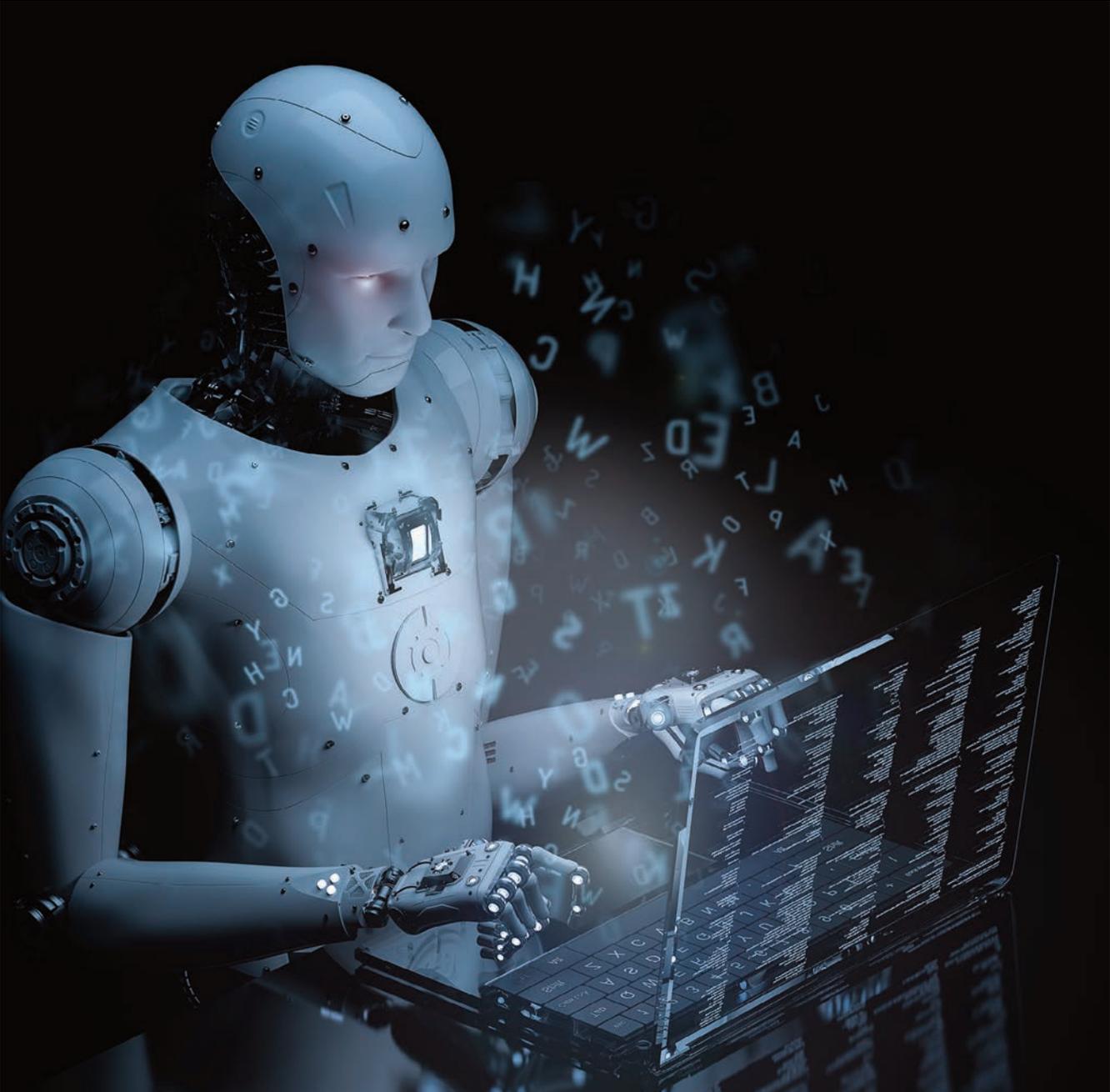


# Glossário



**Aprendizado de máquina (*machine learning*):** subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que as máquinas aprendam e melhorem seu desempenho em uma tarefa específica por meio da experiência.

**Caixa branca (*white box*):** sistema ou modelo de AI cujo funcionamento interno é simples de entender e explicar.

**Caixa preta (*black box*):** sistema ou modelo de AI cujo funcionamento interno é desconhecido ou difícil de entender.

**Direito a uma explicação:** conceito jurídico que sustenta que os indivíduos têm o direito de saber como são tomadas as decisões automatizadas que os afetam e de receber uma explicação compreensível de como funcionam os algoritmos envolvidos.

**Explicabilidade:** capacidade de um sistema de AI de fornecer justificativas claras e compreensíveis de suas previsões ou decisões aos usuários e às partes interessadas. Isso envolve o fornecimento de informações detalhadas e contextualizadas sobre como e por que um modelo de AI chega a uma determinada conclusão, facilitando a confiança e a adoção da tecnologia.

**GPT-4:** quarta geração do modelo Generative Pre-trained Transformer, desenvolvido pela OpenAI Foundation, que é usado para tarefas de processamento de linguagem natural e de geração de texto.

**Inteligência artificial (AI):** campo de estudo que busca desenvolver sistemas capazes de realizar tarefas que normalmente exigem inteligência humana, como aprendizado, raciocínio, percepção e tomada de decisões.

**Inteligência artificial explicável (XAI):** abordagem de IA que busca tornar os modelos de inteligência artificial mais compreensíveis e transparentes para os seres humanos.

**Interpretabilidade:** facilidade com que os seres humanos podem entender o processo de tomada de decisão de um modelo de IA, bem como as relações entre os recursos de entrada e as previsões ou decisões. Um modelo interpretável permite que os usuários entendam como se chega a uma previsão ou decisão específica.

**LIME (*Local Interpretable Model-agnostic Explanations*):** técnica de explicabilidade que ajuda a entender as previsões individuais de um modelo de AI criando aproximações locais interpretáveis.

**Modelo sub-rogado:** modelo interpretável que é treinado para imitar as previsões de um modelo de AI complexo e menos interpretável, como uma rede neural profunda. O objetivo de um modelo sub-rogado é fornecer uma explicação simplificada e compreensível de como o modelo original toma decisões.

**Open AI Foundation:** organização de pesquisa e desenvolvimento de inteligência artificial, atualmente participada da Microsoft, cujo objetivo declarado é garantir que a AI beneficie toda a humanidade.

**Partial Dependence Plot (PDP):** técnica de visualização que mostra o efeito médio de uma característica nas previsões de um modelo de AI, mantendo constantes todas as demais características. Ela ajuda a entender a relação entre características e previsões e a detectar possíveis interações e não linearidades.

**Rede neural profunda:** tipo de algoritmo de machine learning que consiste em várias camadas de neurônios artificiais e é capaz de aprender representações hierárquicas de dados.

**Regulamento Geral de Proteção de Dados (GDPR):** legislação da União Europeia que estabelece regras para a coleta, o armazenamento e o processamento de dados pessoais de cidadãos da UE.

**SHAP (*SHapley Additive exPlanations*):** técnica de explicabilidade que usa valores de Shapley da teoria dos jogos cooperativos para atribuir a importância de cada variável na previsão de um modelo de AI.

**Sparsity:** propriedade de um modelo em que ele considera apenas o subconjunto de variáveis que são realmente relevantes para a estimativa.

**Teste de esquema de Winograd:** teste de compreensão de linguagem natural que avalia a capacidade de uma AI de resolver ambiguidades na linguagem por meio do uso de conhecimento e raciocínio comuns.

**Teste de Turing:** teste proposto por Alan Turing em 1950 que avalia a capacidade de uma máquina de imitar a inteligência humana a ponto de ser indistinguível de um ser humano em uma conversa.

**Transformer:** arquitetura de rede neural introduzida pelo Google Brain em 2017, usada principalmente em tarefas de processamento de linguagem natural (NLP). Os transformers são conhecidos por sua capacidade de lidar com longas sequências de dados e por sua eficiência no treinamento. Eles se baseiam em mecanismos de atenção, que permitem que a rede pondere a importância relativa de palavras ou elementos em uma sequência ao longo do tempo. Os transformers impulsionaram o desenvolvimento de modelos de linguagem de última geração, como o GPT e o BERT, e revolucionaram o campo da NLP.

**Transparência:** abertura e acessibilidade de um sistema de AI em termos de seu desenho, estrutura e processos internos. Um sistema transparente permite que os usuários e as partes interessadas examinem e entendam seus componentes, algoritmos e decisões.