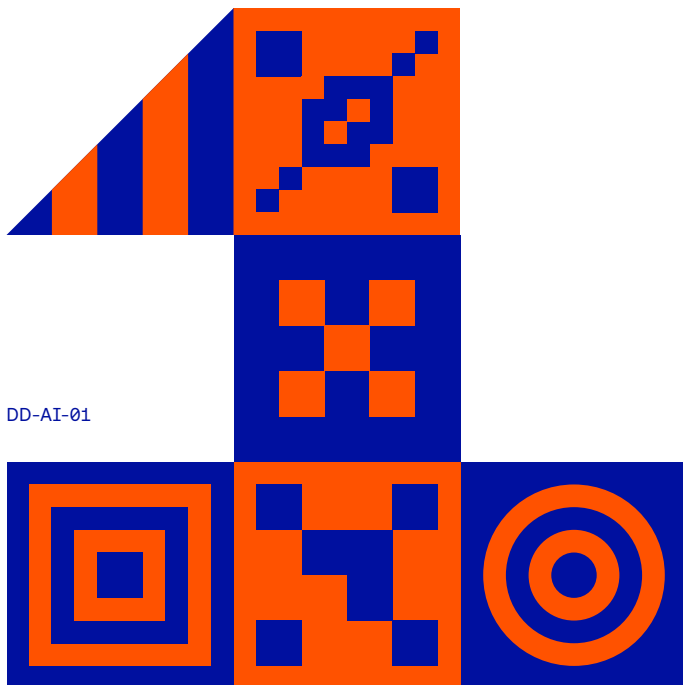
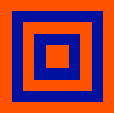


Alguns conceitos básicos





Inteligência Artificial (IA)

Refere-se ao sistema para a análise e tomada de decisões através de operações informatizadas e automatizadas baseadas em um conjunto de dados. O termo Inteligência Artificial é usado para diferentes tecnologias com a característica comum de simular a inteligência humana, aplicada a problemas para oferecer soluções baseadas na previsão ou no reconhecimento de padrões.

A IA é considerada "fraca" ou "limitada" ao executar uma tarefa específica, e "forte" ou "geral" se consegue contextualizar diferentes problemas específicos e executar tarefas independentes a partir dessa contextualização, como uma inteligência humana.

Alguns sistemas de IA, com diferentes níveis de complexidade e autonomia, estão presentes em aplicativos como os assistentes virtuais e os sistemas de reconhecimento de voz, ou em dispositivos especializados como os drones e os veículos autônomos.

Algoritmo

É uma fórmula ou conjunto de regras para articular instruções, procedimentos e processos para resolver um problema ou executar uma tarefa. Na IA os algoritmos são incorporados nos códigos de programação de um sistema informático para operar de forma automatizada, e assim orientar as máquinas sobre como procurar respostas a uma pergunta ou soluções para um problema.



Os problemas associados à tomada de decisões são geralmente atribuídos a algoritmos ou fórmulas mal projetadas, que consideram ou ponderam dados com resultados equivocados, discriminatórios e prejudiciais.

Dados

São unidades de informação, representações alfanuméricas de fatos singulares e de estatísticas que podem ser analisadas por diversos meios. Eles constituem o ingrediente essencial para a aprendizagem de máquina e a aplicação da IA. Por esse motivo, costuma-se pensar que, com mais dados, os sistemas automatizados podem ser melhor treinados para tomar decisões mais precisas.

Dados Pessoais

Consiste em uma unidade de informação que, por si só ou em relação a outras informações, permite identificar uma pessoa física. São dados pessoais: o nome; o número nacional de identidade; o telefone; a idade; o endereço residencial e/ou do trabalho; entre outros. A anonimização dos dados pessoais é, muitas vezes, um processo necessário para o seu tratamento adequado.

Dados Abertos - Open Data

Representa a prática de manter um conjunto ou base de dados disponível para que qualquer pessoa possa utilizá-la, reutilizá-la e compartilhá-la livremente, sem restrição de confidencialidade,



direitos autorais e patentes. Os dados são considerados utilizáveis quando estão em um formato comum que pode ser lido e processado por máquinas.

Conjunto ou Base de Dados - Dataset/Database

São grupos de elementos que guardam uma relação entre si e que armazenam dados como números, datas e palavras que podem ser processados para produzir informações.

A qualidade do banco de dados será diretamente proporcional às capacidades disponíveis para construí-lo, ou seja, para coletar os dados de forma inclusiva e rigorosa. Assim, qualquer incompletude (por exemplo, em relação a pessoas que moram em áreas afastadas), imprecisão, desatualização ou parcialidade na coleta serão reproduzidas no conjunto de dados.

Macrodados - Big Data

Refere-se à gestão de grandes conjuntos de dados ou dados complexos para métodos tradicionais de processamento e armazenamento e, portanto, que exigem uma abordagem diferente para o seu gerenciamento. Dados com muitos casos (linhas) oferecem maior poder estatístico, enquanto dados com maior complexidade (mais atributos ou colunas) podem levar a uma maior taxa de erros se não forem tratados com as ferramentas certas. É relevante considerar dimensões como: o volume dos dados, sua velocidade de geração, a variedade de sua origem, a importância da veracidade dos dados e a avaliação de seu valor na abordagem de possíveis riscos. Este é o modelo conhecido como os 5 V's do Big Data..



Mineração de Dados - Data Mining

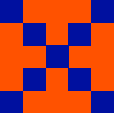
Nome dado ao processo de coleta ou extração de dados de diferentes fontes, normalmente por sistemas automatizados, com o propósito de construir bases de dados ou de identificar padrões nas informações. Embora seja uma forma de processar informações, pode também servir para fazer inferências e identificar pessoas contra sua vontade.

Ciência de Dados - Data Science

Refere-se ao campo científico que integra disciplinas como matemática, estatística, estudo das probabilidades, computação e visualização de dados para extrair conhecimento de um conjunto heterogêneo de dados (imagens, sons, textos, medições físicas, dados genômicos, links de redes sociais, entre outros). Os métodos e ferramentas derivados da aplicação da IA fazem parte deste campo.

Código Aberto

É um modelo de desenvolvimento que consiste em garantir que as pessoas acessem o código-fonte de um programa ou aplicativo gratuitamente. Assim, sua auditabilidade, integração a outros sistemas, reutilização e adaptação em contextos diferentes do original são facilitadas. Está em oposição ao código fechado que, por justificativas de propriedade intelectual ou segredos industriais, é mantido sob reserva ou sem possibilidade de modificação por pessoas não autorizadas. Dessa forma, sua auditabilidade e verificação são limitadas.



Computação em Nuvem - Cloud Computing

É um conjunto de tecnologias que permitem o acesso remoto através da internet a recursos informáticos tais como programas, aplicativos, armazenamento de arquivos e processamento de dados. Deste modo, os recursos estão guardados em grandes centros de processamento de dados. Para acessá-los não são necessários servidores locais nem a instalação de programas nos computadores finais, permitindo a prestação de serviços à distância por meio de "software como serviço - (SaaS)". É utilizada também no desenvolvimento da Internet das Coisas (IoT), sistemas de aprendizagem de máquina e Big Data.

Interface de Programação de Aplicativos - Application Programming Interface (API)

A partir de um conjunto de definições operacionais e protocolos de comunicação, este mecanismo permite programar componentes adicionais em um sistema informático, ou fazê-lo interagir com outros programas.

Bot / Chatbot

O Bot é um programa de computador com vários níveis de complexidade capaz de interagir com as pessoas através de uma conversa via áudio ou texto, normalmente utilizado para facilitar um "diálogo" entre uma pessoa e uma instituição ou sistema.



Aprendizagem de Máquina - Machine Learning

Esta é a área da IA dedicada ao desenvolvimento de técnicas e algoritmos que permitem que um sistema melhore o seu desempenho para resolver perguntas ou problemas propostos. Esta melhora do desempenho é feita a partir de inferências e de forma automatizada mediante o uso constante e a acumulação de informações.

Aprendizagem não supervisionada

É uma forma de aprendizagem de máquina onde o sistema não recebe informações sobre como os dados resultantes devem ser, eles são apenas integrados em um conjunto de dados não estruturados. O algoritmo deverá identificar possíveis padrões de dados, as relações entre eles, e propor como resultado uma estrutura, sem exigir validação humana.

Aprendizagem supervisionada

Consiste em outra forma de aprendizagem de máquina onde as pessoas fornecem os dados de entrada, os de saída e como eles devem ser usados. Isso é feito para que a máquina entenda como relacioná-los e, assim, proponha melhorias que serão validadas por outra pessoa.

Aprendizagem Profunda - Deep Learning

É um ramo da aprendizagem de máquina que visa reproduzir o funcionamento do sistema nervoso humano. Consiste na análise de grandes conjuntos de dados não estruturados para encontrar padrões, sem a necessidade de treinamento prévio. Para isso, ele



trabalha com camadas de processamento que se conectam, por exemplo, como redes neurais: o resultado da primeira camada alimenta a segunda e assim por diante, produzindo resultados cada vez mais abstratos.

Esses algoritmos têm inúmeras aplicações, como o reconhecimento automático de sons, o desenvolvimento de sistemas de visão computacional, o processamento da linguagem natural, os diagnósticos médicos, os veículos autodirigidos, entre outros. É importante considerar todos os riscos das áreas em que esses algoritmos são implementados.

Análise Preditiva

Este método tem como base a observação e estudo dos dados disponíveis até o momento para prever a probabilidade da ocorrência de eventos no futuro. Ele foi implementado em políticas públicas que buscam prevenir a violação de direitos e destinar benefícios sociais. No entanto, seus resultados são discutíveis pelas possíveis parcialidades nos dados disponíveis e pela opacidade no processamento desses dados, portanto, devem estar sujeitos a ampla verificação.

Modelo

É uma representação abstrata do que um sistema de aprendizagem de máquina aprendeu no treinamento a partir de um conjunto de dados. Por exemplo, um modelo de análise de dados climáticos pode ser programado para estabelecer probabilidades que informem a previsão diária.



Árvore de Decisão - Decision Tree

Refere-se a um tipo de modelo preditivo para realizar uma análise decisória onde, através de uma aprendizagem supervisionada, os dados são divididos de maneira iterativa a partir de parâmetros ou critérios predefinidos. O objetivo é ensinar as máquinas a tomar decisões e resolver problemas de categorização e regressão para obter um modelo definitivo.

Regressão Logística - Logistics Regression

É um processo estatístico implementado na aprendizagem de máquina para prever o resultado de uma variável dependente com base na análise de dados anterior. Ele é usado para a classificação binária, ou seja, um resultado ou seu oposto, por meio de um processo de aprendizagem supervisionada. Uma aplicação prática é a detecção de spam em um serviço de e-mail.

Redes Neurais - Neural Networks

Refere-se à base de funcionamento da aprendizagem profunda que consiste em imitar as funções neuronais do cérebro humano. A rede neural trabalha com camadas de processamento, onde cada neurônio é uma equação matemática simples que se conecta a outra, e isso é repetido em toda a camada. Assim, em cada camada, as informações desnecessárias são eliminadas e uma representação cada vez mais simples e precisa dos dados é mantida. Esses tipos de algoritmos exigem grandes conjuntos de dados e permitem modelar vários resultados em simultâneo. Essas redes podem



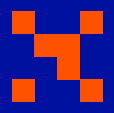
ser utilizadas em atividades como a detecção automatizada de rostos humanos (reconhecimento facial) e a contagem de caracteres.

Processamento de Linguagem Natural - Natural Language Processing (NLP)

Título dado ao campo da IA empregado nos sistemas com os quais as pessoas interagem, como assistentes operados por comandos de voz e com resposta verbal. Está focado no desenvolvimento e otimização de programas que processam a linguagem humana com o propósito de deduzir o seu conteúdo. Dentro do NLP encontra-se o Entendimento de Linguagem Natural (Natural Language Understanding) que consiste em transformar a linguagem humana em um formato legível para as máquinas. Em paralelo, está a Geração de Linguagem Natural (Natural Language Generation) que consiste na produção de narrativas faladas ou escritas a partir de um conjunto de dados.

Sistema de Visão Computacional - Computer Vision System

Esta é uma área científica interdisciplinar dedicada à exploração, reconhecimento e análise de imagens e vídeos. Entre as tarefas realizadas estão o reconhecimento e o rastreamento de objetos, a detecção de eventos e movimentos, entre outros. As técnicas de visão computacional são utilizadas em aplicativos como a detecção de objetos em veículos autônomos e o reconhecimento facial automatizado. O mau



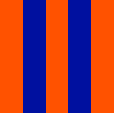
desenvolvimento destes pode ter consequências catastróficas, enquanto um grande aperfeiçoamento pode facilitar trabalhos de vigilância massiva e discriminatória.

Data Brokers

São pessoas ou empresas envolvidas na comercialização de dados pessoais obtidos a partir do uso de serviços na internet e de outras fontes (incluindo outros data brokers). São também orientados para a criação de perfis (por exemplo, para direcionar campanhas de marketing), para o qual coletam e analisam informações obtidas de fontes públicas e privadas de usuáries. Embora concentrem uma grande quantidade de informações pessoais, geralmente operam sem serem detectados por usuáries de tecnologias, que permanecem sem o controle da qualidade desses dados e da quantidade de pessoas ou empresas que os possuem.

Viés de Algoritmo - Algorithmic Bias

É um valor de defasagem da origem de uma fórmula, um defeito que leva a um resultado incorreto ou injusto. Nos sistemas de aprendizagem de máquina, dados os efeitos sociais que o viés pode ter hoje, deve ser entendido a partir da ética da IA, pelo menos com dois significados: o viés indutivo, ou seja, os estereótipos ou preconceitos de quem está projetando um sistema de coleta ou processamento de dados sendo incorporados ao funcionamento do sistema. O segundo significado é o viés de confirmação, o qual é um erro sistemático introduzido por um procedimento de amostragem

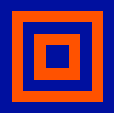


ou de informação disponível, em outras palavras, são os vieses que podem existir na conformação do banco de dados. Para reduzir esses vieses é possível impedir, por meio de um sistema de aprendizagem supervisionada, que os resultados tendenciosos de um algoritmo se generalizem. No entanto, é necessário auditar regularmente um sistema para evitar que novos vieses se manifestem.

Discriminação Algorítmica

É o resultado inadequado, incorreto, irregular e injusto que ocorre após a operação de um sistema automatizado de previsão, classificação ou tomada de decisão. Este resultado tende a prejudicar certas pessoas ou grupos. Os algoritmos costumam ter vieses (ou parcialidades), pois tanto na definição de variáveis como no desenvolvimento do código se imprimem valores das pessoas encarregadas do seu desenho. Também é possível que ocorram vieses no conjunto de dados com o qual os algoritmos são treinados, dependendo dos critérios usados para coletar e processar esses dados. Isso é bastante perceptível em sistemas de aprendizagem de máquina para a tomada de decisões e modelos preditivos. Por exemplo, no processamento de linguagem natural e sistemas de visão computacional quando são implementados na: gestão pública ou assistência social; seleção e contratação de pessoal; alocação de créditos ou benefícios sociais; aplicação da lei; entre outras atividades.

As previsões algorítmicas podem reforçar a discriminação contra grupos tradicionalmente vulnerabilizados, com base



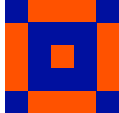
na classe socioeconômica, renda, raça, etnia, gênero, etc. Até agora, esses processos foram caracterizados por sua opacidade. Por isso, há pouco foram estabelecidos mecanismos como a IA explicável (XAI). Eles permitem a verificação, transparência e regulação de sistemas de IA, a fim de reduzir os efeitos desproporcionais e injustos desses sistemas.

Transparência Algorítmica

É o princípio orientado para que os fatores que determinam as decisões tomadas por algoritmos operem de forma transparente e compreensível para as pessoas que os utilizam, regulam e sofrem os efeitos do processamento de dados realizado. Trata-se de um princípio essencial para detectar problemas de funcionamento e para facilitar a prestação de contas na operação de sistemas automatizados.

Proteção de Dados Pessoais

É a normativa que regula os direitos das pessoas sobre as próprias informações. Em particular, a respeito da coleta, gerenciamento e processamento dos dados para conhecer os detalhes do uso, corrigir os erros ou solicitar a eliminação destes. A especificidade e escopo da regulamentação variam conforme a estrutura legal de cada país e eficácia é crucial para o controle de um elemento fundamental dos sistemas de decisão automatizada: as informações relativas às pessoas.



Legalidade

Trata-se de um princípio fundamental do direito público, que orienta as ações da função pública estritamente no âmbito das normas jurídicas determinadas pelos órgãos competentes. A legalidade exige que todas as decisões públicas devem respeitar regras formais de competência, além das regras substantivas sobre direitos fundamentais e legais.

Necessidade e Proporcionalidade

O princípio da necessidade é considerado um princípio da função pública que exige que a medida implementada seja requerida para o resguardo de um direito ou o cumprimento de um objetivo legítimo. Isso é realizado após a avaliação da possibilidade de outras decisões ou soluções também eficazes, mas potencialmente prejudiciais.

A proporcionalidade é somada à análise da necessidade e é um princípio que exige a ponderação equilibrada entre os meios adotados para alcançar fins legítimos. Assim, a concretização desses fins não terá grandes consequências negativas. Isto implica analisar a adequação das medidas para a consecução dos fins.

Nas decisões públicas que utilizam a IA, a aplicação destes princípios envolve: uma decisão informada sobre a adequação de um sistema automatizado para assistir ao cumprimento de funções públicas; os interesses que serão afetados pela sua implementação; a quantidade de informação estritamente necessária para o seu funcionamento; a existência de alternativas; entre outros aspectos.



Ética da Inteligência Artificial - Ethics of Artificial Intelligence

Refere-se aos conjuntos de princípios que tentam influenciar o desenvolvimento e a implementação de sistemas de IA para evitar os seus riscos. Assim, para reduzir os riscos de discriminação algorítmica, diferentes empresas e instituições no mundo contam com princípios éticos para o planeamento, treinamento e implementação de sistemas de IA e de aprendizagem de máquina. Embora existam dezenas de iniciativas sobre ética e princípios, em novembro de 2021 a Unesco publicou uma série de recomendações éticas adotadas por 193 Estados-membros. Ainda que estas considerações não sejam vinculativas, até agora são as normas com mais ampla aceitação mundialmente.

IA Explicável - Explainable AI (XAI)

É um conjunto de processos e métodos que permitem a usuáries entender os resultados e produtos criados por algoritmos de aprendizagem de máquina. A XAI consiste na superação da ideia de caixa oculta onde os sistemas de IA são desenvolvidos. Ela é fundamental para gerar confiança nestes sistemas ao descrever um modelo, o impacto esperado, os possíveis vieses e mecanismos de reparação ante potenciais abusos. Facilita a auditabilidade dos sistemas e também habilita quem os desenvolve para ajustá-los da melhor forma.

Inteligência Artificial e Direitos Humanos

Alguns conceitos básicos

Esta publicação foi realizada pela equipe da Derechos Digitales, no âmbito do projeto "Inteligência Artificial e Inclusão", coordenado por Jamila Venturini, Juan Carlos Lara e Patricio Velasco. A presente edição foi revisada, atualizada e traduzida.

Para mais informações sobre este projeto, visite: <https://ia.derechosdigitales.org>

Texto por Patricio Velasco e Juliana Guerra.

Edição e correções por Juan Carlos Lara e Ileana Silva.

Desenho e diagramação por Comunas Unidas.

Tradução e adaptação por Urgas Traduc.toras.



DD-AI-01



Este trabalho está disponível sob a licença Creative Commons Atribuição 4.0 Internacional
<https://creativecommons.org/licenses/by/4.0/deed.es>