

Inteligência Artificial: riscos, benefícios e uso responsável

ANDRÉ CARLOS PONCE DE LEON FERREIRA DE CARVALHO¹

NOS ÚLTIMOS ANOS ocorreu um crescimento acelerado da presença da Inteligência Artificial (IA) no nosso cotidiano. Não são poucas as situações em que usamos, na maioria das vezes sem darmos conta, modelos criados por algoritmos de IA (Faceli, 2021). E usamos em atividades tão rotineiras como ler mensagens enviadas por e-mail, lavar nossas roupas, dirigir um veículo autônomo ou semiautônomo e decidir a que filme ou episódio de uma série assistir em uma plataforma de *streaming*. Existem também situações em que o uso de IA em um produto ou serviço é claramente mencionado, criando muitas vezes a expectativa de alguma coisa melhor. Isso pode ser observado pelo grande crescimento, nestes últimos anos, no número de equipamentos e serviços que usam como peça de propaganda e marketing a divulgação de que são baseados em IA. Mencionar que usa IA virou quase uma garantia de que um produto ou serviço oferecido por uma empresa é superior aos oferecidos pelas empresas concorrentes. Embora isso não seja necessariamente verdade, tanto no real uso de IA quanto na superioridade em relação aos concorrentes.

Mas por que a IA cresceu tanto? Um dos principais motivos para esse crescimento, senão o principal, é o rápido desenvolvimento de novas tecnologias para extração, armazenamento, transmissão e processamento de dados, que serão aqui denominados os quatro eixos de avanços tecnológicos que provocaram a expansão da IA.

Os avanços no primeiro eixo, extração de dados, ocorreram por meio do desenvolvimento de novos sensores, incluindo câmeras, cada vez mais rápidos, baratos e sofisticados. Progressos significativos no segundo eixo, o de armazenamento de dados, também foram possíveis por causa do uso de novos materiais que aumentaram a capacidade de armazenamento, melhorando sua confiabilidade e reduzindo seu custo. No terceiro eixo, transmissão de dados, os avanços estão associados não apenas à queda no custo e no aumento da capacidade dos meios físicos utilizados, mas também à revolução trazida pelas redes de computadores e pela internet, que conecta mais dispositivos, ou coisas, do que pessoas. No início de seu uso, a internet servia para conectar as pessoas em diferentes partes do mundo, bastando para isso um computador e um ponto de acesso. Em 2008, a internet das pessoas foi superada pela internet das coisas. Desde então, a diferença entre o número de coisas e o número de pessoas conectados à internet cresceu tanto que, atualmente, o número de coisas conectadas pela

internet é mais de dez vezes o número de pessoas. Finalmente, no quarto eixo temos computadores cada vez mais eficientes e acessíveis, que, ao mesmo tempo que reduzem de tamanho, aumentam na rapidez com que processam os dados, a um custo, também, cada vez menor.

Também nestes últimos anos, esses avanços tecnológicos levaram à geração do que é hoje chamado de *big data*, que, apesar de ter um grande número de definições, foi inicialmente descrito por três V: Volume, Variedade e Velocidade. O volume diz respeito à quantidade de dados gerados, a variedade está relacionada aos diferentes formatos dos dados, que incluem imagens, sons e vídeos, muito compartilhados por aplicativos de redes sociais, e às diferentes fontes de onde os dados são gerados, como sensores cada vez mais específicos e sofisticados, como os sensores usados para monitorar o corpo quando uma pessoa está realizando exercícios físicos. Vários outros V foram criados depois para incorporar novos aspectos ao *big data*, como valor, veracidade e validade.

O que está compreendido pela expressão *big data* criou uma grande demanda por ferramentas computacionais capazes de explorar os dados gerados, extraindo conhecimento novo, útil e relevante. Esse conhecimento tem sido cada vez mais utilizado para tomada de decisões.

Tudo isso criou um ambiente favorável para que a IA saísse dos laboratórios de pesquisa para ser incorporada em vários produtos e serviços, que tanto geraram ganhos econômicos como sociais. Empresas conseguiram projetar produtos e serviços mais seguros, baratos e personalizados, órgãos públicos puderam oferecer melhores serviços à população e organizações não governamentais conseguiram fiscalizar com maior eficiência ações que pudessem trazer danos a sociedade.

Para que tudo isso possa ocorrer, a maioria dos países passou a olhar a IA não apenas como capaz de criar oportunidades econômicas e sociais, mas também como por sua relação com segurança e por servir de suporte para a criação de novas tecnologias, relacionada à soberania nacional. Isso tem levado a uma corrida por investimentos em IA por diferentes países.

Protagonismo e investimento em IA

Aumenta com velocidade crescente o efeito da IA nos diversos setores da economia mundial. É importante observar que esse efeito já é discutido faz tempo. Em 1966, o estatístico britânico Irving John Good estimou a contribuição de uma máquina ultrainteligente para a economia britânica. Segundo Good (1966), ela seria de um milhão de vezes o que contribuiu a inteligência do economista John Keynes, um dos principais economistas que o mundo conheceu, à economia da Inglaterra. Good estimou o valor da contribuição de Keynes em 100 bilhões de libras esterlinas.

De acordo com um estudo publicado pela consultoria McKinsey em 2018, os bens e serviços que usam IA podem valer em 2030 cerca de 13 trilhões de dólares.¹ Esse futuro mercado é um dos motivos pelos quais vários países vieram a ter um papel de protagonismo em IA. Os países que buscam a liderança em IA

fazem grandes investimentos monetários, criam ou trazem centros de pesquisa e incentivam adoção de IA nos vários setores de sua economia e órgãos do Executivo, Legislativo e Judiciário. Cientes dos efeitos sociais, esses países avaliam o papel da IA na sociedade e criam regulações para prevenir e evitar abusos.

Esse mesmo estudo prevê que a IA deve aumentar as distâncias de desempenho entre países que são líderes em IA, em geral os países desenvolvidos, e os demais países do mundo. Os países líderes em IA podem aumentar em até 25% seus benefícios econômicos, com os demais países apresentando um aumento de cerca de 10%. Previsão semelhante é feita para empresas, com as empresas líderes em IA praticamente dobrando seu lucro e as empresas que demoram a adotar a IA ficando para trás. Finalmente, o mercado de trabalho para profissionais que conseguem incorporar a IA em suas atividades será maior, e pagará melhores salários, que o mercado para profissionais que realizam tarefas repetidas, que podem ser automatizadas.

Por isso, vários países associam o protagonismo internacional em IA ao fortalecimento da soberania nacional. Para isso, esses países investem pesadamente na captação, na retenção e na formação de talentos na área de IA e na criação de uma estrutura de pesquisa básica e aplicada capaz de gerar e transferir conhecimento para o setor produtivo. Essa transferência, além de agregar valor ao que é produzido no país, estimula a criação de *startups* inovadoras. Com isso, esses países conseguem não apenas dar conta de suas demandas internas, mas também melhorar a balança comercial, trazer novos investimentos e criar postos de trabalhos bem remunerados.

Em 2017, mais de 15 bilhões de dólares foram investidos no mundo em *startups* de IA, 48% desse montante foram investidos na China e 38%, nos Estados Unidos.

Um estudo de janeiro de 2021 aponta os dez países líderes em IA. Dessa lista fazem parte China, Estados Unidos, Reino Unido, Canadá, Rússia, Alemanha, Noruega, Suécia, França e Índia.²

O governo chinês incluiu objetivos ambiciosos para o desenvolvimento da IA em seu 13º plano de cinco anos, que abrange o período de 2016 a 2020. O país quer ser o líder mundial de IA até 2030, investindo desde chips a algoritmos, com foco nas áreas de cidades inteligentes, defesa e manufatura. Para isso, o Conselho de Estado da China, também conhecido como Governo Popular Central, pretende criar, até 2030, um grande mercado doméstico de IA.

Nesse plano, cuja preparação teve apoio de grandes empresas chinesas de tecnologia da informação, Baidu, Alibaba e Tencent, é estimado que em 2030 o valor das empresas chinesas na área de IA atingirá 150 bilhões de dólares. Estima ainda que o valor das empresas que utilizam IA será de 1,5 trilhão de dólares.

A China já conta com grandes empresas que oferecem alternativas para produtos e serviços na área de IA oferecidos por grandes multinacionais do Ocidente. Essas empresas chinesas incluem a Didi, na área de transporte (seme-

lhante a Uber), a Baidu, para busca na Internet (similar a Google), e a Taobao, no ramo de comércio eletrônico (que concorre com a Amazon). Assim como a Google, a Baidu está investindo em carros autônomos. Desde 2020 a Baidu está testando seu veículo autônomo, Apollo, nas ruas da cidade de Beijing. Em 2021, a Baidu iniciou uma parceria com uma fabricante de automóveis chinesa, a Geely Automobile Holdings, sediada em Hong Kong, para o desenvolvimento de veículos elétricos autônomos.³

Além disso, um número crescente de cidades chinesas está desenvolvendo seus próprios programas de investimento em IA. A cidade de Beijing, por exemplo, planeja investir 2,1 bilhões de dólares na construção de um parque de pesquisa para desenvolvimento de tecnologias de IA. O Parque contará com até 400 empresas e tem como focos pesquisa básica, produção de patentes e cidades inteligente.

O Reino Unido também busca protagonismo em IA. De acordo com o braço de pesquisa da consultoria McKinsey, o Instituto Global da McKinsey (MGI, do original em inglês McKinsey Global Institute), o Reino Unido, que já é o país europeu com maior densidade de *startups* de IA na área de saúde, deverá ter aproximadamente 16% do mercado mundial de IA em 2030.⁴

Em 2017, a Câmara dos Lordes do Reino Unido indicou um Comitê de Inteligência Artificial, composto por membros da Câmara de Lordes, para realizar um estudo sobre as implicações éticas, econômicas e sociais do avanço da IA. Em 2018 o Comitê publicou seu estudo em um documento, que recebeu o nome de “IA no Reino Unido: pronto, querendo e capaz?” (do original em inglês *AI in the UK: ready, willing and able?*).⁵ Nesse documento, após argumentar que o Reino Unido já pode assumir um protagonismo mundial na área de IA, apresentou uma série de recomendações para ações necessárias para apoiar o governo a identificar o potencial da IA para a sociedade e a economia britânica, e para proteger a sociedade de futuras ameaças e riscos.

Entre as recomendações está a criação de um fundo para empresas de pequeno e médio portes que trabalham com IA, para dar escala aos seus negócios, a complementação do financiamento de doutorados em IA pelo setor privado, a padronização dos mecanismos para a criação de *startups* em IA a partir de pesquisas realizadas por universidades britânicas e o reconhecimento da importância de especialistas estrangeiros para o sucesso da IA no Reino Unido com o aumento da concessão de vistos para trazer especialistas talentosos.

Atualmente, o governo britânico possui um órgão específico para Inteligência Artificial (Office for Artificial Intelligence), que é vinculado a dois departamentos governamentais: Departamento para Cultura, Mídia e Esporte Digital e Departamento para Negócios, Energia e Estratégia Industrial. Esse órgão, por meio de um comitê de especialistas em Inteligência Artificial, produziu no início de 2021 um relatório intitulado “Roteiro para IA” (do original em inglês *AI Road-map*),⁶ que transmite duas mensagens: a primeira é que o Reino Unido

precisa dobrar os investimentos feitos recentemente em IA; a segunda é que é necessário olhar o horizonte de forma a conseguir se adaptar a disrupções. O texto menciona ainda que o Reino Unido se beneficiará completamente da IA apenas se todos os setores da sociedade tiverem total confiança na ciência e tecnologia, assim como no governo e na regulação que os disponibiliza. Completa falando que essa confiança dependerá da existência de sistemas que garantem completa responsabilização, ética e transparência.

Na França, o presidente Emmanuel Macron anunciou plano de investimento de US\$ 1,8 bilhão até 2022 para transformar a França em um dos líderes mundiais em IA. Esses recursos serão utilizados especialmente para criação de institutos especializados, disponibilização de dados e recrutamento de pesquisadores estrangeiros, além de franceses trabalhando no exterior. É importante mencionar que ao final da presidência de François Holland, o antecessor do presidente Emmanuel Macron, o governo produziu o relatório sobre a importância da IA para a França.

Em 2017, o matemático francês Cédric Villani, ganhador da Medalha Fields, considerada o Prêmio Nobel da Matemática, e um dos deputados da Assembleia Nacional da França, encabeçou, a pedido do primeiro-ministro francês Édouard Philippe, uma força-tarefa discutir a estratégia de IA para a França e a Europa. Essa força-tarefa foi a responsável pela preparação de um relatório com cerca de 150 páginas, intitulado “Por uma Inteligência Artificial que faça sentido: em direção a uma estratégia francesa e europeia”.⁷

Nesse relatório, Villani discorre que, assim como a eletricidade, a IA estará em todo lugar. Fala ainda da crescente presença da IA na economia e na sociedade. Os temas cobertos pelo relatório incluem políticas econômicas baseadas em dados, como estimular uma pesquisa ágil em IA, como antecipar e controlar o impacto da IA nas habilidades profissionais necessárias e no mercado de trabalho, a contribuição da IA para criar uma economia mais ecológica, a ética da IA e por uma IA inclusiva e que respeite a diversidade.

Um efeito colateral do crescimento da IA é que, em todo o mundo, estão surgindo cursos de graduação em IA. Um outro efeito colateral é que vários pesquisadores de renome na área estão saindo das Universidades para trabalhar em grandes empresas ou criar *startups*. Vários alunos de mestrado e de doutorado estão fazendo o mesmo, muitos antes de concluírem seus cursos.

Uma pesquisa internacional recente estima que quase 50% dos consumidores se animam mais a comprar um produto, quando ele tem IA. Isso motiva várias empresas a usar como estratégia comercial um destaque da sigla IA na descrição de seu produto ou serviço. Mesmo não tendo IA tendo mais artificial que inteligência, tendo uma IA muito simples, muito inferior àquela produzida nos grandes centros de pesquisa, ou ainda utilizando IA de forma incorreta.

No Brasil, os movimentos de estímulo ao uso, e mais importante, a produção de pesquisas, inovações e tecnologias de IA internacionalmente competitivas

ocorrem de forma isolada, tímida, sem estratégia e foco claros. O país, infelizmente, está se acomodando e se contentando com o papel de futuro consumidor de produtos e serviços baseados em IA.

Nesse cenário, alguns centros de pesquisa nacionais acabam sendo apenas vitrines para facilitar a venda dos produtos da empresa ou para reduzir os impostos pagos pela empresa. Para realmente contribuir para o desenvolvimento da ciência, tecnologia e inovação no país, um centro de IA nacional tem que ser internacionalmente competitivo, com pesquisas da mesma qualidade daquelas realizadas nos principais centros de pesquisa do exterior.

Para não depender de empresas ou de outros países, e, assim, não perder competitividade, o Brasil precisa criar produtos e formar especialistas com padrão e qualidade internacionais. Isso porque, sem tecnologia nacional, o país se verá obrigado a pagar *royalties* e, mesmo assim, não terá acesso a todas as funcionalidades da ferramenta de IA.

As empresas brasileiras precisam ter capacidade de competir em pé de igualdade com as de qualquer outro país do mundo, exportando conhecimento e gerando empregos e riqueza para o Brasil.

Riscos e receios

Como toda tecnologia, especialmente aquelas que devem seguir cuidadosamente uma metodologia internacionalmente consolidada, a IA pode ser utilizada de forma incorreta, ou até mesmo perigosa. Não são poucos os casos em que o uso incorreto da IA tem causado sérios danos aos consumidores e à reputação de uma empresa.

A preocupação com a automação já existe há muito tempo. Platão, discípulo de Sócrates e um dos principais filósofos gregos, para transmitir os pensamentos de Sócrates, que não deixou nada escrito, escreveu vários diálogos em que Sócrates era o protagonista, conhecidos como diálogos socráticos.

Nesses diálogos, dois ou mais personagens discutiam problemas filosóficos e morais. Em um desses diálogos, Fedro, os personagens Sócrates e Fedro, um aristocrata de Atenas, discutiam sobre a arte da retórica e como ela deve ser praticada. No diálogo, Sócrates questiona os aspectos positivos e negativos da escrita, insistindo que a escrita destrói a memória e enfraquece a mente (Reis, 2016).

Para justificar sua teoria, Sócrates conta uma história em que dois deuses egípcios, Theuth e Thamus, debatem o mérito de apresentar textos escritos ao povo. Nele, Theuth argumenta que eles tornarão os egípcios mais sábios e com uma memória melhor. Thamus, por sua vez, acreditava que isso tornaria as pessoas dependentes dos textos escritos, deixando de usar sua memória.

Dando um salto no tempo, e já falando explicitamente em IA, poucas décadas atrás, a IA já fazia parte do enredo de filmes de ficção científica. Em geral ajudando o vilão ou sendo o próprio vilão. Esse estereótipo pode ter contribuído para que as pessoas tenham receio, quando não medo, da presença cada vez maior da IA em sua vida.

No mesmo artigo em que estima o valor da IA para a economia britânica, Good (1966) fala do risco da IA, em particular em tornar a raça humana redundante. Good (1966, p.33) especula sobre uma consequência de termos máquinas ultra inteligentes.

Seja uma máquina ultrainteligente definida como uma máquina que pode superar de longe todas as atividades intelectuais de qualquer homem, por mais inteligente que seja. Como o projeto de máquinas é uma dessas atividades intelectuais, uma máquina ultrainteligente poderia projetar máquinas ainda melhores; haveria, então, inquestionavelmente, uma “explosão de inteligência” e a inteligência do homem seria deixada para trás. Assim, a primeira máquina ultrainteligente é a última invenção que o homem precisa fazer sempre, desde que a máquina seja dócil o suficiente para dizer nos como mantê-lo sob controle. É curioso que isso raramente é abordado fora da ficção científica. As vezes vale a pena levar a ficção científica a sério.

Em um artigo recente, Price (2018) discute o balanço entre segurança e autonomia em um mundo com máquinas superinteligentes. Good discorre ainda sobre problemas éticos, como se uma máquina pode sentir dor e se a máquina deve ser desmontada se ela se tornar obsoleta. Segundo ele, a IA pode causar problemas sociais, mas também pode resolvê-los. Menciona ainda que os riscos apontados são reais e urgentes, que devem ser discutidos fora da ficção científica. Para ele, não estava claro se os impactos da IA seriam positivos ou negativos.

Uma das principais incertezas sobre nossa relação com máquinas inteligentes é como lidar com conflitos entre máquinas e seres humanos. Havendo uma disputa entre máquinas e seres humanos, é importante observar que enquanto a estrutura e o comportamento dos seres humanos são guiados por lentos processos bioquímicos, as máquinas, com os avanços na ciência de materiais e na capacidade de processamento de dispositivos computacionais, são cada vez mais resistentes e eficientes.

É importante observar que toda novidade, em particular novas tecnologias, desperta esperanças e receios. Um dos exemplos mais antigos foi a descoberta de como dominar o fogo, que pode ser usado tanto para aquecer como para machucar. Da mesma forma, todo avanço científico, tecnológico ou inovador pode trazer benefícios ou malefícios, dependendo da forma como é utilizado. Avanços na biologia permitem, por exemplo, criar organismos que nunca existiram, ou novos vírus capazes de trazer graves riscos à saúde de seres vivos. Por outro lado, permite também criar novos medicamentos capazes de reduzir o sofrimento ou levar à cura. Progressos na química podem apoiar a criação de novos materiais, que podem tanto corroer ou destruir objetos, como causar danos que provocam acidentes que podem machucar ou matar seres vivos. Mas podem também criar próteses que reduzem o desgaste físico e que melhoram a vida das pessoas. Na área jurídica, novas leis ou novas interpretações de leis

existentes podem pôr em risco o meio ambiente, a liberdade e o bem-estar das pessoas. Pesquisas na área de comunicação podem aumentar a disseminação e a assimilação de notícias falsas.

IA no mercado de trabalho

Um dos receios associados a isso é como a IA afetará o mercado de trabalho. Vários estudos mostram que atividades profissionais desaparecerão, sendo substituídas por atividades que até o momento são desconhecidas ou inimagináveis. Com frequência, são divulgadas listas com as profissões com maior probabilidade de desaparecer no futuro.

A IA tem como possíveis benefícios reduzir a exposição das pessoas a situações e atividades de risco e diminuir a necessidade de realizar tarefas cansativas repetitivas e monótonas, liberando os seres humanos para atividades mais agradáveis e desafiadoras. Um efeito colateral é reduzir o número de horas dedicadas ao trabalho, permitindo mais tempo com família e amigos e em atividades de lazer. Fenômeno semelhante ocorreu na Revolução Industrial.

Nos Estados Unidos, no século XIX, a substituição do tear manual pelo tear mecânico automatizado fez que a quantidade de tecido produzido por uma pessoa amentasse em 50 vezes. Com isso, o número de pessoas necessárias para tecer diminuiu em 98%. Como consequência, foi reduzido o preço do tecido, permitindo sua compra por um número maior de pessoas, e gerados mais empregos, por exemplo, no setor de corte e costura. A Revolução Industrial permitiu ainda a redução de horas de trabalho, de 10 a 16 horas por dia para 8 horas por dia.

IA responsável

Existem várias propostas sobre que aspectos definem o uso responsável da IA, propostos por diferentes empresas e grupos de pesquisa. Em 2019, o grupo independente de especialistas em Inteligência Artificial da União Europeia, criado em 2018, publicou um documento intitulado “Orientações éticas para uma IA confiável”, que recebeu mais de 500 comentários quando um primeiro rascunho foi disponibilizado para consulta pública.⁸ O documento define que uma IA responsável, denominada confiável, deve ser legalmente válida (respeita todas as leis e regulações aplicáveis), ética (atende aos princípios e valores éticos) e robusta (sob as perspectivas técnicos e sociais).

A IA Responsável engloba aspectos como justiça, transparência e privacidade. Esses temas serão abordados nas próximas seções, seguidas de um tema que pode prevenir ou controlar efeitos danosos destes aspectos, a regulação da IA.

IA justa

É cada vez maior a presença da IA nos processos de tomada de decisão. Conforme apontado em Grđic-Hlaca et al. (2018), os algoritmos estão cada vez mais tomando decisões que afetam vidas humanas. Para confiar e aceitar essas

decisões, as pessoas devem sentir que elas são justas. Não são poucas as notícias que reportam decisões preconceituosas, com viés, tomadas por modelos gerados por algoritmos de IA. Um caso típico é de uma ferramenta para identificação de criminosos baseada em IA que utilizava informação racial para a tomada de decisão. Com isso, a ferramenta retornava com uma maior probabilidade de ser criminoso pessoas de uma dada raça.

Uma das principais fontes de preconceito está no conjunto de dados utilizados para gerar modelos, como é o caso das aplicações que geram modelos por meio da aplicação de algoritmos aprendizado de máquina aos dados. Por isso, é preciso ter muito cuidado na hora de coletar os dados. Quando um algoritmo de aprendizado de máquina é aplicado a um conjunto de dados, ele busca por padrões nos dados e reforça-os na criação de um modelo que represente os dados. Se os padrões tiverem algum viés, o modelo gerado terá o viés.

A IA justa aumenta a proteção contra a geração de modelos preconceituosos devido a padrões extraídos de dados pessoais (Hajian et al., 2014). Ela busca evitar que decisões tomadas por modelos gerados por algoritmos de IA sejam preconceituosas, que levem em conta aspectos como classe social, credo, doença preexistente, idade, nacionalidade, orientação sexual e raça na tomada de decisão. O uso de variáveis representando esses aspectos, conhecidas como variáveis sensíveis, também podem levar a modelos preconceituosos. Por isso, muitas vezes os valores destas variáveis são usados de forma criptografada.

É importante observar que informações que levam a modelos preconceituosos podem inclusive estar presentes de forma indireta nos dados. Por exemplo, se uma das variáveis do conjunto for renda e se diferentes grupos raciais tiverem diferentes rendas, o grupo racial das pessoas é usado de forma indireta. Uma forma de avaliar se um modelo é preconceituoso é utilizar algoritmos que geram modelos transparentes.

IA transparente

Uma IA transparente permite que modelos gerados por algoritmos de IA para tomada de decisão que afete a vida das pessoas devem ser facilmente interpretados. Assim, a forma como o modelo tomou uma decisão deve ser clara o suficiente para que qualquer pessoa possa compreender. Essas decisões podem ser, por exemplo, se uma pessoa tem direito ou não a um empréstimo financeiro, se uma pessoa está ou não acometida por dada doença, ou se um doente deve seguir um tratamento A ou B. O nível de interpretabilidade dos modelos gerados por algoritmos de IA divide-os em modelos caixa-preta, quando não é possível interpretar como eles tomam uma decisão, caixa-branca, quando qualquer pessoa pode entender como a decisão é tomada, e caixa-cinza, quando o modelo é parcialmente interpretável (Yang, 2017).

A transparência está fortemente associada com a confiança que as pessoas têm nas decisões tomadas por modelos gerados por algoritmos de IA. Quanto maior a transparência de um modelo, maior a confiança que as pessoas têm

nas decisões. A transparência permite ainda validar as decisões tomadas pelos modelos. A importância dessa transparência é reforçada numa linha da IA, chamada IA explicável (XAI, do termo original em inglês *Explainable AI*) (Arrieta, 2020). Um dos objetivos dos pesquisadores desta área é conseguir interpretar o conhecimento utilizado por modelos caixa-preta (Botari, 2020).

De acordo com Weller (2019), a transparência é geralmente uma questão crítica para o emprego efetivo de soluções baseadas em IA. Ainda segundo o autor, tanto a motivação quanto os benefícios da transparência dependem, em grande parte, do contexto, sendo assim difícil avaliar e identificá-la de forma objetiva.

IA que garante a privacidade e proteção de dados

Um outro aspecto importante para o uso responsável da IA é o direito à privacidade e proteção de seus dados. O direito à privacidade de dados pessoais pode ser resumido pelo controle, por parte de um indivíduo, da coleta, do uso e do compartilhamento de seus dados pessoais. Com a grande expansão no volume de dados gerados, tornaram-se comuns notícias de vazamento acidental ou criminoso de dados pessoais.

Para melhor lidar com o crescente uso e geração de dados, a União Europeia atualizou em 2016 sua legislação para proteção de dados, aprovando o Regulamento Geral sobre a Proteção de Dados (GDPR, do original em inglês *General Data Protection Regulation*).⁹ O GDPR regula por meio de várias regras a coleta, o armazenamento e o uso de informação pessoal. Uma dessas é o direito a explicação. De acordo com esse direito, organizações públicas, privadas e sem fins lucrativos que utilizem dados pessoais de cidadãos da União Europeia têm a obrigação de garantir o processamento justo e transparente desses dados. O GDPR garante aos órgãos europeus de controle de uso de dados os poderes de investigar e aplicar multas.

Inspirado no GDPR, o Congresso brasileiro aprovou em 2018 a Lei Geral de Proteção de Dados Pessoais (LGPD), que entrou em vigor em 2020.¹⁰ Leis como o GDPR e a LGPD contribuem para o aumento da transparência da IA. Para fiscalizar o cumprimento da LGPD foi criada em 2018 a Autoridade Nacional de Proteção de Dados (ANPD).

Em um estudo publicado em 2015 (Youyou et al., 2015), os autores aplicaram um algoritmo de aprendizado de máquina a um conjunto de dados gerados por voluntários usando uma conhecida rede social. Os conjuntos de dados eram formados apenas por *likes* dados pelos voluntários. Os modelos gerados pelos algoritmos precisaram ter acesso a apenas dez *likes* para conhecer pessoas melhores que seus colegas de trabalho, 150 para conhecer melhor que um membro da família, e 300 para conhecer melhor que o cônjuge. É importante mencionar que a aplicação de aprendizado de máquina a dados de redes sociais pode trazer vários benefícios, um deles é a redução de assédio nestas redes (Peireira, 2020).

Regulação da IA

A regulação da IA é alvo de debate no mundo inteiro e já está ocorrendo em vários países, frequentemente despertando reações antagônicas. No Brasil, até o final de 2020, contava com quatro propostas de regulação no parlamento, duas na Câmara dos Deputados e duas no Senado Federal. Caso não seja regulada, pode trazer danos a indivíduos, comunidades, sociedade e meio ambiente. Caso seja, pode limitar inovações em IA, reduzindo seus potenciais benefícios.

Etzioni (2018) defende que o que deve ser regulado não é a IA, mas as aplicações da IA. Para justificar sua proposta, argumenta que a regulação é um processo lento, sujeito a distorções e interferências políticas. A regulação de aplicações de IA, ao invés da IA, permite um equilíbrio entre os potenciais benefícios e riscos da IA, ficando no meio termo entre uma regulação que não seja o suficiente e uma regulação excessiva.

Segundo o autor, uma das dificuldades para a regulação é definir o que é realmente IA. O crescente número de aplicações de IA torna a definição cada vez mais difícil. O autor propõe cinco regras ou direções para a regulação da IA:

- Não tornar IA em armas: regular contra uso de armamentos e ciberarmamentos autônomos, que utilizem IA.
- O responsável não é a IA: o uso da IA deve estar sujeito às leis existentes que se aplicam aos seres humanos. Pessoas devem ser responsáveis por danos causados pela IA. Algumas vezes o culpado é quem desenvolveu a IA, em outras, é o dono da ferramenta que usa IA, e outras vezes a culpa deve recair sobre quem está usando a IA.
- Aplicações de IA devem deixar claro que não são pessoas. Assim, *chatbots* e equivalentes devem se identificar como máquinas e se comentários sobre notícias e propagandas forem feitos por máquinas, isso deve estar claro para quem os recebe.
- Ferramentas que possuem IA não devem manter ou divulgar informação sem aprovação explícita da fonte. Para isso, deve estar de acordo com as boas práticas de respeito a privacidade, protegendo as pessoas de mau uso de dados coletados por dispositivos inteligentes (do original em inglês *smart devices*).
- IA não deve aumentar qualquer preconceito que exista nos sistemas atuais, mas sim reduzir. Isso está associado a uma importante linha de pesquisa para uma IA sem preconceitos. Muitas vezes, isso requer cuidado com variáveis sensíveis, que podem ser usadas para julgamentos preconceituosos.

A regulação da IA não é apenas um problema tecnológico, que possa ser facilmente atualizado para lidar com a criação de e com novas tecnologias. É também um problema de ciências sociais, requerendo para isso a contribuição de especialistas em ciências humanas, que possam melhor avaliar os diferentes aspectos que afetam a vida das pessoas. A escrita de uma boa regulação deve

passar por discussões e debates com a sociedade, para considerar os diferentes ângulos, ser clara e continuar válida no futuro.

Um exemplo típico de uma regulação que pode dar margens a diferentes interpretações é a aprovação de uma lei na Califórnia proibindo motociclistas de andarem sem capacete. Logo depois de entrar em vigor, um policial parou um motociclista que estava sem usar um capacete na cabeça. O motociclista sugeriu ao policial que lesse novamente a lei. A lei da Califórnia exigia que motociclistas usassem um capacete à prova de quebra e que o capacete estivesse firmemente apertado. O guarda não pôde multar o motociclista, pois ele tinha um capacete a prova de quebra e firmemente apertado, ao seu joelho

A regulação da IA é alvo de intensos debates, inclusive no Brasil.

Conclusões

Independentemente de nossos desejos, a IA já está se tornando íntima de todos nós. O que temos que decidir agora não é mais se teremos ou não a IA, mas como teremos a IA. Para isso, temos que, prestando atenção em experiências passadas, de acesso e inclusão social a novas tecnologias, garantir que a IA veio para beneficiar a todos nós e que os riscos que receamos possam ser evitados. Isso é inseparável de um uso responsável da IA, que é uma IA justa, transparente e que respeite a privacidade das pessoas. Como toda nova tecnologia, a razão de sua existência deve ser melhorar a vida das pessoas, não deixando ninguém para trás.

Notas

- 1 2 Notes from the ai frontier modeling the impact of ai on the world economy. Disponível em: <<https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx>>. Acesso em: 2 mar. 2021.
- 2 AI Artificial Intelligence Investment by Top 10 Countries. Disponível em: <<https://www.analyticsinsight.net/artificial-intelligence-investment-by-top-10-countries/>>. Acesso em: 2 mar. 2021.
- 3 Chinese search giant Baidu to create an electric vehicle company. Disponível em: <<https://www.cnbc.com/2021/01/11/chinese-search-giant-baidu-to-create-an-electric-vehicle-company.html>> Acesso em: 2 mar. 2021.
- 4 Artificial intelligence in the United Kingdom: Prospects and challenges. Disponível em: <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Artificial%20intelligence%20in%20the%20United%20Kingdom%20Prospects%20and%20challenges/Artificial-intelligence-in-the-United-Kingdom-VF2.ashx>>. Acesso em: 2 mar. 2021.
- 5 AI in the UK: ready, willing and able?. Disponível em: <https://ec.europa.eu/jrc/communities/sites/jrcctties/files/ai_in_the_uk.pdf> Acesso em: 4 mar. 2021.

- 6 AI Roadmap. Disponível em: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949539/AI_Council_AI_Roadmap.pdf>. Acesso em: 2 mar. 2021.
- 7 AI for humanity. Disponível em: <<https://www.aiforhumanity.fr/en/>>. Acesso em: 2 mar. 2021.
- 8 Ethics guidelines for trustworthy AI, 2019. Disponível em: <<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>>. Acesso em: 2 mar. 2021.
- 9 Complete guide to GDPR compliance. Disponível em: <<https://gdpr.eu/>>. Acesso em: 2 mar. 2021.
- 10 Lei Geral de Proteção de Dados Pessoais. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709compilado.htm>. Acesso em: 2 mar. 2021.

Referências

- ARRIETA, A. B. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, v.58, p.82-115, 2020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1566253519308103?via%3Dihub>>.
- BOTARI, T. et al. Local Interpretation Methods to Machine Learning Using the Domain of the Feature Space. 30th European Conference on Machine Learning and Principles and 23rd Practice of Knowledge Discovery in Databases, PKDD/ECML 2019, Workshops (1), Communications in Computer and Information Science, v.1167, p.241-52, 2020. Springer. Disponível em: <https://doi.org/10.1007/978-3-030-43823-4_21>.
- ETZIONI, O. Point: Should AI technology be regulated?: yes, and here's how. *Communications ACM*, v.61, n.12, p.30-2, 2018. Disponível em: <<https://cacm.acm.org/magazines/2018/12/232893-point-should-ai-technology-be-regulated/fulltext>>.
- FACELI, K. et al. *Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina*. 2. edição. GrupoGen, 2021. Disponível em: <<https://www.grupogen.com.br/e-book-inteligencia-artificial-uma-abordagem-de-aprendizado-de-maquina>>.
- GOOD, I. J. Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers*, v.6, p.31-88, 1966. Disponível em: <[https://doi.org/10.1016/S0065-2458\(08\)60418-0](https://doi.org/10.1016/S0065-2458(08)60418-0)>.
- GRGIC-HLACA et al. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In: PROCEEDINGS OF THE 2018 WORLD WIDE WEB CONFERENCE (WWW '18)., Geneva, p.903-12. DOI:<https://doi.org/10.1145/3178876.3186138>
- HAJIAN, S. et al. i. Fair pattern discovery. In: PROCEEDINGS OF THE 29TH ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING (SAC '14). Association for Computing Machinery, New York, USA, p.113-20, 2014. DOI:<https://doi.org/10.1145/2554850.2555043>

O'KEEFE, K.; BRIEN, D. O. *Ethical Data and Information Management: Concepts, Tools and Methods*. s. l.: Kogan Page Ltd., 2018. Disponível em: <<https://dl.acm.org/doi/abs/10.5555/3265200>>.

PEREIRA, F. S. F. et al. Gradient Boosting Machine and LSTM Network for Online Harassment Detection and Categorization in Social Media. 30th European Conference on Machine Learning and Principles and 23rd Practice of Knowledge Discovery in Databases, PKDD/ECML 2019, Workshops (2), Communications in Computer and Information Science, v.1168. p.314-320, 2020 Springer. Disponível em: <https://doi.org/10.1007/978-3-030-43887-6_25>.

PRICE, H, VOLD, K. Living with AI. *Research Horizons*, v.I, n.35, p.20-1, February 2018. University of Cambridge. Disponível em: <https://www.cam.ac.uk/system/files/issue_35_research_horizons.pdf>.

REIS, M. C. G. *Fedro*. São Paulo: Penguin Companhia, 2016. Disponível em: <<https://www.companhiadasletras.com.br/detalhe.php?codigo=85146>>.

YANG, Z. et al. Investigating Grey-Box Modeling for Predictive Analytics in Smart Manufacturing. Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 2B: 43rd Design Automation Conference. Cleveland, Ohio, USA. August 6–9, 2017. V02BT03A024. ASME. Disponível em: <<https://doi.org/10.1115/DETC2017-67794>>.

YOUYOU, W. et al. Computers judge personalities better than humans. *Proceedings of the National Academy of Sciences*, v.112, n.4, p.1036-1040, jan. 2015. Disponível em: <<https://www.pnas.org/content/112/4/1036>>.

WELLER A. Transparency: Motivations and Challenges. In: SAMEK, W. et al. (Ed.) *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Lecture Notes in Computer Science, v.11700, p.23-40, Springer, 2019. Disponível em: <https://link.springer.com/chapter/10.1007%2F978-3-030-28954-6_2>.

RESUMO – Estamos usando tecnologias baseadas em Inteligência Artificial em um número crescente de sistemas e ferramentas. A Inteligência Artificial pode tornar reduzir a necessidade da presença humana em muitas atividades perigosas, monótonas e cansativas, nos liberando para atividades menos perigosas e mais desafiadoras e estimulantes. Ao mesmo tempo, a Inteligência Artificial pode aumentar riscos existentes e trazer novos riscos. Para evitar ou reduzir esses riscos, é necessário o desenvolvimento de novos algoritmos de Inteligência Artificial, ou seu uso de maneiras novas e inovadoras, levando em consideração questões éticas, sociais e legais.

PALAVRAS-CHAVE: Inteligência Artificial, Aprendizado de Máquina, IA Responsável.

ABSTRACT – We are using Artificial Intelligence-based technologies in an increasing number of systems and tools. Artificial Intelligence can reduce the need for human presence in many dangerous, monotonous and tiring activities, freeing us for less dangerous and more challenging and stimulating activities. At the same time, Artificial Intelligence can increase existing risks and introduce new risks. To avoid or reduce these risks, new Ar-

tificial Intelligence algorithms must be developed or used in new and innovative ways, taking into account ethical, social and legal issues.

KEYWORDS: Artificial Intelligence, Machine Learning, Responsible AI.

André Carlos Ponce de Leon Ferreira de Carvalho é professor titular do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (ICMC-USP), campus São Carlos. Bolsista de Produtividade em Pesquisa 1A do CNPq e vice-presidente da Sociedade Brasileira de Computação (SBC). @ – andre@icmc.usp.br / <https://orcid.org/0000-0002-4765-6459>.

Recebido em 5.3.2021 e aceito em 12.3.2021.

¹ Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos, São Paulo, Brasil.