XX CODAIP

CONGRESSO DE DIREITO DE AUTOR E INTERESSE PÚBLICO.

Os limites da aplicação do conceito de reprodução às IAs generativas

Marcelo Frullani Lopes

Advogado, Mestre e Doutor / Universidade de São Paulo (USP)

Realização:

GEDAI Grupo de Estudos de Direito Autoral e Industrial Universidade Federal do Paraná

UFPR UNIVERSIDADE FEDERAL DO PARANÁ

PPGD PROGRAMA DE PÓS GRADUAÇÃO EM DIREITO Mestrado, Doutorado e Pós-doutorado

APRESENTAÇÃO DO PALESTRANTE



Marcelo Frullani Lopes

Advogado, Mestre e Doutor / Universidade de São Paulo (USP)

Minicurrículo

Graduado pela Faculdade de Direito da Universidade de São Paulo (USP). Especialista em Direito e Tecnologia da Informação pela Escola Politécnica da USP. Mestre e Doutor em Filosofia e Teoria Geral do Direito pela USP. Autor do livro “Obras geradas por inteligência artificial: desafios ao conceito jurídico de autoria.

https://br.linkedin.com/in/marcelofrullani • https://www.frullanilopes.adv.br • marcelo@frullanilopes.adv.br

RECORTE DO TEMA DA PESQUISA

- Aplicabilidade e limites do conceito jurídico de reprodução na IA generativa
- Foco no treinamento, no modelo treinado e na geração de resultados
- Análise conceitual: identificar a reprodução antes de discutir a licitude dos usos
- Hipótese: aplicação parcial, especialmente limitada quanto ao próprio modelo

PERGUNTA CENTRAL

Em que medida o conceito jurídico de reprodução se aplica ao treinamento, ao modelo treinado e à geração de resultados?

O CONCEITO CLÁSSICO DE REPRODUÇÃO

- Conceito criado quando as cópias eram físicas
- Direito exclusivo de autorizar a produção de exemplares
- Fixação cria o primeiro exemplar. Reprodução gera cópias a partir da primeira fixação
- A reprodução juridicamente relevante para o direito autoral se dá sobre expressões, não sobre ideias

CONCEITO-CHAVE

“Quando se fala em ‘direito de reprodução’ só se tem em vista a produção de exemplares.”
Ascensão (1997, p. 174)

O IMPACTO DA DIGITALIZAÇÃO

- As obras passam a ter representação numérica em bits. A cópia se dissocia do exemplar físico.
- A diminuição dos custos de reprodução e distribuição. Cópias digitais mantêm a qualidade do original.
- Muitas cópias são realizadas com o uso cotidiano dos computadores e da internet, muitas vezes sem percepção pelo usuário.
- O conceito jurídico se amplia para o armazenamento eletrônico, inclusive temporário, com limitações.

A BANALIZAÇÃO DAS CÓPIAS NO AMBIENTE DIGITAL

- **Reprodução temporária**

  RAM e cache geram cópias transitórias para executar programas ou acessar páginas em ambiente virtual. A lei brasileira prevê limites à exclusividade no art. 30, § 1º.

- **Web scraping**

  Extração automatizada de dados da web, que pode envolver cópias de obras protegidas. A licitude depende do uso concreto e das limitações aplicáveis.

- **Mineração de textos e dados**

  Análise automatizada de grandes volumes de dados para identificar padrões e correlações, que geralmente depende da realização de cópias e da inclusão de obras em bases de dados.

O REAPROVEITAMENTO DE OBRAS NO AMBIENTE DIGITAL

- **Remix**

  Recombinação e transformação de obras preexistentes para criar algo novo. A digitalização facilita o uso de obras alheias como insumo criativo.

- **Memes**

  Imagens que usuários replicam e transformam continuamente. A circulação pode gerar valor para o criador e diluir a autoria em contribuições coletivas.

- **Projetos criativos colaborativos**

  Em projetos de software livre, usuários podem copiar, modificar e aperfeiçoar programas conforme suas licenças. A reprodução integra um processo de criação coletiva.

AS FUNÇÕES DA REPRODUÇÃO NO AMBIENTE DIGITAL

- **Circulação de informações**

  Cópias técnicas sustentam o acesso e o funcionamento da rede, bem como o desenvolvimento de serviços, produtos e pesquisas. A reprodução integra a infraestrutura digital.

- **Insumo para a criação**

  O acesso a obras e o baixo custo das cópias favorecem o reaproveitamento criativo. Remix, memes e projetos colaborativos combinam reprodução e transformação criativa.

- **A fronteira entre reprodução e criação**

  A cópia assume funções técnicas e criativas. A fronteira entre reprodução e criação torna-se menos nítida.

AS FUNÇÕES DA REPRODUÇÃO NA IA GENERATIVA

- **Digitalização: cópias com funções distintas**

  Algumas cópias viabilizavam a circulação de informações no ambiente digital. Outras serviam de insumo para a criação.

- **IA generativa: as mesmas cópias, duas funções**

  As mesmas cópias cumprem as duas funções: viabilizam a circulação de informações e servem de insumo para a criação.

- **A fronteira entre reprodução e criação**

  A sobreposição dessas funções nas mesmas cópias torna ainda menos nítida a fronteira entre reprodução e criação.

A INTELIGÊNCIA ARTIFICIAL GENERATIVA

- **Aprendizado de máquina**

  O treinamento ajusta parâmetros de um modelo matemático a partir de dados. Esses parâmetros refletem padrões estatísticos que o modelo utiliza para fazer inferências sobre novas entradas.

- **Modelo fundacional**

  Modelo treinado em larga escala com dados variados, capaz de servir de base a diversas tarefas. Pode passar por refinamento para aplicações específicas e integrar sistemas de IA generativa.

MODELOS DE LINGUAGEM E DIFUSÃO DE IMAGENS

- **Grandes modelos de linguagem (LLMs)**

  Aprendem relações estatísticas em grandes conjuntos de textos, divididos em tokens. A partir do contexto, geram texto ao estimar probabilidades para o próximo token, repetindo o processo a cada etapa.

- **Modelos de difusão de imagens**

  No treinamento, aprendem a remover ruído adicionado às imagens. Na geração, partem de ruído aleatório e o reduzem progressivamente, sob orientação do prompt, para produzir uma imagem.

DOIS CONCEITOS FUNDAMENTAIS

- **Representação abstrata de dados**

  O modelo aprende características e correlações estatísticas dos dados e as representa matematicamente. Essas representações permitem generalizar e gerar novos conteúdos, sem guardar cada obra como um arquivo individualizado.

- **Memorização de dados de treinamento**

  Informações específicas dos exemplos de treinamento podem permanecer nos parâmetros e permitir a reconstrução de trechos exatos ou quase exatos. Isso pode aparecer nos resultados por extração dirigida ou regurgitação, mesmo sem intenção do usuário.

REPRODUÇÃO NA FASE DE TREINAMENTO

- **Cópias em diferentes etapas**

  Nas fase de coleta e pré-processamento, geralmente há copias de expressões protegidas, seja para download ou para mudanças de formato.
  Na fase de treinamento em sentido estrito (ajuste de parâmetros), a ocorrência de novas cópias é controversa.
  Apesar de haver reprodução, não há relação direta entre o número de cópias utilizadas a importância da obra para o desempenho do modelo.

- **Reprodução e análise da licitude**

  Mesmo sem comunicação ao público, as cópias integram o conceito jurídico de reprodução.
  Se não houver autorização do titular de direitos autorais, a licitude depende da avaliação da aplicação de exceções e limitações ou do *fair use*, por exemplo.

REPRODUÇÃO NO MODELO DE IA GENERATIVA

- **A tese das “cópias recuperáveis”**

  Alguns itulares de direitos autorais e parte da doutrina tratam a memorização e a geração de *outputs* semelhantes como indícios de que há cópias fixadas nos parâmetros do modelo.
  Desenvolvedores contrapõem o aprendizado de padrões estatísticos ao armazenamento de obras.

- **Conclusão da tese para o direito brasileiro**

  O modelo contém representações abstratas. A memorização, por si só, não permite equipará-lo a cópias compactadas ou a uma obra derivada.
  Eventuais violações nos *outputs* exigem demonstração concreta da semelhança. Análise precisa ser feita caso a caso.

REPRODUÇÃO NA GERAÇÃO DE OUTPUTS

- **Reprodução no processo de recuperação (RAG)**

  O sistema de IA consulta bases de dados ou fontes externas para responder aos *prompts* e normalmente envolve cópias de expressões protegidas.
  Essas cópias têm relevância jurídica. Sua licitude exige análise própria, distinta da avaliação do *output*.

- **Similaridade e responsabilidade caso a caso**

  Deve-se comparer cada *output* com a expressão protegida da obra anterior. O uso de ideias ou fatos não basta para caracterizar reprodução. Além disso, não se pode dizer que todos os *outputs* gerados sejam cópias ou obras derivadas.
  A responsabilidade depende das características do sistema, dos filtros e da conduta do usuário, sem atribuição automática.

RESUMO DA APLICAÇÃO DO CONCEITO A CADA FASE

- **Treinamento e modelo**

  Treinamento: as cópias integram o conceito de reprodução, mas sua quantidade não revela a importância de cada obra para o modelo.
  Modelo: representações abstratas e memorização não equivalem ao armazenamento de cópias individualizadas.

- **Geração de outputs**

  RAG: as cópias realizadas na recuperação de conteúdo integram o conceito jurídico de reprodução.
  Outputs: a reprodução exige comparação caso a caso com obras preexistentes. A origem no treinamento não torna todo resultado uma cópia ou obra derivada.

RESUMO DA APLICAÇÃO DO CONCEITO A CADA FASE

- **Treinamento e modelo**

  Treinamento: as cópias integram o conceito de reprodução, mas sua quantidade não revela a importância de cada obra para o modelo.
  Modelo: representações abstratas e memorização não equivalem ao armazenamento de cópias individualizadas.

- **Geração de outputs**

  RAG: as cópias realizadas na recuperação de conteúdo integram o conceito jurídico de reprodução.
  Outputs: a reprodução exige comparação caso a caso com obras preexistentes. A origem no treinamento não torna todo resultado uma cópia ou obra derivada.

PROPOSTAS NORMATIVAS E LIMITES DO CONCEITO

Tese central: há cópias no treinamento e em certos *outputs*, mas o modelo não armazena obras individualizadas.

- ## Medidas compatíveis com a tese

  Exceções de TDM e opt-out: podem disciplinar as cópias no treinamento, mas não resolvem eventuais reproduções nos outputs;
  Transparência: divulgar os dados de treinamento ajuda a comprovar o uso de obras. Não garante remuneração nem impede infrações;
  Medidas preventivas: filtros, refinamento e alinhamento podem reduzir reproduções indevidas nos *outputs*, sem eliminar todo o risco.

- ## Propostas que exigem delimitação

  Remuneração: licenças para o treinamento são coerentes. Contar cópias não mede a contribuição de cada obra. Tributos ou fundos podem redistribuir valor sem esse critério.
  Medidas reativas: excluir obras da base para futuros treinos e filtrar *outputs* são medidas possíveis, mas “remover obras” do modelo parte de premissa inadequada.

XX CODAIP

CONGRESSO DE DIREITO DE AUTOR E INTERESSE PÚBLICO

Muito obrigado!

Marcelo Frullani Lopes

marcelo@frullanilopes.adv.br

https://br.linkedin.com/in/marcelofrullani /
https://www.frullanilopes.adv.br

INSERIR
QR CODE

[HASHTAG OFICIAL]

Realização:

GEDAI
Grupo de Estudos de Direito Autoral e Industrial

UFPR

PPGD
PROGRAMA DE PÓS GRADUAÇÃO EM DIREITO
Mestrado, Doutorado e Pós-doutorado