Centre for Information Policy Leadership

O CIPL é um laboratório global de ideias sobre privacidade de dados e cibersegurança no escritório de advocacia Hunton Andrews
Kurth LLP e é financiado pelo escritório e por 90 empresas parceiras que são
líderes em setores centrais da economia global. A missão do CIPL é envolver-se em liderança de ideias e desenvolver melhores práticas que garantam tanto proteções de privacidade efetivas quanto o uso responsável de informações pessoais nesta era moderna da informação. O trabalho do CIPL facilita o envolvimento
construtivo entre líderes de negócios, profissionais de privacidade e segurança, autoridades reguladoras e
legisladores de todo o mundo. Para mais informações, visite o site do CIPL em **http://www.informationpolicycentre.com/.**

Projeto CIPL AI

- Para saber mais sobre o Projeto sobre Inteligência Artificial e Proteção de Dados do CIPL: Entregando Responsabilidade de IA Sustentável na Prática, consulte **https://www.informationpolicycentre.com/ai-project.html**
- Para ler o primeiro relatório de IA do CIPL sobre Inteligência Artificial e Proteção de Dados em Tensão, consulte **https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ai_first_report_-_artificial_intelligence_and_data_protection_in_te....pdf**
- Se estiver interessado(a) em se unir ao CIPL e participar neste projeto de IA, entre em contato com Bojana Bellamy no email **bbellamy@HuntonAK.com** ou Michelle Marcoot no email **mmarcoot@HuntonAK.com**.

Centre for Information Policy Leadership
HUNTON ANDREWS KURTH

Escritório em Washington, DC
2200 Pennsylvania Avenue
Washington, DC 20037
+1 202 955 1563

Escritório em Londres
30 St Mary Axe
London EC3A 8EP
+44 20 7220 5700

Escritório em Bruxelas
Rue des Colonies 11
1000 Brussels
+32 2 643 58 00