

# A regulamentação da IA está chegando

por François Cadelon, Rodolphe Charme di Carlo, Midas De Bondt, e Theodoros Evgeniou

Data da publicação (Setembro–Outubro 2021)



Li Sun

**Resumo** Durante anos, a preocupação pública com o risco tecnológico se concentrou no uso indevido de dados pessoais. Mas, à medida que as empresas incorporam cada vez mais inteligência artificial em produtos e processos, a... [mais](#)

Aviso: este texto foi traduzido com o uso de tradução automática e pode conter erros. Responda a esta [pesquisa](#) para nos enviar seus comentários e obtenha mais informações em nossas [perguntas frequentes](#).

[Read in English](#)

Durante a maior parte da última década, as preocupações do público com a tecnologia digital se concentraram no potencial abuso de dados pessoais. As pessoas se sentiam desconfortáveis com a forma como as empresas podiam rastrear seus movimentos on-line, muitas vezes coletando números de cartão de crédito, endereços e outras informações críticas. Eles acharam assustador serem seguidos pela web por anúncios claramente acionados por suas pesquisas ociosas e se preocuparam com roubo de identidade e fraude.

Essas preocupações levaram à aprovação de medidas nos Estados Unidos e na Europa garantindo aos usuários da Internet algum nível de controle sobre seus dados e imagens pessoais, principalmente, o Regulamento Geral de Proteção de Dados (GDPR) de 2018 da União Europeia. Obviamente, essas medidas não encerraram o debate sobre o uso de dados pessoais pelas empresas. Alguns argumentam que restringi-lo prejudicará o desempenho econômico da Europa e dos Estados Unidos em relação a países menos restritivos, principalmente a China, cujos gigantes digitais prosperaram com a ajuda do acesso imediato e levemente regulamentado a informações pessoais de todos os tipos. (Recentemente, no entanto, o governo chinês começou a limitar o liberdade das empresas digitais—conforme demonstrado pelas grandes multas impostas ao Alibaba.) Outros apontam que há muitas evidências de que uma regulamentação mais rígida colocou empresas europeias menores em considerável desvantagem em relação a rivais norte-americanos de maior porte, como Google e Amazon.

Mas o debate está entrando em uma nova fase. À medida que as empresas incorporam cada vez mais a inteligência artificial em seus produtos, serviços, processos e tomadas de decisão, a atenção está se voltando para a forma como os dados são usados pelo software, principalmente por algoritmos complexos e em evolução que podem diagnosticar um câncer, dirigir um carro ou aprovar um empréstimo. A UE, que está novamente na liderança (em seu livro branco de 2020 “Sobre inteligência artificial — uma abordagem europeia de excelência e confiança” e sua proposta de 2021 para um quadro jurídico de IA), considera que a regulamentação é essencial para o desenvolvimento de ferramentas de IA nas quais os consumidores possam confiar.

O que tudo isso significará para as empresas? Estamos pesquisando como regulamentar Algoritmos de IA e como implementar Sistemas de IA que se baseiam nos princípios-chave subjacentes às estruturas regulatórias propostas, e temos ajudado empresas de todos os setores a lançar e ampliar iniciativas orientadas por IA. Nas páginas seguintes, nos baseamos nesse trabalho e no de outros pesquisadores para explorar os três principais desafios que os líderes de negócios enfrentam ao integrar a IA em suas tomadas de decisão e processos,



tentando garantir que ela seja segura e confiável para os clientes. Também apresentamos uma estrutura para orientar os executivos nessas tarefas, com base em parte nos conceitos aplicados ao gerenciamento de riscos estratégicos.

## **Resultados injustos: os riscos do uso da IA**

Os sistemas de IA que produzem resultados tendenciosos estão nas manchetes. Um exemplo bem conhecido é o da Apple algoritmo de cartão de crédito, que foi acusada de discriminar mulheres, desencadeando uma investigação pelo Departamento de Serviços Financeiros de Nova York.

Mas o problema surge de muitas outras formas: por exemplo, em algoritmos de publicidade on-line onipresentes, que podem segmentar espectadores por raça, religião ou gênero, e no rastreador automático de currículos da Amazon, que filtrava candidatas do sexo feminino. UMA estudo recente publicado em *Ciência* mostrou que as ferramentas de previsão de risco usadas na área da saúde, que afetam milhões de pessoas nos Estados Unidos todos os anos, exibem um viés racial significativo. Outro estudo, publicado no *Jornal de Medicina Interna Geral*, descobriram que o software usado pelos principais hospitais para priorizar os receptores de transplantes renais discriminava pacientes negros.

## **A IA aumenta a escala potencial de preconceito: qualquer falha pode afetar milhões de pessoas, expondo as empresas a ações judiciais coletivas.**

Na maioria dos casos, o problema decorre dos dados usados para treinar a IA. Se esses dados forem tendenciosos, a IA adquirirá e poderá até mesmo amplificar o viés. Quando a Microsoft usou tweets para treinar um chatbot a interagir com usuários do Twitter, por exemplo, ela teve que derrubar o bot no dia seguinte à sua publicação por causa de suas mensagens racistas e inflamatórias. Mas não basta

simplesmente eliminar informações demográficas, como raça ou sexo, dos dados de treinamento, porque, em algumas situações, esses dados são necessários para corrigir preconceitos.

Em teoria, pode ser possível codificar algum conceito de justiça no software, exigindo que todos os resultados atendam a determinadas condições. A Amazon está experimentando um métrica de imparcialidade chamado disparidade demográfica condicional, e outras empresas estão desenvolvendo métricas semelhantes. Mas um obstáculo é que não há uma definição consensual de justiça, nem é possível ser categórico sobre as condições gerais que determinam resultados equitativos. Além disso, as partes interessadas em qualquer situação podem ter noções muito diferentes do que constitui justiça. Como resultado, qualquer tentativa de projetá-lo no software será complicada.

Ao lidar com resultados tendenciosos, os reguladores recorreram principalmente à legislação antidiscriminação padrão. Isso é viável desde que haja pessoas que possam ser responsabilizadas por decisões problemáticas. Mas com a IA cada vez mais presente, a responsabilidade individual é prejudicada. Pior ainda, a IA aumenta a escala potencial de preconceito: qualquer falha pode afetar milhões de pessoas, expondo as empresas a ações judiciais coletivas de proporções históricas e colocando sua reputação em risco.

O que os executivos podem fazer para evitar esses problemas?

Como primeiro passo, antes de tomar qualquer decisão, eles devem aprofundar sua compreensão dos riscos, explorando quatro fatores:

**O impacto dos resultados.** Alguns algoritmos tomam ou afetam decisões com consequências diretas e importantes na vida das pessoas. Eles diagnosticam condições médicas, por exemplo, selecionam candidatos a empregos, aprovam empréstimos imobiliários ou recomendam sentenças de prisão. Em tais circunstâncias, pode ser sensato evitar o uso da IA ou, pelo menos, subordiná-la ao julgamento humano.

No entanto, a última abordagem ainda requer uma reflexão cuidadosa. Suponha que um juiz tenha concedido liberdade antecipada a um infrator contra uma recomendação de IA e essa



pessoa tenha cometido um crime violento. A juíza estaria sob pressão para explicar por que ela ignorou a IA. O uso da IA pode, portanto, aumentar a responsabilidade dos tomadores de decisão humanos, o que pode fazer com que as pessoas passem a usar os algoritmos com mais frequência do que deveriam.

Isso não quer dizer que a IA não tenha seus usos em contextos de alto impacto. As organizações que dependem de tomadores de decisão humanos ainda precisarão controlar o preconceito inconsciente entre essas pessoas, o que a IA pode ajudar a revelar. Em última análise, a Amazon decidiu não aproveitar a IA como ferramenta de recrutamento, mas sim usá-la para detectar falhas em sua recrutamento atual abordagem. A conclusão é que a imparcialidade dos algoritmos em relação à tomada de decisão humana precisa ser considerada ao escolher se deseja usar a IA.

**A natureza e o escopo das decisões.** Pesquisas sugerem que o grau de confiança na IA varia de acordo com o tipo de decisão para a qual ela é usada. Quando uma tarefa é percebida como relativamente mecânica e limitada — pense em otimizar um cronograma ou analisar imagens — o software é considerado pelo menos tão confiável quanto os humanos.

Mas quando as decisões são consideradas subjetivas ou as variáveis mudam (como nas sentenças legais, em que as circunstâncias atenuantes dos infratores podem ser diferentes), o julgamento humano é mais confiável, em parte devido à capacidade de empatia das pessoas. Isso sugere que as empresas precisam se comunicar com muito cuidado sobre a natureza e o escopo específicos das decisões às quais estão aplicando a IA e por que ela é preferível ao julgamento humano nessas situações. Este é um exercício bastante simples em muitos contextos, mesmo aqueles com consequências graves. Por exemplo, em diagnósticos automáticos de exames médicos, as pessoas podem facilmente aceitar a vantagem que o software treinado em bilhões de pontos de dados bem definidos tem sobre os humanos, que podem processar apenas alguns milhares.



Por outro lado, aplicar a IA para fazer um diagnóstico sobre saúde mental, em que os fatores podem ser comportamentais, difíceis de definir e específicos para cada caso, provavelmente seria inapropriado. É difícil para as pessoas aceitarem que as máquinas possam processar situações altamente contextuais. E mesmo quando as variáveis críticas foram identificadas com precisão, a forma como elas diferem entre as populações geralmente não é totalmente compreendida, o que nos leva ao próximo fator.

**Complexidade operacional e limites de escala.** Um algoritmo pode não ser justo em todas as regiões e mercados. Por exemplo, selecionar consumidores para obter descontos pode parecer equitativo em toda a população dos EUA, mas ainda assim mostrar preconceito quando aplicado a, digamos, residentes de Manhattan, se o comportamento e as atitudes dos consumidores em Manhattan não corresponderem às médias nacionais e não se refletirem no treinamento do algoritmo. As estatísticas médias podem mascarar a discriminação entre regiões ou subpopulações, e evitá-la pode exigir a personalização de algoritmos para cada subconjunto. Isso explica por que qualquer regulamentação destinada a diminuir os preconceitos locais ou de pequenos grupos provavelmente reduzirá o potencial de vantagens de escala da IA, que geralmente é a motivação para usá-la em primeiro lugar.

O ajuste às variações entre os mercados adiciona camadas aos algoritmos, aumentando os custos de desenvolvimento. A personalização de produtos e serviços para mercados específicos também aumenta significativamente os custos de produção e



monitoramento. Todas essas variáveis aumentam a complexidade e a sobrecarga organizacional. Se os custos ficarem muito altos, as empresas podem até mesmo abandonar alguns mercados. Por causa do GDPR, por exemplo, alguns desenvolvedores, como a Gravity Interactive (fabricante dos jogos Ragnarok e Dragon Saga), optaram por parar de vender seus produtos na UE por algum tempo. Embora a maioria já tenha encontrado uma maneira de cumprir o regulamento (Dragon Saga foi relançado em maio passado na Europa), os custos incorridos e as oportunidades perdidas são importantes.

**Capacidades de conformidade e governança.** Para seguir as regulamentações de IA mais rigorosas que estão no horizonte (pelo menos na Europa e nos Estados Unidos), as empresas precisarão de novas processos e ferramentas: auditorias de sistema, documentação e protocolos de dados (para rastreabilidade), monitoramento de IA e treinamento de conscientização sobre diversidade. Várias empresas já testam cada novo algoritmo de IA em uma variedade de partes interessadas para avaliar se sua produção está alinhada com os valores da empresa e é improvável que levante questões regulatórias. Google, Microsoft, BMW e Deutsche Telekom estão desenvolvendo políticas formais de IA com compromissos com segurança, justiça, diversidade e privacidade. Algumas empresas, como a Federal Home Loan Mortgage Corporation ( Freddie Mac), até nomearam diretores de ética para supervisionar a introdução e a aplicação de tais políticas, em muitos casos apoiando-os com conselhos de governança ética.

### **Transparência: explicando o que deu errado**

Assim como o julgamento humano, a IA não é infalível. Os algoritmos inevitavelmente tomarão algumas decisões injustas ou até inseguras.

Quando as pessoas cometem um erro, geralmente há uma investigação e uma atribuição de responsabilidade, o que pode impor penalidades legais ao tomador de decisão. Isso ajuda a organização ou a comunidade a entender e corrigir decisões injustas e a criar confiança com suas partes interessadas. Então, devemos exigir — e podemos até mesmo esperar — que a IA explique suas decisões também?

Os reguladores certamente estão se movendo nessa direção. O GDPR já descreve “o direito... de obter uma explicação da decisão tomada” por algoritmos, e a UE identificou a explicabilidade como um fator-chave para aumentar a confiança na IA em seu livro branco e proposta de regulamentação de IA.

Mas o que significa obter uma explicação para decisões automatizadas, para as quais nosso conhecimento de causa e efeito geralmente é incompleto? Foi Aristóteles quem apontou que, quando essa é a situação, a capacidade de explicar como os resultados são obtidos pode ser menos importante do que a capacidade de reproduzir os resultados e verificar empiricamente sua precisão — algo que as empresas podem fazer comparando as previsões da IA com os resultados.

Os líderes de negócios que estão considerando aplicativos de IA também precisam refletir sobre dois fatores:

**O nível de explicação necessário.** Com algoritmos de IA, as explicações podem ser amplamente classificadas em dois grupos, adequados a diferentes circunstâncias.

Explicações globais são explicações completas para todos os resultados de um determinado processo e descrevem as regras ou fórmulas que especificam relações entre variáveis de entrada. Eles geralmente são necessários quando a justiça processual é importante, por exemplo, nas decisões sobre a alocação de recursos, porque as partes interessadas precisam saber com antecedência como elas serão tomadas.

**Devemos exigir — e podemos até mesmo esperar — que a IA explique suas decisões? Os reguladores certamente estão se movendo nessa direção.**

Fornecer uma explicação global para um algoritmo pode parecer simples: tudo o que você precisa fazer é compartilhar sua fórmula. No entanto, a maioria das pessoas não tem as habilidades avançadas em



matemática ou ciência da computação necessárias para entender essa fórmula, muito menos determinar se as relações especificadas nela são apropriadas. E no caso do aprendizado de máquina, em que o software de IA cria algoritmos para descrever relações aparentes entre variáveis nos dados de treinamento, falhas ou vieses nesses dados, não no algoritmo, podem ser a causa final de qualquer problema.

Além disso, as empresas podem nem mesmo ter uma visão direta do funcionamento de seus algoritmos, e responder às restrições regulatórias para obter explicações pode exigir que elas olhem além de seus departamentos de dados e TI e talvez até mesmo para especialistas externos. Considere que as ofertas de grandes provedores de software como serviço, como Oracle, SAP e Salesforce, geralmente combinam vários componentes de IA de fornecedores terceirizados. E seus clientes às vezes escolhem e combinam soluções baseadas em IA. Mas todos os componentes de um produto final e como eles se combinam e se interconectam precisarão ser explicáveis.

Explicações locais ofereça a justificativa por trás de um resultado específico — digamos, por que um candidato (ou classe de candidatos) teve seu empréstimo negado enquanto outro recebeu um. Eles geralmente são fornecidos pelos chamados algoritmos de IA explicáveis, que têm a capacidade de informar ao destinatário de uma saída os motivos para ela. Eles podem ser usados quando os indivíduos precisam saber apenas por que uma determinada decisão foi tomada sobre eles e não têm, ou não podem, ter acesso às decisões sobre os outros.



As explicações locais podem assumir a forma de declarações que respondem à pergunta: Quais são as principais características do cliente que, se fossem diferentes, teriam mudado o resultado ou a decisão da IA? Por exemplo, se a única diferença entre dois candidatos é que um tem 24 anos e o outro 25, a explicação seria que o primeiro candidato teria recebido um empréstimo se tivesse mais de 24 anos. O problema aqui é que as próprias características identificadas podem ocultar preconceitos. Por exemplo, pode acontecer que o CEP do candidato seja o que faz a diferença, com candidatos sólidos de bairros negros sendo penalizados.

**As compensações envolvidas.** Os algoritmos mais poderosos são inerentemente opacos. Veja o Ant Group do Alibaba na China, cuja unidade MyBank usa IA para aprovar empréstimos para pequenas empresas em menos de três minutos sem intervenção humana. Para fazer isso, ele combina dados de todo o ecossistema do Alibaba, incluindo informações sobre vendas de suas plataformas de comércio eletrônico, com aprendizado de máquina para prever riscos de inadimplência e manter classificações de crédito em tempo real. Porque Software da Ant usa mais de 3.000 entradas de dados, articular claramente como ele chega a avaliações específicas (sem falar em fornecer uma explicação global) é praticamente impossível. Muitos dos aplicativos de IA mais interessantes exigem entradas algorítmicas em uma escala similar. Condições de pagamento personalizadas em mercados B2B, subscrição de seguros e carros autônomos são apenas algumas das áreas em que requisitos rigorosos de explicabilidade da IA podem prejudicar a capacidade das empresas de inovar ou crescer.

As empresas enfrentarão desafios para introduzir um serviço como o da Ant em mercados em que consumidores e reguladores valorizam muito os direitos individuais, principalmente a União Europeia e os Estados Unidos. Para implantar essa IA, as empresas precisarão ser capazes de explicar como um algoritmo define semelhanças entre clientes, por que certas diferenças entre dois clientes potenciais podem justificar tratamentos diferentes e por que clientes semelhantes podem obter explicações diferentes sobre a IA.



As expectativas de explicações também variam de acordo com a geografia, o que representa desafios para os operadores globais. Eles poderiam simplesmente adotar os requisitos de explicabilidade mais rigorosos em todo o mundo, mas isso poderia claramente colocá-los em desvantagem para os players locais em alguns mercados. Os bancos que seguissem as regras da UE teriam dificuldade em produzir algoritmos tão precisos quanto os da Ant na previsão da probabilidade de inadimplência do mutuário e, conseqüentemente, poderiam ter que ser mais rigorosos quanto aos requisitos de crédito. Por outro lado, a aplicação de vários padrões de explicabilidade provavelmente será mais complexa e cara, porque uma empresa estaria, em essência, criando algoritmos diferentes para mercados diferentes e provavelmente precisaria adicionar mais IA para garantir a interoperabilidade.

Existem, no entanto, algumas oportunidades. Os requisitos de explicabilidade podem oferecer uma fonte de diferenciação: as empresas que podem desenvolver algoritmos de IA com capacidades explicativas mais fortes estarão em uma posição melhor para conquistar a confiança dos consumidores e dos reguladores. Isso pode ter conseqüências estratégicas. Se o Citibank, por exemplo, pudesse produzir uma IA explicável para crédito de pequenas empresas tão poderosa quanto a da Ant, certamente dominaria os mercados da UE e dos EUA e poderia até mesmo ganhar uma posição no próprio território da Ant. A capacidade de comunicar a justiça e a transparência das decisões das ofertas também é um potencial diferencial para as empresas de tecnologia. A IBM desenvolveu um produto que ajuda as empresas a fazer isso: Watson OpenScale, uma plataforma de análise de dados baseada em IA para empresas.

O ponto principal é que, embora exigir que a IA forneça explicações para suas decisões possa parecer uma boa maneira de melhorar sua equidade e aumentar a confiança das partes interessadas, isso tem um preço alto, que nem sempre vale a pena pagar. Nesse caso, a única opção é voltar a encontrar um equilíbrio entre os riscos de obter alguns resultados injustos e os retornos de uma produção geral mais precisa, ou abandonar o uso da IA.

## Aprendendo e evoluindo: um terreno em constante mudança

Uma das características distintivas da IA é sua capacidade de aprender; quanto mais imagens rotuladas de vacas e zebras um algoritmo de reconhecimento de imagem é alimentado, maior a probabilidade de reconhecer uma vaca ou uma zebra. Mas existem desvantagens em aprendizagem contínua: Embora a precisão possa melhorar com o tempo, as mesmas entradas que geraram um resultado ontem podem registrar outro amanhã porque o algoritmo foi alterado pelos dados recebidos nesse ínterim.

Ao descobrir como gerenciar algoritmos que evoluem — e se devem permitir o aprendizado contínuo em primeiro lugar — os líderes de negócios devem se concentrar em três fatores:

**Riscos e recompensas.** As atitudes dos clientes em relação à evolução da IA provavelmente serão determinadas por um cálculo pessoal de risco-retorno. Na precificação de seguros, por exemplo, os algoritmos de aprendizado provavelmente fornecerão resultados mais adequados às necessidades do cliente do que qualquer coisa que os humanos possam oferecer, portanto, os clientes provavelmente terão uma tolerância relativamente alta a esse tipo de IA. Em outros contextos, o aprendizado pode não ser uma preocupação. A IA que gera recomendações de filmes ou livros, por exemplo, poderia evoluir com bastante segurança à medida que mais dados sobre as compras e as opções de visualização de um cliente chegassem.

Mas quando o risco e o impacto de um resultado injusto ou negativo são altos, as pessoas aceitam menos a evolução da IA. Certos tipos de produtos, como dispositivos médicos, podem ser prejudiciais aos usuários se forem alterados sem qualquer supervisão. É por isso que alguns reguladores, principalmente a Administração de Alimentos e Medicamentos dos EUA, autorizaram o uso de apenas algoritmos “bloqueados”—que não aprendem toda vez que o produto é usado e, portanto, não mudam — neles. Para essas ofertas, uma empresa pode executar duas versões paralelas do mesmo algoritmo: uma usada apenas em P&D que aprende continuamente e uma versão bloqueada para uso comercial aprovada pelos reguladores. A versão comercial pode ser substituída em uma certa frequência por uma nova versão baseada em uma versão em constante melhoria — após a aprovação regulatória.

Os reguladores também temem que o aprendizado contínuo possa fazer com que os algoritmos discriminem ou se tornem inseguros de maneiras novas e difíceis de detectar. Em produtos e serviços com os quais a injustiça é uma grande preocupação, você também pode esperar uma maior atenção sobre a capacidade de evolução.

**Complexidade e custo.** A implantação da IA de aprendizado pode aumentar os custos operacionais. Primeiro, as empresas podem estar executando vários algoritmos em diferentes regiões, mercados ou contextos, cada um dos quais respondeu aos dados e ambientes locais. As organizações podem então precisar criar novas funções e processos de sentinela para garantir que todos esses algoritmos estejam operando adequadamente e dentro das faixas de risco autorizadas. Os diretores de risco podem precisar expandir seus mandatos para incluir o monitoramento de processos autônomos de IA e a avaliação do nível de risco legal, financeiro, de reputação e físico que a empresa está disposta a assumir com a IA evoluível. As empresas também devem equilibrar a descentralização com as práticas padronizadas que aumentem a taxa de aprendizado de IA. Eles podem criar e manter um backbone de dados global para impulsionar as soluções digitais e de IA da empresa? Até que ponto seus próprios sistemas estão prontos para armazenamento e processamento descentralizados? Eles estão preparados para responder às ameaças à segurança cibernética? A produção precisa se aproximar dos clientes finais ou isso exporia as operações a novos riscos? As empresas podem atrair talentos suficientes com experiência em IA para as posições corretas de liderança nos mercados locais? Todas essas perguntas devem ser respondidas com atenção.

**Entrada humana.** Novos dados ou mudanças ambientais também podem fazer com que as pessoas ajustem suas decisões ou até alterem seus modelos mentais. Um gerente de recrutamento, por exemplo, pode tomar decisões diferentes sobre o mesmo candidato em dois momentos diferentes se a qualidade dos candidatos concorrentes mudar, ou até mesmo porque ela está cansada na segunda vez. Como não há regulamentação para impedir que isso aconteça, pode-se argumentar que é permitido que a IA evolua como resultado de novos dados. No entanto, seria necessário convencer as pessoas a esse ponto



## **Os reguladores temem que o aprendizado contínuo possa fazer com que os algoritmos discriminem ou se tornem inseguros de maneiras novas e difíceis de detectar.**

O que as pessoas podem aceitar com mais facilidade é a IA complementada de forma inteligente pela tomada de decisões humanas. Conforme descrito no artigo da HBR de 2020 “Uma maneira melhor de integrar a IA” (em coautoria com Theodoros Evgeniou), os sistemas de IA podem ser implantados como “treinadores”, fornecendo feedback e sugestões aos funcionários (por exemplo, negociadores de títulos financeiros em uma empresa de gestão de ativos). Mas não é uma via de mão única: muito do valor da colaboração vem do feedback que os humanos dão aos algoritmos. O Facebook, na verdade, adotou uma abordagem interessante para monitorar e acelerar o aprendizado de IA com seu Plataforma Dynabench. Ele encarrega especialistas humanos de procurar maneiras de induzir a IA a produzir um resultado incorreto ou injusto usando algo chamado coleta dinâmica de dados adversariais.

Quando os humanos aprimoram ativamente a IA, eles podem liberar valor com bastante rapidez. Em uma palestra recente do TED, o BCG Sylvain Duranton descreveu como um varejista de roupas economizou mais de 100 milhões de \$ em apenas um ano com um processo que permitiu que compradores humanos inserissem sua experiência em IA que previa tendências de roupas.

...

Dado que a crescente dependência da IA, especialmente do aprendizado de máquina, aumenta significativamente os riscos estratégicos que as empresas enfrentam, as empresas precisam assumir um papel ativo na elaboração de um livro de regras para algoritmos. À medida que as análises são aplicadas a decisões como

aprovações de empréstimos ou avaliações de reincidência criminal, as reservas sobre preconceitos ocultos continuam aumentando. A opacidade inerente à programação complexa subjacente ao aprendizado de máquina também está causando consternação, e aumenta a preocupação sobre se as ferramentas de IA desenvolvidas para uma população podem tomar decisões com segurança sobre outras populações. A menos que todas as empresas, incluindo aquelas não diretamente envolvidas no desenvolvimento da IA, se envolvam precocemente com esses desafios, elas correm o risco de corroer a confiança em produtos baseados em IA e desencadear regulamentações desnecessariamente restritivas, o que prejudicaria não apenas os lucros comerciais, mas também o valor potencial que a IA poderia oferecer aos consumidores e à sociedade.

A version of this article appeared in the [September–October 2021](#) issue of Harvard Business Review.

## FC

**François Candelon** is a senior partner and managing director at Boston Consulting Group (BCG), and the global director of the BCG Henderson Institute, where his research is focused on the impact of technologies on business and society.

## RC

**Rodolphe Charme di Carlo** is a partner in the Paris office of the Boston Consulting Group.

## MB

**Midas De Bondt** is a project leader in the Brussels office of the Boston Consulting Group.

## TE