

Conclusão

Com a programação correta, um computador pode se tornar um teatro, um instrumento musical, um livro de referência, um oponente de xadrez. Nenhuma outra entidade no mundo, a não ser o ser humano, tem uma natureza tão adaptável e universal.

Daniel Hillis⁷⁴



Este estudo apresentou a Inteligência Artificial Explicável (XAI), seus fundamentos, contexto e técnicas para melhorar a interpretabilidade dos modelos. Foram discutidos os principais desafios enfrentados pelos modelos de inteligência artificial em termos de interpretabilidade e como a tecnologia pode ajudar a resolvê-los, incluindo um estudo de caso desenvolvido com o ModelCraft™ para demonstrar como essas técnicas podem ser empregadas para entender e explicar os modelos de AI.

A disciplina de AI, e dentro dela a XAI, cresceu em importância globalmente nos últimos anos, já que o desenvolvimento de tecnologias de AI de alto desempenho se tornou uma prioridade para muitos setores, da saúde à segurança, dos serviços financeiros à energia, entre outros. A interpretabilidade surge como uma necessidade para entender e aprimorar os modelos de AI, o que é particularmente complexo para determinadas técnicas.

Como vimos, os modelos de AI podem enfrentar dificuldades para explicar os resultados ou a lógica por trás de suas decisões. Isso ocorre porque esses modelos usam técnicas de aprendizagem profunda e algoritmos complexos para aprender com os dados, que muitas vezes são difíceis de interpretar, o que apresenta desafios na avaliação dos modelos de AI e na confiabilidade de seus resultados.

Como resultado, o framework regulatório de AI está evoluindo rapidamente, e espera-se que as organizações se adaptem aos novos requisitos de transparência, explicabilidade e equidade no uso de modelos de AI. Isso implica a necessidade de uma abordagem holística que integre a interpretabilidade e a explicabilidade na organização e nos processos de cada empresa, abrangendo técnicas de interpretabilidade, gestão de riscos de modelos, colaboração interdisciplinar e treinamento em XAI para profissionais envolvidos no desenvolvimento e na aplicação da AI, entre outros.

Concluindo, a interpretabilidade dos modelos de inteligência artificial é uma área emergente de pesquisa e espera-se que continue a se desenvolver e a crescer em importância à medida que os modelos de AI se tornam mais complexos, a regulamentação continua a proliferar e seu uso se estende a domínios mais altamente sensíveis.

⁷⁴Daniel Hillis (n. 1956), inventor, empresário e cientista americano, pioneiro da computação paralela e de seu uso no campo da inteligência artificial, com mais de 300 patentes publicadas.