

Existe ética no treinamento das IA

Nos últimos anos, todas as grandes empresas de tecnologia vêm desenvolvendo e lançando sistemas de inteligência artificial, visando explorar um mercado com enorme potencial. Nesse contexto, surgem questionamentos relevantes: de que forma e em quais ambientes essas IAs são treinadas? Quais bancos de dados são empregados nesse processo? Esses conjuntos de dados são de acesso público ou há, em alguma medida, apropriação indevida de propriedade intelectual?

Um caso recente ilustra a complexidade ética envolvendo a inteligência artificial na reprodução de estilos artísticos consagrados. Trata-se do estilo de animação característico dos estúdios Ghibli, responsáveis por obras como *A Viagem de Chihiro* e *O Menino e a Garça*. Com o lançamento de uma nova ferramenta de geração de imagens por IA pela OpenAI, usuários em redes sociais buscaram replicar o estilo único criado por Hayao Miyazaki, compartilhando imagens geradas pela tecnologia.

Em um vídeo amplamente difundido de 2016, Miyazaki expressa seu posicionamento crítico em relação à arte produzida por IA qualificando-a como um “insulto à própria vida”. Reconhecido por sua animação meticulosamente elaborada quadro a quadro, ele afirma: “Estou completamente enojado. Se você realmente quer fazer coisas assustadoras, pode ir em frente e fazer, mas eu nunca desejaria incorporar essa tecnologia ao meu trabalho.”

O episódio evidencia debates contemporâneos sobre ética na IA, especialmente em relação ao treinamento de modelos que reproduzem características exclusivas de obras artísticas pré-existent, levantando questões sobre propriedade intelectual e apropriação de estilos criativos consagrados.

Um caso recente ilustra questões éticas e legais envolvendo o uso de dados protegidos por direitos autorais no treinamento de modelos de inteligência artificial. A Meta, empresa proprietária do Instagram, Facebook e WhatsApp, foi processada nos Estados Unidos por supostamente ter baixado ilegalmente mais de 80 TB de dados via torrents. Documentos internos indicam que o próprio CEO, Mark Zuckerberg, teria aprovado a utilização de versões pirateadas desses conteúdos para o treinamento dos modelos de IA da empresa, mesmo após alertas da equipe executiva de IA sobre a natureza ilícita do conjunto de dados, descrito como “sabidamente pirateado”.

As comunicações internas destacam que o uso desse banco de dados poderia enfraquecer a posição da empresa em negociações com órgãos reguladores. Conforme registrado nos documentos: “A cobertura da mídia sugerindo que usamos um conjunto de dados que sabemos ser pirateado, como o LibGen, pode minar nossa posição de negociação com os órgãos reguladores.”

O episódio evidencia que a principal preocupação da Meta não estava centrada na ética do uso de conteúdo sem o consentimento dos autores, mas sim em preservar sua posição estratégica frente aos órgãos reguladores. Tal situação ressalta a necessidade de reflexão crítica sobre os limites legais e morais no desenvolvimento de tecnologias de inteligência artificial, especialmente quando envolvem materiais protegidos por direitos autorais.

Um caso recente evidencia questões éticas e legais relativas ao uso de dados de usuários no treinamento de sistemas de inteligência artificial. O Google foi processado sob a alegação de ter coletado informações de milhões de usuários sem seu consentimento, com o objetivo de desenvolver e treinar seus produtos de IA. A

ação, movida pela Clarkson Law Firm, afirma que a empresa “tem roubado secretamente tudo o que já foi criado e compartilhado na internet por centenas de milhões de americanos” e utilizado esses dados para aprimorar seus modelos.

O processo também indica que o Google teria apropriado “praticamente toda a nossa pegada digital”, incluindo “trabalhos criativos e escritos”, para construir seus produtos de inteligência artificial. A denúncia faz referência a uma atualização recente da política de privacidade da empresa, a qual declara explicitamente que informações publicamente acessíveis podem ser utilizadas no treinamento de modelos e ferramentas de IA.

Em resposta, o Google afirmou que sua política “há muito tempo é transparente quanto ao uso de informações disponíveis publicamente na web aberta para treinar modelos de linguagem em serviços como o Google Tradutor. Esta atualização mais recente apenas esclarece que serviços mais recentes, como o Bard, também estão incluídos.”

Este episódio ressalta a complexidade ética e jurídica envolvida na coleta e utilização de dados para o treinamento de sistemas de inteligência artificial, evidenciando a necessidade de regulamentações claras e da proteção da privacidade e dos direitos autorais dos usuários.

Uma questão central que emerge no contexto do treinamento de sistemas de inteligência artificial refere-se à ética envolvida nesses processos. Até que ponto é aceitável sacrificar princípios éticos em função do lucro em uma corrida tecnológica cada vez mais competitiva? Observa-se que, no estágio atual de desenvolvimento das grandes empresas de tecnologia, há pouca preocupação com o uso de conteúdos protegidos por direitos autorais ou com a coleta não autorizada de dados de usuários. O que parece predominar é a ênfase em manter os sistemas de IA atualizados e relevantes, a fim de atrair investidores e maximizar retornos financeiros.

Diante desse cenário, algumas soluções poderiam ser implementadas para conciliar inovação e ética. Entre elas, destacam-se: a criação de bancos de dados nos quais os autores possam disponibilizar suas obras mediante remuneração; a contratação de artistas para a criação de conteúdos originais, assegurando pagamento sempre que suas obras forem utilizadas; ou a utilização de materiais em domínio público. Mesmo que a remuneração seja apenas uma pequena fração do valor gerado, tal medida representaria um avanço significativo. No entanto, enquanto a lógica do lucro prevalecer sobre considerações éticas, é provável que práticas questionáveis continuem a ocorrer no desenvolvimento de tecnologias de inteligência artificial.

Fontes:

<https://www.theguardian.com/technology/2025/jan/10/mark-zuckerberg-meta-books-ai-models-sarah-silverman>

<https://www.cnnbrasil.com.br/tecnologia/entenda-a-polemica-envolvendo-ia-e-o-studio-ghibli/>

<https://forbes.com.br/forbes-tech/2024/12/openai-na-berlinda-por-violacao-de-direito-autoral-veja-como-se-proteger/>

https://www.cnnbrasil.com.br/tecnologia/google-e-processado-por-roubar-dados-de-usuarios-para-treinar-suas-ferramentas-de-ia/#goog_rewarded