

# Ética & Inteligência Artificial

Marcelo Fantini RA 108341  
Rubens de Castro Pereira RA 217146

24 de março de 2023

## 1. Defina Ética em Inteligência Artificial.

A ética em Inteligência Artificial (IA) representa um conjunto de valores e princípios que devem ser respeitados por todos os atores envolvidos no ciclo de vida de sistemas de IA com destaque ao respeito, proteção e promoção dos direitos humanos, liberdades fundamentais e dignidade humana, diversidade e inclusão [1].

## 2. Apresente uma notícia recente de um problema de Ética em IA.

Escreva sua resposta aqui!

## 3. Apresente um artigo científico recente de uma solução para um problema de Ética em IA.

Escreva sua resposta aqui! O artigo [2] tem como propósito tratar dos riscos potenciais e uso indevido de geração de conteúdo por IA (GCIA), auxiliar na eliminação de obstáculos e promover entregas seguras e éticas de conteúdos gerados por IA. Os tipos de conteúdos podem ser bastante variados podendo ser imagens, textos, áudios e vídeos.

Com o uso extensivo da GCIA tem surgido preocupações relativas à privacidade, preconceito, toxicidade, desinformação, propriedade intelectual e potencial uso indevido diante das pessoas. Uma questão relevante surgiu recentemente com a disponibilização de novas funcionalidades no ChatGPT as quais permitem fazer a depuração de código fonte de programas de computador ou elaborar trabalhos escolares/acadêmicos. Esses resultados podem gerar riscos potenciais, pois os modelos de GCIA produzem trabalhos replicando conteúdos com os quais foram treinados devido à elevada capacidade de memorização. O conjunto de dados utilizados para treinamento frequentemente tem origem e direitos autorais desconhecidos, muitas das vezes não passam por uma análise de curadoria cuidadosa. A maioria dos modelos de GCIA decodificadores de textos são treinados com grandes quantidades de dados obtidos da internet, os quais podem conter desvios (bias) relacionados a temas sociais, podem ser tóxicos, e outras limitações inerentes aos grandes modelos de linguagens.

Para que os modelos de GCIA sejam responsáveis, estes devem considerar o seguinte escopo: privacidade, vies (tendências), toxicidade, desinformação, proteção da propriedade intelectual. Além disso, deve contemplar também a robustez dos sistemas, explicabilidade (feedback), código fonte aberto, consentimento, créditos e compensação, e ambiente amigável para o seu uso.

Privacidade Os modelos base e os modelos generativos de conteúdos possuem a O vazamento de informações

Tendências/bias

Toxicidade

Desinformação

Proteção da propriedade intelectual

Discussão

Conclusão

## Referências

- [1] Scientific United Nations Educational and Cultural Organization (UNESCO). Recommendation on the ethics of artificial intelligence, 2022.
- [2] Chen Chen, Jie Fu, and Lingjuan Lyu. A pathway towards responsible ai generated content, 2023.