

Panorama setorial da Internet

Inteligência Artificial em perspectiva

Por Carolina Bigonha¹

#TechforGood

À medida que a Quarta Revolução Industrial ganha ritmo, novas tecnologias moldam o futuro da produção, da mobilidade, do comércio, da informação e do entretenimento, dos sistemas financeiros, da saúde e do bem-estar, da educação, do consumo, da economia e de muitas outras esferas da vida cotidiana. Combinada com outras tecnologias emergentes – como Internet das coisas, *blockchain*, robótica avançada, novos materiais –, Inteligência Artificial (IA) já funciona como um grande motor para a transformação digital das organizações, dos governos, das indústrias e das nossas vidas. Agricultura inteligente, veículos elétricos autônomos e conectados, soluções de resposta a desastres naturais, cidades inteligentes, previsão e modelagem climática para o gerenciamento de oceanos e de florestas são apenas alguns exemplos de aplicações diretas de IA para impactar positivamente desafios mundiais.

Quando combinadas com os métodos tradicionais de estatística e de cálculo de indicadores, técnicas de IA também podem contribuir

para os atuais processos de geração de dados em pontos como precisão, cobertura, custo e velocidade. Além de sofisticar a análise e a detecção de padrões, ampliando o entendimento e a capacidade de tomada de decisão, a inclusão de dados e realidades, os quais antes não poderiam ser acessados, coletados ou compreendidos eficientemente, contribui diretamente para a missão da Agenda 2030 da Organização das Nações Unidas de "não deixar ninguém para trás". Exemplos de IA para mensuração incluem: novas oportunidades no processamento e análise de imagens de satélite – como mapeamento de escolas em regiões isoladas, detecção de desmatamento em tempo real, entendimento de fluxos urbanos –; uso de mídias digitais para prever epidemias e tendências sociais, previsão de padrões de saneamento e de consumo para melhorar a provisão de água potável e de saneamento; entendimento de padrões e otimização de consumo energético; e muitas outras possibilidades.

¹ Empreendedora, Cientista de Dados e especialista em Produto, Carolina é co-fundadora da Hekima, empresa de tecnologia focada em soluções de Inteligência Artificial. É Global Shaper, uma iniciativa criada pelo Fórum Econômico Mundial, e co-fundadora do Em Perspectiva, uma iniciativa que visa promover uma narrativa positiva, inclusiva, ética e responsável para Inteligência Artificial no Brasil. [carolina.bigonha@hekima.com]

Enquanto lideranças e cientistas de dados do mundo todo constroem produtos, projetos e desenham soluções com potencial de afetar diretamente a forma como nós vivemos e nos relacionamos, há uma importante discussão sobre o melhor uso da Inteligência Artificial para a sociedade. Tornar éticas e inclusivas tais tecnologias e aplicações requer uma colaboração proativa entre cientistas de dados, sociedade civil, formuladores de políticas, governos, setor privado, investidores, especialistas. A Quarta Revolução Industrial também trará grandes mudanças no mercado de trabalho: novos formatos e tipos de emprego surgirão e uma parcela significativa dos atuais diminuirá drasticamente.

Este artigo conceitua Inteligência Artificial e coloca em perspectiva as principais oportunidades e cuidados necessários para que essa revolução seja uma transformação positiva para a sociedade.

O que é Inteligência Artificial?

Inteligência Artificial é um campo de estudo que surgiu na década de 50, cujo objetivo principal é o estudo e a construção de sistemas capazes de exibir comportamentos normalmente associados às pessoas, como aprendizado e resolução de problemas. Algumas linhas de estudo de Inteligência Artificial são mais focadas em reproduzir a maneira como pensamos e raciocinamos, ao passo que outras se concentram no entendimento e na simulação de comportamento.

Antes restritas aos grandes centros de pesquisa, tecnologias de Inteligência Artificial estão hoje inseridas no mercado, nos produtos que consumimos e em vários aspectos de nossas vidas. A crescente popularidade de tais tecnologias está diretamente ligada à abundância e ao barateamento da infraestrutura para processamento, aos avanços em algoritmos, à maior disponibilidade de dados, à disponibilidade dessas tecnologias em código aberto e até à maior conectividade do nosso mundo atual.

A capacidade aumentada de processamento destrava um ponto chave: a velocidade para a tomada de decisão, gargalo para qualquer tipo de automação. Quanto mais avançadas e acessíveis se tornam as tecnologias de processamento e de armazenamento de dados, mais poderosos ficam os sistemas de Inteligência Artificial, simplesmente porque eles conseguem tomar decisões mais complexas em tempo hábil. Imagine o poder computacional necessário para um carro autônomo decidir, em milésimos de segundos, se ele deve desviar para a direita ou para a esquerda, em virtude de um obstáculo na pista, determinar sua velocidade atual ou perceber outros veículos na estrada, para citar somente alguns fatores.

Se os algoritmos são o motor, certamente dados são o combustível dessa revolução tecnológica. É marcante em nossa atualidade o crescente volume de dados produzidos e de informação disponíveis, cuja diversidade passa por diferentes naturezas de informação – de pessoas, de organizações, de governos, de transações, de comportamento, de eventos – e até diferentes tipos de captação: coletores da web, sensores de imagem, som, luz, movimento, vídeo, aceleração, gravidade, temperatura, e muitos outros. Nessa abundância, reside todo o potencial de Inteligência Artificial e, principalmente, de um de seus subconjuntos de técnicas muito utilizado, chamado Aprendizado de Máquina.

Se os algoritmos são o motor, certamente dados são o combustível dessa revolução tecnológica.

Aprendizado de Máquina

Subconjunto de Inteligência Artificial, as técnicas de Aprendizado de Máquina também não são tão recentes: foram criadas na década de 80. São programas capazes de aprender a realizar uma tarefa não a partir de instruções explícitas, como na programação tradicional, mas por meio de uma experiência. Quanto maior a quantidade, a qualidade e a diversidade de dados – experiências – disponíveis, mais complexas podem ser as tarefas aprendidas e executadas por esses algoritmos.

Existem muitos tipos de sistemas de Aprendizado de Máquina e diferentes formas de categorizá-los: se há necessidade ou não de interferência humana – algoritmos supervisionados, não supervisionados, semi supervisionados, *reinforcement learning* –, se o algoritmo pode aprender ou não em tempo real – aprendizado em *batch* ou *online* –, se ele entende um padrão a partir de dados de treino e cria um modelo preditivo ou se simplesmente compara dados novos com os dados conhecidos – aprendizado baseado em instância ou em modelos – e até se é um aprendizado estatístico ou neural – como regressão linear ou *deep learning*.

Contribuindo para a popularidade de IA, há uma disponibilidade cada vez maior de ferramentas para implementação de sistemas de Aprendizado de Máquina. Ao longo do tempo, grandes provedores de tecnologia e até comunidades de código aberto disponibilizaram uma rica gama de bibliotecas e de ferramentas úteis para construção de sistemas, de forma gratuita e aberta, por exemplo TensorFlow², PyTorch³ e scikit-learn⁴; além disso, há também um amplo conjunto de plataformas comerciais disponíveis. Esse acesso facilitado às ferramentas pulverizou a criação de soluções de IA, por isso hoje, em poucas horas e com poucos recursos, pode-se criar um sistema que reconhece se uma foto contém uma pessoa ou um cachorro, por exemplo.

O que torna Aprendizado de Máquina, consequentemente Inteligência Artificial, tão interessante é a diversidade de desafios que esse conjunto de técnicas é capaz de endereçar. São técnicas bastante poderosas, que funcionam muito bem em cenários complexos, em que o contexto muda de forma muito dinâmica, a automatização requer longas, minuciosas (e, certas vezes, impraticáveis) listas de regras, ou há um volume de dados intratável com programação tradicional.

O que torna Aprendizado de Máquina, consequentemente Inteligência Artificial, tão interessante é a diversidade de desafios que esse conjunto de técnicas é capaz de endereçar.

² Para saber mais, acesse <<https://www.tensorflow.org/>>

³ Para saber mais, acesse <<https://pytorch.org/>>

⁴ Para saber mais, acesse <<http://scikit-learn.org/stable/>>

Hoje, em poucas horas e com poucos recursos, pode-se criar um sistema que reconhece se uma foto contém uma pessoa ou um cachorro, por exemplo.

Inteligência Artificial centrada em pessoas

Construir um bom sistema de Inteligência Artificial abrange muito mais que a tecnologia em si. Modelar o desafio, entender que tipo de abordagem é mais ou menos apropriada para cada situação, observar e avaliar minuciosamente a matéria-prima – os dados – e tomar a melhor decisão diante das restrições são apenas algumas considerações feitas na construção de uma solução de Inteligência Artificial.

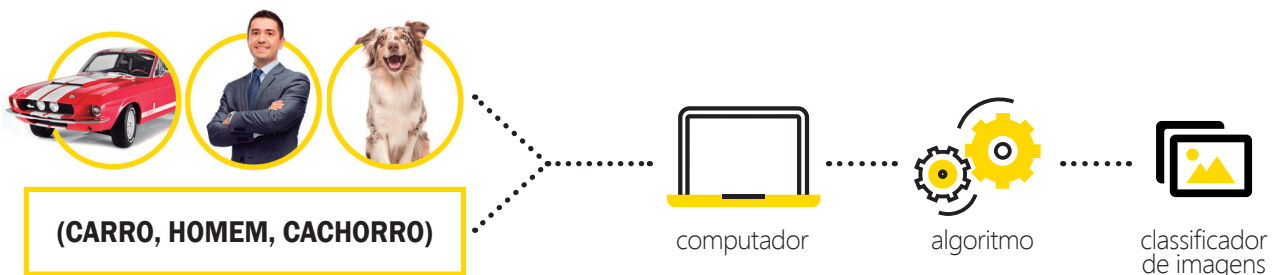
A inserção de algoritmos em processos decisórios importantes, tais como seleção de pessoas para contratação, avaliação para cessão de crédito, diagnósticos médicos, dentre outros, levanta uma série de reflexões quanto à qualidade de tais decisões. Questionamentos incluem desafios de transparência, inclusão, privacidade e até *accountability*. Quem é o culpado por uma decisão ruim de um algoritmo? O que leva um modelo a escolher uma pessoa ou outra?

O ponto fundamental desse novo paradigma de tomada de decisão automatizada é: a tecnologia em si é apenas um dos vários elementos atuantes. Fei-Fei Li (2018), professora de Ciência da Computação em Stanford, co-fundadora do AI4All e Conselheira do Google AI, escreveu recentemente um artigo apontando que, apesar do nome, não há nada “artificial” em IA, visto ser feita por humanos, com a intenção de se comportar como humanos e para afetar humanos. Então, se queremos que essa tecnologia tenha um papel positivo na sociedade, a conversa deve ser guiada também por preocupações humanas. Nesse contexto, Fei-Fei Li (2018) sugere o conceito de Inteligência Artificial centrada em pessoas.

Na criação de modelos de Aprendizado de Máquina, uma das formas mais comuns de implementação de IA, há pelo menos três pontos em que pessoas intervem diretamente: seleção e preparação de dados; desenho da solução e definição de sucesso; e intenção de uso do sistema. Para cada um desses pontos de interferência, há uma série de cuidados que precisam ser tomados para que a IA construída seja segura, justa e inclusiva.

ESQUEMA SIMPLIFICADO DE UM SISTEMA DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE IMAGENS

Dados e resultados servem como entrada; um algoritmo é criado para um determinado uso.



DESTAQUE PARA TRÊS IMPORTANTES PONTOS DE INTERVENÇÃO HUMANA NO PROCESSO



1. Seleção e preparação de dados

O primeiro ponto diz respeito à definição e tratamento do conjunto de dados utilizado como treinamento para criação do modelo.

O desafio do viés. Ao treinar um modelo de Aprendizado de Máquina, é muito importante que o conjunto de dados seja representativo e completo em relação ao objetivo de aplicação. Sistemas de Inteligência Artificial têm relação direta com os dados que recebem como entrada e pode haver grandes diferenças na qualidade do resultado, dependendo de quem projetou a solução e dos atributos considerados pelo modelo.

Um estudo recente por Joy Buolamwini e Timnit Gebru (2018) mostrou que há grande diferença na acurácia do reconhecimento facial, dependendo da raça e do gênero. Os autores destacam que sistemas de IA têm o grau de inteligência permitido pelos dados, evidenciando que uma das bases de reconhecimento facial amplamente utilizadas por diversos algoritmos contém aproximadamente 75% de imagens de homens e mais de 80% de imagens de brancos. Como o algoritmo pouco conhece exemplos de mulheres negras, conseqüentemente há uma acurácia muito menor no reconhecimento dessas pessoas.

Desigualdade social, heranças históricas e até diferenças culturais e geográficas podem também causar desafios de representatividade e provocar a reprodução de estereótipos. Algoritmos são um reflexo do comportamento humano e, até quando não expressos explicitamente, certos vieses podem ser influentes no comportamento do sistema. Um exemplo comum são associações implícitas entre nomes femininos e família, e entre nomes masculinos e carreira, como mostra estudo de Caliskan, Bryson e Narayanan (2017).

Se não observados cuidadosamente, o viés de bases de treino pode resultar em modelos de Aprendizado de Máquina preconceituosos. Quando se trata de pessoas, podem ficar evidentes vieses relacionados a gênero, raça e outros aspectos socio-demográficos, como acontece com a aplicação de predição de futuros criminosos feita nos Estados Unidos, notoriamente tendenciosa em relação a pessoas negras⁵.

A ferramenta Facets⁶ exemplifica o processo de investigação que deve ser conduzido por Cientistas de Dados para a construção de sistemas eficazes de Aprendizado de Máquina. Os criadores da ferramenta reforçam que bons profissionais devem ir muito além do treino do algoritmo: devem atuar como detetives, a fim de entender melhor seu modelo, as bases de dados que os alimentam e até acompanhar os resultados em produção.

Privacidade como prioridade. Algoritmos de Inteligência Artificial, por definição, necessitam de um grande volume de dados como entrada. A constante busca por mais customização e mais qualidade no resultado dos modelos torna as empresas preocupadas cada vez mais em coletar e armazenar toda e qualquer informação que possa ser útil e agregar mais valor aos seus serviços e plataformas. Na maioria das vezes, esse valor é derivado de um maior conhecimento por parte da empresa a respeito do consumidor, o qual muitas vezes disponibiliza seus dados pessoais em diferentes plataformas digitais.

Dados pessoais abrangem informações que indicam identidade digital, conteúdo e comportamento em redes sociais, rastros de navegação, como *cookies* ou informação geolocalizada e até aqueles derivados de análise, inferência, segmentação. Devido à sofisticação das técnicas de IA e à crescente precisão com a

Desigualdade social, heranças históricas e até diferenças culturais e geográficas podem também causar desafios de representatividade e provocar a reprodução de estereótipos. Algoritmos são um reflexo do comportamento humano e, até quando não expressos explicitamente, certos vieses podem ser influentes no comportamento do sistema.

⁵ Para saber mais, acesse <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>

⁶ Para saber mais, acesse <<https://pair-code.github.io/what-if-tool/>>

O controle do uso dos dados deve ser do usuário, não da empresa, já que qualquer solução de dados construída deve permitir consentimento informado – claro o suficiente para que o usuário realmente entenda como, quando e para que seus dados serão utilizados –, de forma granular – que permita diferentes níveis de consentimento – e dinâmico – a fim de que uma concessão de uso seja revogada pelo usuário.

qual indivíduos podem ser analisados e até identificados, diversas discussões surgiram acerca do tema. O escândalo Cambridge Analytica⁷ é um exemplo do poder desses dados.

Recentemente, entraram em vigor na Europa leis regulamentando o uso responsável dos dados de terceiros – a chamada General Data Protection Regulation (GDPR)⁸. O Brasil também está estruturando um conjunto similar de leis⁹. Conforme aponta um relatório do World Economic Forum (2017), alguns usuários estão dispostos a trocar elementos de seus dados pessoais em troca de valor agregado em serviços e plataformas: consentimento é o principal elemento a ser considerado durante um projeto de solução de dados, além de ser o ponto central das legislações vigentes na Europa e, em desenvolvimento, no Brasil.

O controle do uso dos dados deve ser do usuário, não da empresa, já que qualquer solução de dados construída deve permitir consentimento informado – claro o suficiente para que o usuário realmente entenda como, quando e para que seus dados serão utilizados –, de forma granular – que permita diferentes níveis de consentimento – e dinâmico – a fim de que uma concessão de uso seja revogada pelo usuário.

Também do ponto de vista técnico, há formas de se garantir o valor agregado de customização a partir de dados, sem infringir questões associadas à privacidade. Uma das técnicas que podem ser utilizadas é o Aprendizado Federado, que consiste em um algoritmo que aprende localmente, sem que os dados saiam do dispositivo do usuário.

2. Definição do sucesso e desenho da solução

O segundo ponto engloba a seleção da métrica de avaliação de desempenho do modelo e a definição do que é uma boa decisão do sistema.

Definindo sucesso. Outro ponto importante de influência humana na construção de modelos de Aprendizado de Máquina é a definição da métrica de sucesso, ou seja, a métrica pela qual o desempenho do modelo será avaliado. A acurácia nem sempre é o melhor indicador de desempenho de um determinado algoritmo, principalmente em bases de dados que apresentam viés. É importante escolher uma métrica que considere os fatores humanos e sociais envolvidos no processo.

Considere, por exemplo, um modelo que concede crédito e prediz a pontuação de uma pessoa, dada a probabilidade de ela pagar o empréstimo. O que representa o sucesso do sistema? Mais lucro? Maior número de pessoas recebendo empréstimo? Maior número de pagantes? O estudo por Wattenberg, Viégas e Hardt (2016) mostra que há diversas formas de tratar um mesmo problema e que exercitar diferentes métricas pode ajudar a reduzir a desigualdade no uso de Inteligência Artificial.

⁷ Para saber mais, acesse:

<<https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm>>

⁸ General Data Protection Regulation (regulamento geral sobre a proteção de dados) é um regulamento do direito europeu sobre privacidade e proteção de dados pessoais, aplicável a todos os indivíduos na União Europeia e no Espaço Económico Europeu. Regulamenta também a exportação de dados pessoais para fora da UE e EEE.

⁹ Para saber mais, acesse:

<<http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=548066>>

Ataques. Como qualquer outra tecnologia, modelos de Inteligência Artificial também estão sujeitos a ataques. Ataques adversariais são modificações pequenas em instâncias feitas, de modo a não atrapalhar a percepção de um ser humano, mas suficiente para enganar o algoritmo. Um exemplo de ataque adversarial seria realizar pequenas modificações em imagens para confundir os algoritmos de IA e forçar uma saída de interesse, como um pequeno adesivo em uma placa de "pare" que pode induzir uma leitura errada para o classificador de imagens, por exemplo uma sinalização de "limite de velocidade".

EXEMPLO DE ATAQUE ADVERSARIAL EM IMAGEM: UM PEQUENO QUADRADO AMARELO CONFUNDE O ALGORITMO DE INTERPRETAÇÃO DE PLACAS DE TRÂNSITO



Ao colocar um sistema de Aprendizado de Máquina em produção, o cientista de dados precisa considerar a sensibilidade da tomada de decisão automatizada. Deve também aplicar práticas de segurança e proteção a potenciais efeitos colaterais da tecnologia e, em casos apropriados, criar condições de teste em ambientes simulados. Após a integração do modelo, monitorar sua operação.

Confiança e transparência. Como saber se podemos ou não confiar em uma decisão automática? O algoritmo está fazendo o que esperamos? E, mesmo se o resultado for o esperado, como verificar se o sistema foi correto no processo de avaliação?

Confiança é um fator chave na interação entre pessoas e máquinas e, na maioria das vezes, a reflexão passa pelo entendimento do que acontece por trás da automação. Em algoritmos mais simples, como uma Árvore de Decisão, é fácil entender o racional de funcionamento do modelo, já que se trata de uma sequência de regras. No entanto, para algoritmos mais sofisticados, como *deep learning*, o modelo pode ser bastante complexo, logo é muito difícil entender a decisão do algoritmo.

Por meio de técnicas como LIME¹⁰ e SHAP¹¹ é possível criar explicações para classificações, mesmo em algoritmos que sejam originalmente “caixa-preta”. Isso ajuda o decisor a entender a motivação do algoritmo por trás de uma determinada classificação.

¹⁰ Do inglês Local Interpretable Model – Agnostic Explanations. Para saber mais, acesse: <<https://arxiv.org/pdf/1602.04938.pdf>>

¹¹ Do inglês Shapley Additive explanations. Para saber mais, acesse: <<https://arxiv.org/pdf/1705.07874.pdf>>

Além de ajudar na compreensão e na validação do resultado das predições de IA, aumentar a transparência dos algoritmos pode levar à descoberta de novos fatos. Após obterem sucesso na aplicação de *deep learning* para diagnóstico de retinopatia diabética¹², pesquisadores do Google¹³ decidiram abrir a "caixa-preta" do algoritmo. Aplicando técnicas de atenção, para entender o racional que o algoritmo preditor utilizou para realizar o diagnóstico, a equipe observou padrões ligados aos vasos sanguíneos e identificou que imagens do olho poderiam também prever outros indicativos de doenças cardiovasculares. É uma nova forma de realizar descobertas científicas, podendo ajudar cientistas a gerar hipóteses mais direcionadas e impulsionar uma ampla gama de pesquisas futuras.

Também faz parte da discussão as implicações da tomada de decisão moral pelas máquinas. Quando decisões críticas são tangibilizadas em algoritmos e em programação, dilemas morais rapidamente ficam em evidência. Um exemplo desse desafio são os diferentes cenários de decisão de um carro autônomo: diante de uma decisão impossível, em que uma ou mais vidas podem estar em jogo, qual a melhor decisão técnica? O experimento *Moral Machines* (Awad, 2017) é um dos estudos que têm sido conduzidos sobre o assunto, o qual tenta entender a perspectiva humana sobre decisões morais feitas por máquinas.

Uma estratégia bastante utilizada é construir a solução a fim de que o sistema de IA seja um instrumento de tomada de decisão informada, não o decisor em si. Nas aplicações de Inteligência Artificial para saúde, explicar como o algoritmo fez a previsão dá ao médico as informações necessárias e a confiança para tomar a melhor decisão. Apesar de a tecnologia ter tornado o processo mais rápido e preciso, a palavra final continuou sendo do médico especialista.

3. Intenção de uso

Por último, coloca-se em questão a finalidade de uso do sistema de Inteligência Artificial construído. Uma mesma tecnologia pode ter aplicações benéficas ou não, dependendo da intenção de quem a opera. Reconhecimento de imagem, por exemplo, é uma técnica que pode ajudar uma pessoa com deficiência visual a entender o conteúdo de imagens, um fazendeiro a diagnosticar a qualidade de uma plantação, e médicos a realizarem diagnósticos complexos. Em contrapartida, o reconhecimento de imagem foi também a tecnologia utilizada em experimentos para inferir a orientação sexual de uma pessoa a partir dos traços do rosto¹⁴.

Garantir que as soluções criadas pelas pessoas utilizando Inteligência Artificial serão éticas, benéficas e seguras para a sociedade é um desafio real, que tange a criação de códigos de ética; o equilíbrio de forças, como mercado, infraestrutura tecnológica, normas e leis; e o envolvimento de múltiplas partes interessadas.

¹² Para saber mais, acesse <<https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html>>

¹³ Para saber mais, acesse <<https://ai.googleblog.com/2018/02/assessing-cardiovascular-risk-factors.html>>

¹⁴ Para saber mais, acesse <<https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html>>

O futuro que queremos com Inteligência Artificial

À medida que oportunidades para Inteligência Artificial se multiplicam, maior é a necessidade de engajar líderes da indústria, pesquisadores, especialistas, governantes e sociedade civil para reflexões e ações em prol de soluções que beneficiem a sociedade como um todo. Princípios éticos e padrões técnicos ajudarão a garantir que o desenho e o desenvolvimento de tais tecnologias sejam orientados pela preocupação com seu efeito em pessoas. Mas há também outras perspectivas e iniciativas necessárias.

Diante do impacto das automações no mercado de trabalho e também para garantir uma participação ativa da sociedade civil no avanço dessas tecnologias, será importante um trabalho igualmente intenso em educação, particularmente com relação a STEM (Ciência, Tecnologia, Engenharia e Matemática), e em treinamento de habilidades intrinsecamente humanas, como criatividade, resiliência, flexibilidade, resolução de problemas complexos, dentre outras; portanto, governos, empresas e universidades terão um papel fundamental nesse processo.

Inteligência Artificial veio para ficar e para transformar o mercado, a sociedade e a vida das pessoas. Tão diversas quanto o público que essa tecnologia vai impactar devem ser as pessoas envolvidas em sua construção, com diferentes vivências, trajetórias e formações. O nosso futuro com IA deve ser o resultado da conexão entre tecnologia e ciências humanas, tecnologia e especialidades, tecnologia e sociedade.

Tão diversas quanto o público que essa tecnologia vai impactar devem ser as pessoas envolvidas em sua construção, com diferentes vivências, trajetórias e formações. O nosso futuro com IA deve ser o resultado da conexão entre tecnologia e ciências humanas, tecnologia e especialidades, tecnologia e sociedade.

REFERÊNCIAS

- Awad, Edmon. (2017). *Moral Machine: Perception of moral judgment made by machines*. Recuperado de: <https://www.media.mit.edu/publications/moral-machine-perception-of-moral-judgment-made-by-machines/>
- Buolamwini, J. & Gebru, T. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1-15. Recuperado de: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- Caliskan, A., Bryson, J. J. & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186. Recuperado de: <http://science.sciencemag.org/content/356/6334/183>
- Li, Fei-Fei. (2018). *How to make A. I. that's good for people*. Recuperado de: <https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html>
- Wattenberg, M., Viégas, F. & Hardt, M. (2016). *Attacking discrimination with smarter machine learning*. Recuperado de: <https://research.google.com/bigpicture/attacking-discrimination-in-ml/>
- World Economic Forum. (2017). *Shaping the Future Implications of Digital Media for Society Valuing Personal Data and Rebuilding Trust*. Recuperado de: http://www3.weforum.org/docs/WEF_End_User_Perspective_on_Digital_Media_Survey_Summary_2017.pdf

Entrevista



Sara Rendtorff-Smith

é pesquisadora no Instituto de Tecnologia de Massachusetts (MIT).

Desafios de Governança em Inteligência Artificial

Sara Rendtorff-Smith é Pesquisadora Líder de Governança orientada por Dados e Políticas para IA no Instituto de Tecnologia de Massachusetts (MIT). Nesta entrevista, comenta sobre as oportunidades de Inteligência Artificial, bem como alguns dos problemas associados ao seu uso e a forma de lidar com eles.

P.S._ Como é possível lidar com erros, vieses e resultados não esperados relacionados a Inteligência Artificial?

S.R.S._ Antes de mais nada, é importante considerar dois aspectos: em primeiro lugar, como problemas como erros, vieses e resultados não esperados ocorrem; e em segundo lugar, como podemos auditá-los e investigar de que maneira os erros foram cometidos após o fato. É importante notar que tais problemas não esperados estão localizados diretamente na interseção da interação homem-máquina. Dessa forma, erros e vieses são geralmente introduzidos pelos seres humanos, à medida que selecionam dados e fazem escolhas relacionadas a desenho e custo.

Um exemplo útil para explorar esses conceitos diz respeito aos veículos autônomos. Os engenheiros programarão o veículo para fazer o *trade off* entre o risco para motorista e o risco para qualquer pessoa com a qual o veículo possa colidir. A tecnologia de reconhecimento visual também pode ser inadvertidamente treinada de forma que tenha um desempenho melhor com certos dados demográficos do que com outros. Essa tem sido uma consequência não intencional em que o setor de tecnologia usa dados demográficos viesados, por exemplo em que mulheres e pessoas negras estão sub-representadas. Isso, por sua vez, leva a uma baixa acurácia de reconhecimento de imagem para essas populações. Essas questões são frequentemente levantadas no contexto da responsabilidade legal, mas também são importantes para desenhar *frameworks*, a fim de prevenir e mitigar erros, vieses e resultados não esperados.

No caso de erros verdadeiros e consequências não esperadas, por outro lado, torna-se importante que os sistemas de IA sejam auditáveis, de maneira que permitam aos pesquisadores identificar a origem de um erro ou consequência adversa. Isso faz parte de uma discussão muito mais ampla sobre sistemas de inteligência artificial “caixa-preta” e seu lugar na sociedade.

P.S._ Qual é o papel das diferentes partes interessadas na construção de soluções éticas de IA? Como os usuários finais devem se envolver nessa discussão?

S.R.S._ IA está no centro das relações sociotécnicas, portanto requer o envolvimento ativo e informado das várias partes interessadas para a construção de soluções éticas de IA. Acadêmicos e desenvolvedores devem ser treinados para estarem cientes de sua própria responsabilidade relacionada ao desenvolvimento de sistemas de IA com impacto direto na sociedade. Os governos e os cidadãos, por sua vez, têm uma responsabilidade crítica na definição de quais valores devem governar a implantação de sistemas de IA na esfera pública e como as questões de responsabilidade devem ser reguladas. Os usuários finais ou aqueles indivíduos e comunidades diretamente impactados por uma tecnologia específica de IA podem ser envolvidos em três estágios críticos do desenvolvimento de IA: participando do desenho, do teste e da auditoria de novos sistemas de IA.

P.S._ Como IA pode ser usada para a formulação de políticas baseadas em evidências? Quais são os desafios para isso?

S.R.S._ Uma das áreas de políticas mais proeminentes em que IA tem sido aproveitada para promover a tomada de decisão baseada em evidências é o setor de justiça criminal. Nos EUA, mais de 60 departamentos de polícia atualmente implementam algum tipo de sistema de policiamento preditivo baseado em IA. Enquanto isso, um número crescente de jurisdições tem adotado escores de risco gerados por sistemas de IA para ajudar em vários pontos de decisão em toda a cadeia da justiça criminal, incluindo liberação pré-julgamento, sentença pós-condenação e liberdade condicional.

Exemplos importantes dos desafios associados a tal uso incluem o viés na avaliação da previsão de comportamento futuro de réus e pessoas encarceradas com base na raça, assim como o excesso de policiamento dos bairros de minorias como resultado de práticas de policiamento historicamente tendenciosas. O irônico é que a estratégia foi adotada principalmente para mitigar o viés humano na tomada de decisão judicial. Esses e muitos outros desafios tendem a emergir em parte como resultado da natureza “caixa-preta” do tipo particular de tecnologia de IA implementada, predominantemente em sistemas de aprendizado de máquina e de *deep learning* (aprendizado profundo).

Esses sistemas, particularmente quando falamos de sistemas de aprendizado não supervisionado, são, por natureza, obscuros, o que dificulta a manutenção de princípios fundamentais de transparência, explicabilidade e *accountability*. Outro desafio tem sido que os governos geralmente adquirem e implantam tecnologia proprietária de IA desenvolvida por atores da indústria privada.

O laboratório do MIT em que trabalho está atualmente conduzindo pesquisas sobre como outra forma de inteligência artificial, baseada em computação probabilística, pode ajudar os formuladores de políticas a identificar

padrões em dados, incluindo dependências entre variáveis, semelhanças sobre como a sociedade encara os temas éticos dentro e fora do tema de IA. No Brasil, ética ainda não é um tema muito discutido nas escolas e nas universidades: o que mais se observa são discussões civis, contratuais. Não se discute, por exemplo, se personalidade de um robô é uma boa ideia ou não. Dada a novidade do tema, existem poucas pessoas que estão se debruçando academicamente sobre IA, com um olhar crítico, o que dificulta ainda mais a produção nesse tema. Como o Brasil é muito incipiente nesses debates sobre ética e IA, ainda não temos muita produção nesse sentido. Vivemos num contexto de fetichização da tecnologia. Incorporamos novas tecnologias em nosso cotidiano sem termos uma visão crítica sobre elas, enxergando somente do ponto de vista da utilidade, sem pensar em questões de privacidade, segurança e até impactos éticos dessas novas tecnologias.

P.S._ Há muita discussão sobre a accountability de IA. Como isso deve acontecer? Quais partes interessadas devem ser envolvidas nesse processo?

S.R.S._ Esta é realmente uma discussão muito ativa e oportuna acontecendo no momento. Se estamos falando de IA sendo implantada de uma maneira que afeta o público em geral ou ainda de uma instância onde o governo está aproveitando a tecnologia IA como parte de seu sistema de governança, o limiar de *accountability* é naturalmente maior do que quando uma empresa privada implementa IA para suas operações internas.

Outras abordagens à *accountability* de IA incluem abordagens *upstream*, como ensinar desenvolvedores e acadêmicos sobre suas responsabilidades éticas e treiná-los para abordá-las já no estágio de desenho ou nas abordagens intermediárias, tais como um código de conduta para a inovação ética ou a supervisão pública de testes de produtos pré-implantação.

Entrevista II

Ética, Direito e Inteligência Artificial

Eduardo Magrani, advogado, coordenador do Instituto de Tecnologia e Sociedade do Rio (ITS Rio), comenta sobre os desafios que o desenvolvimento de Inteligência Artificial traz em matéria de ética, legislação e futuro do trabalho.

P.S._ Quais são os principais desafios em termos éticos e legais que devem ser considerados no uso de IA?

E.M._ Há um *gap* cultural sobre como a sociedade encara os temas éticos dentro e fora do tema de IA. No Brasil, ética ainda não é um tema muito discutido nas escolas e nas universidades: o que mais se observa são discussões civis, contratuais. Não se discute, por exemplo, se personalidade de um robô é uma boa ideia ou não. Dada a novidade do tema, existem poucas pessoas que estão se debruçando academicamente sobre IA, com um olhar crítico, o que

dificulta ainda mais a produção nesse tema. Como o Brasil é muito incipiente nesses debates sobre ética e IA, ainda não temos muita produção nesse sentido. Vivemos num contexto de fetichização da tecnologia. Incorporamos novas tecnologias em nosso cotidiano sem termos uma visão crítica sobre elas, enxergando somente do ponto de vista da utilidade, sem pensar em questões de privacidade, segurança e até impactos éticos dessas novas tecnologias.

P.S._ *Fora do Brasil, quais são as questões mais preocupantes sobre a intersecção de ética e IA, em termos de legislação?*

E.M._ Existe uma discussão forte sobre como os programadores e os engenheiros de *software* dão *inputs* éticos e morais nas máquinas. A tecnologia não é neutra, os algoritmos não são neutros, e muitas vezes têm uma moralidade intrínseca àquele artefato técnico, inserida por um programador. Então, quais são os valores que os engenheiros de *software* vão colocar nesses artefatos técnicos, cada vez mais autônomos e mais imprevisíveis também? Uma outra discussão remete à reparação de danos: se um robô autônomo produz um dano que não foi imaginado pela empresa, pelo programador, quem deveria responder por isso?

P.S._ *É possível conceber uma legislação que resguarde os direitos humanos sem limitar o desenvolvimento dessas tecnologias?*

E.M._ Quando você diz que tem que ser uma “legislação orientada pelos direitos humanos” não é algo tão óbvio quanto parece e isso é o que eu defendo. Mas tem também quem defenda o avanço tecnológico em prol da eficiência e não necessariamente da proteção dos direitos humanos. Vemos muito isso ainda hoje no Brasil: grande parte do avanço tecnológico que a gente vive hoje não é orientado pela proteção da privacidade, da segurança, da dignidade humana, conforme aponta nossa Constituição. Esse avanço é orientado meramente pela eficiência e utilidade. Mas, eu defendo que o avanço da IA deve ser guiado por uma visão humano-cêntrica, pois a gente tem de sair de uma visão do direito iluminista, em que só se enxerga humanos como dotados de poderes de ação. O direito deve avançar para enxergar também agentes não humanos que interagem conosco e geram efeitos democráticos, constitucionais, civis; enxergando esses diferentes agentes, humanos e não humanos, vamos poder desenhar regulamentações melhores para esse cenário de hiperconectividade ou de IA.

P.S._ *Quais são os desafios atuais em termos de legislação e AI, considerando nosso contexto nacional e o cenário internacional?*

E.M._ O problema mais óbvio é de jurisdição; o GDPR na União Europeia é uma lei muito forte. Só pode haver fluxo internacional de dados e informações entre países com a mesma adequação de normas de privacidade, o que não é o caso do Brasil. Então como controlamos o fluxo de dados entre países que têm diferentes regulações, por exemplo, de privacidade? Esse é um grande desafio hoje, que tem a ver com o cenário de Internet das Coisas, no qual há cada vez mais coisas inteligentes gerando dados o tempo todo, os quais são transferidos entre as empresas e até internacionalmente.

Outro problema é [a ausência de uma] alfabetização digital forte, uma capacitação em temas digitais. Não temos isso nas escolas, não costumamos ter isso nas universidades. Então o cidadão desconhece completamente mui-



Eduardo Magrani
é coordenador
no Instituto de
Tecnologia e
Sociedade do Rio
(ITS Rio).

tos riscos e vulnerabilidades, conseqüentemente ele não gera pressão política para ter legislação melhor ou para observar o que está acontecendo em outros países e tentar incorporar no Brasil. Aqui, o tema de direito digital ainda é muito pouco discutido, é muito incipiente, então essa ausência de regulações adequadas transparece um cenário de abusividade muito grande.

P.S._ Há boas práticas ou diretrizes a seguir nos temas de ética e IA?

E.M._ Hoje, há várias iniciativas que tentam fazer um mapeamento de boas práticas, como o Institute of Electrical and Electronics Engineers (IEEE)¹⁵, que publicou um relatório sobre princípios que deveriam orientar práticas de IA¹⁶. Há também a a Declaração de Montreal para uma IA Responsável¹⁷, de 2017, e na Europa estão publicando uma série de *guidelines* sobre isso, mas ainda não no Brasil.

É possível aplicar diretrizes internacionais em diferentes contextos nacionais se for seguida uma orientação dos direitos humanos. E como essas declarações atuam num nível bem macro, é possível incorporá-las em diferentes ordenamentos nacionais.

P.S._ Sobre o futuro do emprego, há uma preocupação sobre as “máquinas roubando os empregos” pela automatização, pelo avanço da tecnologia. Qual sua opinião em relação a isso, como a sociedade pode se preparar para se adaptar a essas transformações, sem deixar ninguém para trás?

E.M._ Vivemos a era da robótica. Em trabalhos mecânicos muito repetitivos, não há dúvida de que os humanos serão substituídos. O que o ser humano deve fazer? Adaptar-se e buscar incorporar novas habilidades que os robôs não conseguem fazer, para se manter no mercado de trabalho. O problema é que no cenário brasileiro muitas vezes faltam recursos para dar essa capacitação para as pessoas. No Brasil, podemos ter um *gap* geracional grande de pessoas que não tiveram tempo ou recursos para se adaptar ao novo cenário tecnológico, ao passo que outros países vão fazer isso mais rapidamente para evitá-lo.

Não tenho dúvidas de que a IA, essa nova era da automação, vai gerar novas demandas também. Já é possível observar a quantidade de áreas que trabalham com filtragem de algoritmos, análises de *big data* para incrementar seus negócios; isso gera novas demandas, inclusive dos clientes. O problema são aqueles que não vão ter tempo ou recursos para se adaptar. A capacitação em direito digital e em novas tecnologias são os principais caminhos possíveis para amenizar esse *gap*.

¹⁵ O IEEE, associação dedicada ao avanço da inovação e excelência tecnológica em benefício da humanidade, é a maior sociedade profissional técnica do mundo. Ela foi projetada para atender profissionais envolvidos em todos os aspectos dos campos elétricos, eletrônicos e de computação e áreas afins da ciência e da tecnologia subjacentes à civilização moderna.

¹⁶ O sumário executivo, em português, pode ser acessado em <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_executive_summary_portuguese_v1.pdf>

¹⁷ Disponível em inglês em <<https://www.montrealdeclaration-responsibleai.com/the-declaration>>

Relatório de Domínios

A dinâmica dos registros de domínios no Brasil e no mundo

O Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) monitora mensalmente a quantidade de nomes de domínios registrados entre os 16 maiores domínios de topo de código de país (do inglês, *country code Top-level Domain* – ccTLD) no mundo. Somados, eles ultrapassam 97,8 milhões de nomes de domínios registrados. Em outubro de 2018, os domínios registrados sob o .tk (Tokelau) chegaram a 21,19 milhões. Em seguida, aparecem Alemanha (.de), China (.cn) e Reino Unido (.uk) com, respectivamente, 16,21 milhões, 11,11 milhões e 9,83 milhões de registros¹⁸. O Brasil atingiu a marca de 4 milhões de registros sob o .br e continua ocupando a sétima posição. Na 16ª posição, com 1,93 milhão de registros, está a Espanha (.es), como observado na Tabela 1.

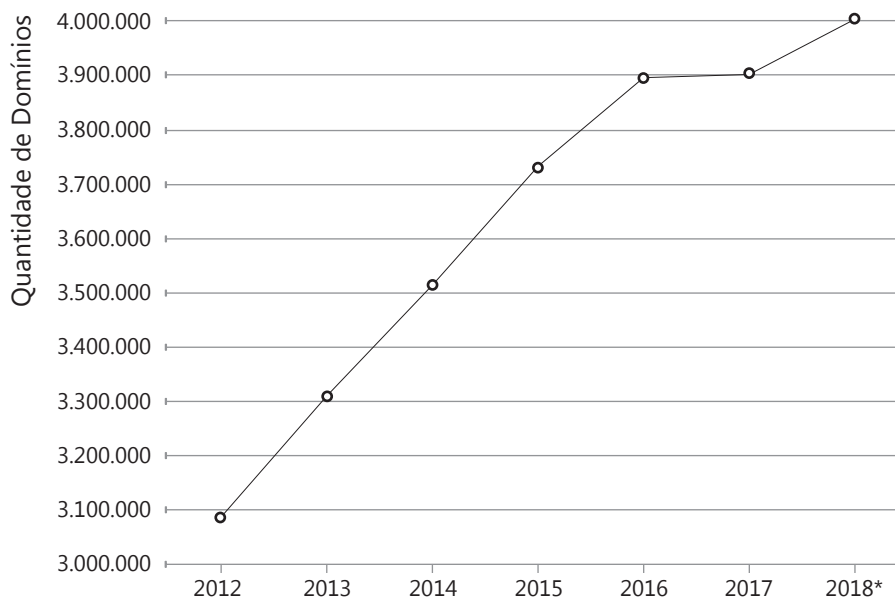
Tabela 1 – REGISTRO DE NOMES DE DOMÍNIOS NO MUNDO – OUTUBRO/2018

Posição	ccTLD	Domínios	Ref.	Fonte
1	Tokelau (.tk)	21.196.458	Out/18	research.domaintools.com/statistics/tld-counts/
2	Alemanha (.de)	16.211.149	Out/18	www.denic.de/
3	China (.cn)	11.113.061	Out/18	research.domaintools.com/statistics/tld-counts/
4	Reino Unido (.uk)	9.832.194	Ago/18	www.nominet.uk/uk-register-statistics-2018/
5	Países Baixos (.nl)	5.815.231	Out/18	www.sidn.nl
6	Rússia (.ru)	5.059.321	Out/18	www.cctld.ru
7	Brasil (.br)	4.002.574	Out/18	registro.br/estatisticas.html
8	União Europeia (.eu)	3.661.862	Out/18	research.domaintools.com/statistics/tld-counts/
9	França (.fr)	3.252.404	Out/18	www.afnic.fr/en/resources/statistics/detailed-data-on-domain-names/
10	Austrália (.au)	3.162.708	Out/18	www.auda.org.au
11	Itália (.it)	3.145.839	Out/18	www.nic.it/
12	Canadá (.ca)	2.779.778	Out/18	www.cira.ca/
13	Polónia (.pl)	2.585.723	Out/18	www.dns.pl/english/zonestats.html
14	Suíça (.ch)	2.164.079	Jun/18	www.nic.ch/reg/cm/wcm-page/statistics/index.html?lid=em*
15	Estados Unidos (.us)	1.950.307	Out/18	research.domaintools.com/statistics/tld-counts/
16	Espanha (.es)	1.931.707	Out/18	www.dominios.es

¹⁸ É importante destacar que o período de referência de cada ccTLD não é o mesmo em todos os casos, embora seja o mais atualizado.

O Gráfico 1 apresenta o desempenho do .br desde o ano de 2012.

Gráfico 1 – TOTAL DE REGISTROS DE DOMÍNIOS AO ANO DO .BR (2012-2018)



Série histórica do número de registros sob o ccTLD .br

*Dado referente ao mês de outubro de 2018.

Fonte: Registro.br

No mês de outubro, os cinco principais domínios genéricos (do inglês, *generic Top-Level Domain* – gTLD) totalizam mais de 169 milhões de registros. O .com destaca-se, com 137,48 milhões de registros, conforme apontado na Tabela 2.

Tabela 2 – PRINCIPAIS GTLDS – OUTUBRO/2018

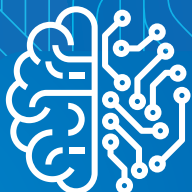
Posição	gTLD	Domínios	Fonte
1	.com	137.486.743	research.domaintools.com/statistics/tld-counts/
2	.net	13.985.155	research.domaintools.com/statistics/tld-counts/
3	.org	10.385.742	research.domaintools.com/statistics/tld-counts/
4	.info	5.440.794	research.domaintools.com/statistics/tld-counts/
5	.biz	2.252.993	research.domaintools.com/statistics/tld-counts/

Fonte: DomainTools.com. Recuperado em 9 de outubro, 2018, de <http://research.domaintools.com/statistics/tld-counts/>

Tire suas dúvidas

Muita novidade?

Preparamos uma lista com os principais termos e conceitos associados a Inteligência Artificial e Ética, confira!



Inteligência Artificial

Campo de estudo que visa a construção de sistemas capazes de exibir comportamentos normalmente associados às pessoas, como aprendizado e resolução de problemas. Algumas linhas de estudo de Inteligência Artificial são mais focadas em reproduzir a maneira como pensamos e raciocinamos, enquanto outras se concentram no entendimento e na simulação de comportamento.



Aprendizado de Máquina

Conjunto de técnicas contidas em Inteligência Artificial, Aprendizado de Máquina é a ciência e a arte de utilizar computadores de forma que eles aprendam a realizar tarefas a partir de experiências (dados). É um campo que proporciona a computadores a habilidade de aprender sem necessidade de programação explícita.



Accountability

Sistemas de Inteligência Artificial estão sujeitos a resultados inesperados ou falhas que podem causar danos. Em tais casos, é essencial que haja capacidade forense para garantir que acidentes ou falhas semelhantes não ocorram novamente e para determinar responsabilidade e prestação de contas. Desafios de *accountability* incluem: entender quem é responsável pelo comportamento de um sistema automático, ser capaz de supervisionar ou auditar decisões automáticas e até determinar responsabilidade e implicações jurídicas e legais.



Transparência

Termo relacionado a quão interpretável ou passível de explicação é um sistema de IA. Algumas técnicas utilizadas são tão complexas que funcionam como "caixa-preta" e deixam o processo decisório automático opaco até mesmo para quem construiu o sistema. Técnicas e padrões de projeto visando a transparência ajudam a determinar o quanto um sistema de IA é confiável, a entender se ele funcionará bem e até a explicar seu racional de funcionamento. O objetivo é ter decisões de modelos não somente boas, mas também interpretáveis.



Bias (ou viés) em decisões automáticas

Algoritmos são um reflexo do comportamento humano; até quando não expressos explicitamente, certos vieses presentes nos dados e nas decisões de projeto podem ser influentes no comportamento do sistema. Por isso, em alguns casos, decisões baseadas em algoritmos podem reproduzir ou reforçar um viés negativo, com padrões de discriminação e manutenção de estereótipos, como herança de decisões preconceituosas de pessoas ou simplesmente pelos dados refletirem aspectos culturais, históricos e socio-demográficos existentes na sociedade.



Lei Brasileira de Proteção de Dados - Lei No 13.709

Lei que regulamenta o uso, a proteção e a transferência de dados pessoais no Brasil. Garante maior controle dos cidadãos sobre suas informações pessoais: exige consentimento explícito para coleta e uso dos dados – pelo poder público e setor privado –, e obriga a oferta de opções para o usuário visualizar, corrigir e excluir esses dados. Também proíbe o tratamento dos dados pessoais para a prática de discriminação ilícita ou abusiva (cruzamento de informações de uma pessoa ou grupo para subsidiar decisões comerciais, políticas públicas ou atuação de órgão público).

GDPR - General Data Protection Regulation



Regulamento do direito europeu sobre a proteção de dados pessoais, cujo objetivo é oferecer ao usuário maior controle e transparência sobre as informações pessoais armazenadas em bancos de dados das empresas. O indivíduo deve ter direito de saber quais informações ele fornece aos serviços de que usufrui. Além disso, a entidade deve explicar por que requisita determinados dados do cliente, e para qual finalidade elas serão usadas.



Fairness

Conceito que engloba técnicas e práticas para garantir que sistemas de IA possam tomar decisões justas e com menos vieses indesejados. Decisões automáticas relacionadas a crédito, hipoteca, seguros, educação, mercado de trabalho, e outros podem conter vieses e reproduzir um comportamento injusto ou discriminatório de pessoas com base em atributos individuais não relacionados a mérito. O objetivo de *fairness* é manter a alta precisão dos algoritmos de aprendizado e, ao mesmo tempo, reduzir o grau em que eles discriminam indivíduos pertencentes a determinado grupo.

/Créditos

ARTIGO PRINCIPAL

Carolina Bigonha
(Hekima/Em Perspectiva)

PREPARAÇÃO E EDIÇÃO

Alexandre Barbosa
(Cetic.br)
José Márcio Martins Júnior
(Cetic.br)
Stefania L. Cantoni
(Cetic.br)
Tatiana Jereissati
(Cetic.br)

TRADUÇÃO DA ENTREVISTA I

Stefania L. Cantoni
(Cetic.br)

REVISÃO EM PORTUGUÊS

Érica Santos Soares de Freitas

AGRADECIMENTOS

Carolina Bigonha
(Hekima/Em Perspectiva)

Eduardo Magrani
(ITS Rio)

Sara Rendtorff-Smith
(Instituto de Tecnologia de
Massachusetts - MIT)

PROJETO GRÁFICO E DIAGRAMAÇÃO

Comunicação NIC.br



Organização
das Nações Unidas
para a Educação,
a Ciência e a Cultura

cetic.br

Centro Regional de Estudos
para o Desenvolvimento da
Sociedade da Informação
sob os auspícios da UNESCO

nic.br

Núcleo de Informação
e Coordenação do
Ponto BR

egi.br

Comitê Gestor da
Internet no Brasil

CREATIVE COMMONS
Atribuição
Uso Não Comercial
Não a Obras Derivadas
(by-nc-nd)





POR UMA INTERNET CADA VEZ MELHOR NO BRASIL

CGI.BR, MODELO DE GOVERNANÇA MULTISSETORIAL

www.cgi.br

nic.br cgi.br