XX CODAIP

CONGRESSO DE DIREITO DE AUTOR E INTERESSE PÚBLICO.

MODEL COLLAPSE E REMUNERAÇÃO DE AUTORES

Dr. Marcos da Cunha e Souza

Professor PPGD / UNINTER

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



Dr. Marcos da Cunha e Souza

Professor do PPGD/ UNINTER

Minicurrículo

Advogado, professor do Corpo Docente do PPGD da UNINTER, consultor da Comissão de Direito Internacional da OAB/PR. É mestre e doutor em Direito pela PUC/PR, tendo sido indicado ao prêmio CAPES. Foi, durante 15 anos, assessor da Procuradoria Regional da República no Rio de Janeiro.

E-mail: marcoscsouza@ymail.com
LinkedIn: www.linkedin.com/in/marcos-da-cunha-e-souza-57a305412

MODEL COLLAPSE

A busca do conceito. Exemplos.

MODEL COLLAPSE

Em 2024, Ilia Shumailov – Departmento de Ciência da Computação, University of Oxford - e outros 5 (*) publicaram na revista Nature um artigo demonstrando empiricamente que determinados processos de treinamento recursivo podem provocar deterioração progressiva da qualidade dos modelos de IA. Esse fenômeno recebeu a denominação de Model Collapse.

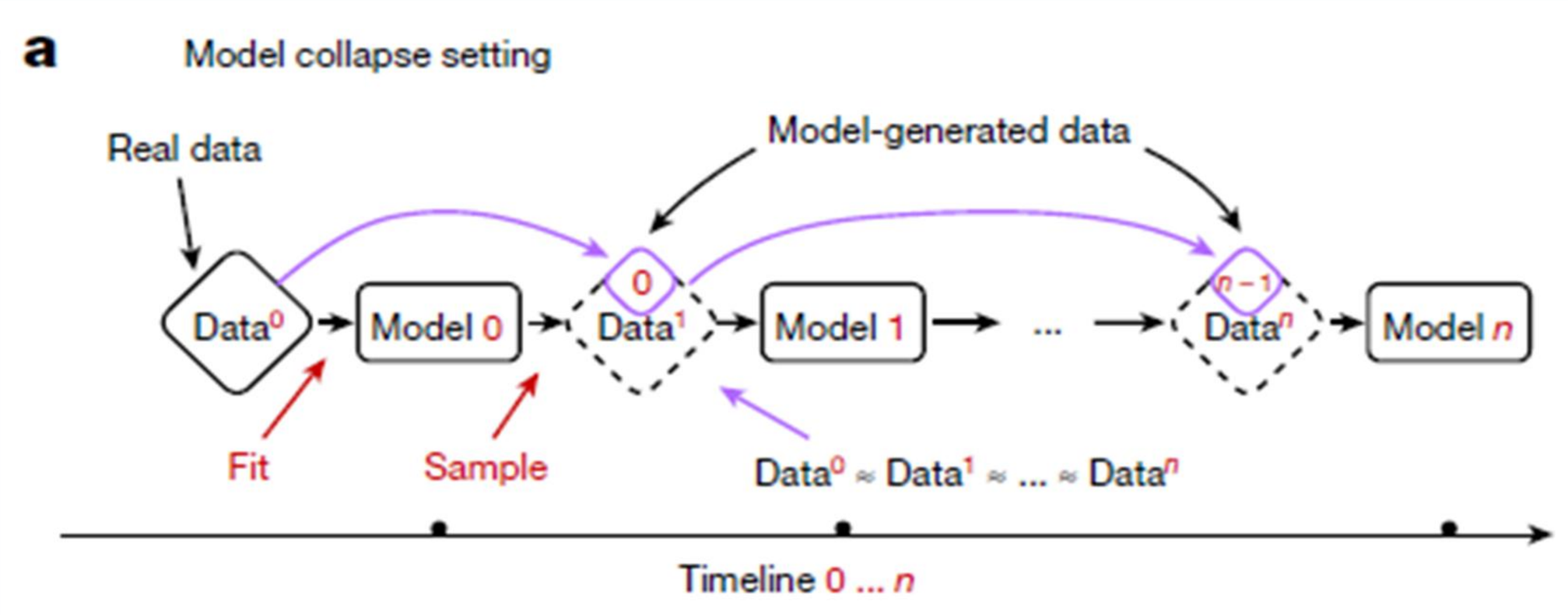
(*) I. Shumailov, et al., AI models collapse when trained on recursively generated data. Nature 631 (8022), 755–759 (2024), doi:10.1038/s41586-024-07566-y.

ESTRUTURA DO MODEL COLLAPSE (1ª ABORDAGEM)

- Bases de dados criadas por seres humanos ao longo de séculos são usadas para treinar modelos de IA → Utilização em massa da IA pelos usuários → Criação e disponibilização de dados sintéticos ao público → IA passa a ser treinada cada vez mais por dados sintéticos, que ganham peso estatístico → Desaparecimento progressivo da diversidade (caudas)

- **Supressão das “caudas” (“tails”): conhecimentos estatisticamente minoritários perdem representação, independentemente de sua veracidade, relevância histórica ou valor científico. Também pode gerar respostas erradas, pela degradação da base estatística dos dados.**

ESTRUTURA DO MODEL COLLAPSE (2ª ABORDAGEM)



Fonte: I. Shumailov, et al., AI models collapse when trained on recursively generated data. Nature 631 (8022), 755–759 (2024), doi:10.1038/s41586-024-07566-y.

CONCEITO DE MODEL COLLAPSE

- “O colapso do modelo é um processo degenerativo que afeta gerações sucessivas de modelos generativos treinados por aprendizado de máquina, no qual os dados por eles produzidos acabam contaminando o conjunto de dados utilizado para treinar a geração seguinte. Ao serem treinados com dados contaminados, esses modelos passam a perceber a realidade de maneira equivocada” (Shumailov et al., 2024, p. 755).

COLAPSO INICIAL E COLAPSO TARDIO

- **Colapso Inicial (early model collapse):** o modelo começa a perder informações sobre as caudas da distribuição estatística.

- **Colapso tardio (late model collapse) :** o modelo converge para uma distribuição que apresenta pouca semelhança com a distribuição original, frequentemente com uma variância substancialmente reduzida e empobrecimento do resultado.

Um exemplo concreto:

EXEMPLO DE SUPRESSÃO DE CAUDAS (1)

- Os modelos de linguagem aprendem como falamos partindo de regularidades estatísticas presentes nos textos de treinamento. Ao gerar respostas, as IAs tendem a reproduzir estruturas linguisticamente prováveis, gerando um homogeneização estilística, com a perda de representação de eventos raros.

- Consequência: Textos sobre assuntos completamente diferentes passam a apresentar estruturas argumentativas semelhantes, assim como expressões idiomáticas repetitivas. Exemplos bem conhecidos, verificáveis no Chat GPT: "A questão central não está em X , mas em Y” ; "O verdadeiro problema não é X, mas Y.“ ; “não é apenas... mas também”.

EXEMPLO DE SUPRESSÃO DE CAUDAS (2)

- Pesquisas do o professor japonês Masaru Yamada (Universidade de Rikkyo), permitem identificar o Model Collapse no uso da IA para a tradução de textos, causando o desaparecimento de expressões linguísticas raras. A situação se agrava porque a IA está tomando o lugar de tradutores profissionais, dificultando a sobrevivência de eventos raros.

EMPOBRECIMENTO DO RESULTADO

- Indo além do tema, nossa linguagem natural poderá ser afetada pela leitura de textos produzidos por IA. Talvez isso já esteja acontecendo.

O PARADOXO

Para evitar o colapso, a IA precisa dos autores humanos que ela, acidentalmente, está prejudicando profissionalmente.

DIREITO AUTORAL E MODEL COLLAPSE

Se a produção intelectual humana original deixar de ser economicamente sustentável (ex: tradutores, desenhistas, arquitetos, designers) poderemos perder justamente os autores que renovam, corrigem e preservam a diversidade cultural e as informações minoritárias.

É nesse ponto que a crise de remuneração autoral e o model collapse podem se encontrar.

PODE O FORTALECIMENTO DO DIREITO AUTORAL EVITAR O COLAPSO?

A NECESSIDADE DA PRODUÇÃO HUMANA

- Shumailov et al (2024, p. 755-756) observam que a degradação dos dados é mais lenta quando dados humanos são mantidos nos sucessivos modelos de treinamento.

Contudo, diante do avanço da IA, como garantir a continuidade de uma adequada/suficiente quantidade de produção 100% humana?

MASARU YAMADA PROPÕE QUATRO MEDIDAS

- A mais plausível seria via modelo semelhante ao ECAD brasileiro, para distribuição de remuneração aos tradutores, **pela contribuição que deram ao treinamento da IA**. Problema: como não podemos identificar os textos que treinaram a IA, como distribuir os direitos autorais?
  O problema foge do modelo tradicional de Direito Autoral, pois (geralmente) não há como identificar plágio. Não há reprodução da forma de expressão dos autores e tradutores.

SERÁ NECESSÁRIO AMPLIAR AS FRONTEIRAS DO DIREITO AUTORAL?

A União Europeia tem se preocupado com a proteção dos direitos autorais dentro do novo contexto tecnológico. Exemplos:

AI Act - Regulamento (UE) 2024/1689: Art. 53 – "... respeitar e proteger os direitos de propriedade intelectual...". Inclui procedimentos para identificar e respeitar esses direitos.

Diretiva (UE) 2019/790: Art. 4º §1º - permite exceção/limitação aos direitos autorais para reprodução e extração de obras legalmente acessíveis para fins de mineração de textos e dados.

SERÁ NECESSÁRIO AMPLIAR AS FRONTEIRAS DO DIREITO AUTORAL?

- **Resolução do Parlamento Europeu P10_TA(2026)0066**: voltada a aumentar a competitividade da UE no desenvolvimento de IA.
  "... sendo de sublinhar a importância de um setor criativo remunerado de forma justa enquanto fonte de dados de treino de IA de elevada qualidade."

Resolução do Parlamento Europeu P10_TA(2026)0066

- “Considera que os direitos de autor atualmente existentes são insuficientes para enfrentar o desafio do licenciamento de material protegido por direitos de autor para IA generativa; apela para a criação de um quadro jurídico adicional (...) incluindo um mercado de licenciamento funcional que restabeleça o poder de negociação dos titulares de direitos e soluções de proteção viáveis”

A solução ainda é vaga, mas está sendo buscada.

CONSIDERAÇÕES FINAIS

> *A proteção dos autores é mecanismo de justa distribuição da riqueza gerada pela IA, além de estímulo para garantir os insumos humanos necessários à própria preservação da Inteligência Artificial.*

REFERÊNCIA BÁSICA:

I. SHUMAILOV, Ilia et al.. AI models collapse when trained on recursively generated data. Nature 631 (8022), 755–759 (2024), doi:10.1038/s41586-024-07566-y.
II. YAMADA, Masaru. Translators as Invisible Teachers of AI: Copyright, Translation Memory, and the Political Economy of Linguistic Data. arXiv, versão 2, 2026. DOI: https://doi.org/10.48550/arXiv.2605.24842

XX CODAIP

CONGRESSO DE DIREITO DE AUTOR E INTERESSE PÚBLICO

Muito obrigado!

Marcos da Cunha e Souza

marcoscsouza@ymail.com

LinkedIn: www.linkedin.com/in/marcos-da-cunha-e-souza-57a305412

[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