

abpi.empauta.com

Associação Brasileira da Propriedade Intelectual
Clipping da imprensa

Brasília, 23 de junho de 2025 às 07h57
Seleção de Notícias

O Globo Online | BR

Direitos Autorais

'Estamos pressionando governos para proteger direitos autorais', diz CEO da HarperCollins ...	3
<small>ECONOMIA E NEGÓCIOS RUAN DE SOUSA GABRIEL</small>	

UOL Notícias | BR

23 de junho de 2025 | Direitos Autorais

Plano brasileiro para turbinar IA ignora conceito básico da tecnologia	6
--	---

'Estamos pressionando governos para proteger direitos autorais', diz CEO da HarperCollins

ECONOMIA E NEGÓCIOS



Líder do segundo maior conglomerado editorial do mundo elogia Bienal do Rio e diz que a inteligência artificial traz oportunidades, como a tradução de 'best-sellers' a um custo menor e a produção de audiolivros

Brian Murray está confiante: ao longo dos séculos, a indústria do livro sobreviveu a todas as novas tecnologias - e não será agora, diante da "tsunami" da inteligência artificial (IA), que vai submergir. O executivo manifesta preocupação em relação ao respeito aos **direitos** autorais e deixa claro que nunca usará a tecnologia para gerar livros, mas também vê no avanço da IA "oportunidades incríveis" para as editoras, como a possibilidade de produzir audiolivros e traduzir best-sellers a um custo menor.

Segundo maior conglomerado editorial do planeta, presente em 15 países, a HarperCollins opera no Brasil há uma década. Por aqui, destaca-se pelo catálogo de obras para público evangélico (agrupados no selo Thomas Nelson) e por dominar a lista de mais vendidos com livros de colorir desde o fim do ano passado. Murray veio ao país para ver de perto o fenômeno na Bienal do Livro do Rio, que termina hoje.

- Foi inspirador ver tantos jovens e crianças tão animados com os livros - disse.

Em entrevista por vídeo, o CEO falou da força do mercado cristão, do esforço para reduzir a dependência da China na impressão e por que considera a Amazon uma frenemy, mistura de amiga e inimiga.

Qual é a importância do Brasil para a estratégia global da HarperCollins?

Enorme. O Brasil é um país que cresce, com uma classe média ascendente e cada vez mais infraestrutura de distribuição. Crescemos mais em países emergentes do que em mercados já consolidados, como alguns países de língua inglesa. Quando viemos para o Brasil, o mercado vivia de altos e baixos, mas se fortaleceu. Estamos satisfeitos e vamos continuar investindo para ficar ainda mais fortes nos próximos dez, 20 anos.

Pensam em adquirir editoras brasileiras?

É possível. Nossa estratégia global é metade crescimento orgânico e metade baseada na compra de boas editoras. Procuramos editoras que façam o que ainda não fazemos, adicionando novas áreas à empresa.

Como vê a onda dos livros de colorir, impulsionada pela HarperCollins no Brasil?

Estávamos acompanhando atentamente, mas só percebi a magnitude do fenômeno chegando aqui. Anos atrás, os livros de colorir já haviam feito sucesso, mas agora isso é graças às redes sociais. Um fenômeno pode começar em qualquer lugar e alguns funcionam em várias línguas. É fascinante! Vamos falar com as sucursais da HarperCollins no mundo todo sobre os livros de colorir.

Amazon vai usar conteúdo do New York Times em suas plataformas de inteligência artificial A HarperCollins vai usar conteúdo do New York Times em suas plataformas de inteligência artificial

Continuação: 'Estamos pressionando governos para proteger direitos autorais', diz CEO da HarperCollins

perCollins Brasil é líder no segmento de livros evangélicos. Qual é a importância desse mercado?

Provavelmente somos a maior editora de livros religiosos do mundo. Nosso portfólio cristão nos abriu muitas portas na América Latina. Na Europa, foram os romances da Harlequin. O Brasil é o mercado onde a diversidade do catálogo da Harper Collins está melhor representada: ficção, não ficção, literatura infantojuvenil, religiosa.

Os audiolivros têm demorado a decolar no Brasil. Por quê?

O audiolivro é um desafio para alguns mercados, como o Reino Unido, a Europa e certamente o Brasil, porque falta catálogo. Nos Estados Unidos, produzimos audiolivros há 70 anos. Construímos um catálogo grande, primeiro em discos de vinil, depois em fitas cassete e CDs, que agora está sendo digitalizado. As editoras precisam investir mais no formato, o que é caro. Para criar um mercado, é preciso ter bastante conteúdo nas línguas locais e, idealmente, vários distribuidores.

Como as redes sociais mudaram o negócio do livro?

As redes sociais, principalmente o TikTok, ajudam a vender livro porque se dirigem diretamente aos leitores. Estamos aprendendo uma nova gramática de venda. Se um influenciador gosta de um livro nosso, é capaz de transformá-lo num best-seller. Sempre foi difícil e caro fazer um autor novo pegar. Hoje, graças ao algoritmo, esses livros encontram leitores que gostam de histórias parecidas.

Como observa os avanços da inteligência artificial?

Na HarperCollins valorizamos narrativas criadas por humanos, apoiamos quem conta e ilustra histórias, jamais vamos usar IA para gerar livros. Empresas de tecnologia acham que têm o direito de roubar trabalho criativo alheio para treinar suas IAs. A Meta estava usando livros disponíveis em sites piratas. Isso é as-

sustador!

Estamos pressionando os governos para proteger os **direitos** autorais. Só nos EUA, apoiamos cerca de 40 processos contra empresas que se apropriam de conteúdo sem autorização. Há quem diga que a IA é uma tsunami, mas olhando para trás eu vejo que os livros sobreviveram a todas as novas tecnologias. A IA também pode trazer oportunidades incríveis para as editoras e para os autores.

Quais oportunidades?

A tradução de livros e criação de audiolivros a um custo menor. Demora muito para traduzir um livro. A IA é boa para uma primeira tradução, que fica pronta em minutos, e depois é editada. É claro que grandes romances e livros importantes vão continuar sendo traduzidos por pessoas, mas há livros mais comerciais e de não ficção, colados no noticiário, que precisam estar disponíveis em várias línguas na mesma data. Não dá para esperar meses para publicar a tradução.

Os tradutores vão resistir.

Claro. Como disse, há livros que, por suas características ou a pedido dos autores, vão continuar nas mãos de tradutores. Mas há projetos em que é do interesse comercial do autor ter uma tradução mais rápida. O futuro é esse. É claro que haverá resistência, mas não vamos forçar ninguém a nada.

Por que a HarperCollins firmou um acordo autorizando o uso de seu catálogo para alimentar IA?

A maioria das empresas de tecnologia estava roubando nosso conteúdo. Quisemos estabelecer um precedente para mostrar o valor dos livros. Estávamos dispostos a um acordo, desde que pagassem e respeitassem os **direitos** autorais. Os autores puderam escolher se participariam ou não e encontramos uma empresa de tecnologia disposta a um acordo. Foi um precedente importante para o mercado editorial, por-

Continuação: 'Estamos pressionando governos para proteger direitos autorais', diz CEO da HarperCollins

que mostrou que os **direitos** autorais não travam a inovação.

A guerra tarifária desencadeada pelo presidente dos EUA, Donald Trump, afetou a indústria do livro?

Felizmente ainda não. A economia como um todo tem sido impactada pela incerteza e pelo aumento do custo de vida nos EUA. Ninguém sabe direito que tarifas estão em vigor, então é difícil tomar decisões. Há algum tempo estamos tentando ser menos dependentes da China para impressão, porque eles não estavam muito felizes com o conteúdo dos nossos livros. O preço chinês é melhor, mas há riscos e estamos firmando parcerias no mundo todo, inclusive no Brasil.

No Brasil, discute-se a limitação dos descontos de lançamentos para ajudar as livrarias na competição com a Amazon. É uma boa ideia?

A questão é complexa. Para o consumidor, o ideal é garantir a maior variedade possível de opções de compra em todos os canais de venda, físicos e digitais. Nos EUA, a pressão é sempre pelo menor preço, o que cria uma situação desafiadora para as livrarias. Apesar disso, as livrarias independentes cresceram nos últimos anos.

Quase metade do varejo de livros no Brasil é digital e dominado pela Amazon. Não é arriscado depender tanto de um único comprador?

Nos EUA é a mesma coisa. Chamo a Amazon de *fre-nemy* (mistura de amigo e inimigo, em inglês): é eficiente, mas se quiser pode acabar com a nossa indústria. Ainda assim, o mercado editorial brasileiro melhorou muito nos últimos anos por causa dela.

Em países como Brasil, Índia e México, a infraestrutura de distribuição era pouco desenvolvida, o que dificultava a expansão do negócio do livro. A venda on-line ajudou muito. Claro, é melhor não depender de um único player, e o poder público deve incentivar a entrada de novos concorrentes no mercado.

O que você gosta de ler?

Leio muita notícia e muita não ficção. Quase nunca tenho o luxo de ler um livro inteiro. Leio um monte de partes de livros (risos).

Plano brasileiro para turbinar IA ignora conceito básico da tecnologia



A versão final do Plano Brasileiro de Inteligência Artificial (PBIA) foi publicada na semana passada, quase um ano após a apresentação de seu esboço durante a 5ª Conferência Nacional de Ciência, Tecnologia e Inovação. Os avanços entre as duas versões são dignos de nota e reconhecimento da equipe envolvida. Temos agora um plano mais detalhado, que nos permite fazer perguntas e nos engajarmos em reflexões mais aprofundadas e interessantes. Uma dessas reflexões tem como ponto de partida as referências do texto à palavra linguagem.

A primeira delas aparece no terceiro dos quatro grandes objetivos previstos pelo plano: "desenvolver modelos de linguagem de grande escala (LLM) [do acrônimo em Inglês, Large Language Models] para inteligência artificial em português, baseados em dados nacionais". Para colocar os pingos nos Is, ou, nesse caso, nos Ls: a tradução de language nesse contexto não é linguagem, mas língua.

E, não, esse texto não é um lamento de um purista que toma por ofensa pessoal uma escolha tradutória infeliz --ainda que eu tenha me dedicado, nos últimos 15 anos, a construir datasets de treinamento e modelos computacionais para o Português Brasileiro. É uma tentativa de um linguista de sumarizar e didatizar aquilo que mais de um século de estudos em Linguística nos ensinou sobre como as línguas funcionam, o que elas são, porque elas não devem ser tomadas como sinônimos da capacidade cognitiva para a linguagem e porque isso importa no contexto

do PBIA.

Línguas humanas são um produto social da capacidade cognitiva para a linguagem. Isso significa que o fato de falarmos Português Brasileiro depende, em parte, da nossa configuração cerebral, desenvolvida ao longo da evolução da espécie, e, em parte, às condições sócio-históricas em que utilizamos essa capacidade para adquirir uma língua específica.

Línguas são, portanto, dispositivos culturais compartilhados entre os membros de uma comunidade e, como todo produto da cultura, elas permitem a expressão de ideias, valores e julgamentos que fazem sentido dentro daquela comunidade. A capacidade cognitiva para a linguagem, portanto, comporta um aspecto mais transcultural, mais genericamente aplicável a todo e qualquer ser humano, porque está ancorada em aspectos biológicos. A aplicação dessa capacidade na aquisição de uma língua específica a ancora em um momento histórico, no qual emergem os valores que uma sociedade tem. Uma língua específica também permite aos membros dessa comunidade discutirem esses valores, engajarem-se no debate democrático e, eventualmente, propor novos valores que sejam menos excludentes ou que contemplem uma diversidade maior de subgrupos dessa sociedade.

E o que isso tem a ver com o PBIA?

Tudo. A centralidade dos LLMs, ou Grandes Modelos de Língua, no debate atual sobre IA é inescapável. Eles estão na base de uma infinidade de aplicações de IA e são, de forma esperada, citados em diversas das ações estratégicas previstas no plano: do chatbot para melhoria do acesso de idosos à saúde ao sistema de fiscalização para a Receita Federal. LLMs são, em si mesmos, o objeto de uma das ações do plano, intitulada IA baseada em dados nacionais (LLM

Continuação: Plano brasileiro para turbinar IA ignora conceito básico da tecnologia

em português). A ação prevê apoio para a curadoria de datasets nacionais para o desenvolvimento de LLMs para o português brasileiro e lista como desafio "criar e aprimorar bases de dados nacionais para treinamento de IA, reduzindo a dependência externa e contemplando a diversidade e as especificidades do Brasil".

Não fica claramente delineado no plano o que se toma por curadoria e por dependência externa, mas é possível ainda assim discutir ambas as questões. Primeiramente, cabe pontuar que qualquer modelo de língua, para ser treinado, requer curadoria de dados. O que se discute, portanto, não é a existência da curadoria, mas, sim, a sua extensão e profundidade. O simples fato de escrever um script que navegue por sites terminados em ".br" coletando todo e qualquer texto --independentemente do conteúdo ou de questões de **direito** autoral-- para ser usado no treinamento de um modelo é curadoria. Trata-se de uma curadoria responsável? Certamente não, mas é uma escolha humana que terá impactos no resultado do modelo. Isso é curadoria de dados.

No extremo oposto, tem-se iniciativas de curadoria de dados que buscam não só escolher os dados coletados, respeitando **direitos** autorais e filtrando conteúdos ofensivos, mas também anotar esses dados com metadados (etiquetas) que permitam um treinamento mais eficiente. Essa noção de curadoria requer que se entenda o segundo L da sigla LLM como língua e não como linguagem. Porque a curadoria se dá a partir de uma língua específica que representa uma cultura específica, no caso, a brasileira. Ler o L como linguagem abre caminho, por exemplo, para adotar como solução técnicas de transferência de aprendizagem, em que aquilo que um modelo aprendeu com base em dados de uma língua seja transferido para outra para a qual se disponibilizam menos dados de treinamento. Ignorar a distinção entre língua e linguagem, portanto, compromete a formulação adequada de estratégias de curadoria dos dados que serão usados no treinamento de LLMs.

E aqui o termo dependência externa entra em cena. Em grandes modelos multilíngues é muito difícil escapar da transferência de aprendizagem. E, para além da dependência externa, usar dados de outras línguas traz outros graves danos. A título de exemplo, considere-se o caso em que preconceitos e enviesamentos negativos de outra cultura podem acabar contaminando LLMs desenvolvidos para o Brasil.

Em um trabalho recente de um consórcio de pesquisadores, construiu-se um dataset para avaliação de LLMs quanto à propagação de preconceitos: o SHADES. O dataset cobre 16 línguas de diversas regiões do mundo, inclusive o Brasil, e permite avaliar se os modelos reproduzem preconceitos de raça, gênero, status social, profissão, idade e nacionalidade em larga escala. Os achados, reportados em um artigo apresentado em maio em uma prestigiosa conferência internacional, mostram não só que diversos modelos multilíngues propagam preconceitos comuns na cultura representada por aquela língua, como também propagam aqueles que só são comuns em outras culturas totalmente diversas.

E não é só isso. O CVQA, um dataset construído por outro consórcio internacional de pesquisadores, linguistas entre eles, mostrou as limitações de grande modelos multimodais para realizar tarefas que requerem conhecimento culturalmente ancorado. LLMs, como são chamados, trabalham não só com línguas mas também com imagens.

O dataset é composto por centenas de perguntas que só podem ser respondidas se o modelo interpretar corretamente a imagem que acompanha o texto. No caso da cultura brasileira, o CVQA tem perguntas sobre como as pessoas comem paçoca de rolha, ou sobre o que a cultura popular pensa que pode ocorrer se um chinelo de borracha for deixado de cabeça para baixo no chão.

Isso significa que não há o que ser feito? Pelo contrário: isso significa que o PBIA está certo em as-

Continuação: Plano brasileiro para turbinar IA ignora conceito básico da tecnologia

sociar à meta de criação de uma IA nacional fundacional soberana a curadoria de dados brasileiros. Mas é preciso definir que o que se quer é uma curadoria humana responsável, que possa gerar IAs igualmente responsáveis e explicáveis, capazes de representar a diversidade linguística e cultural do Brasil.

Para isso, será inescapável trazer para o processo quem entende da natureza dos dados a serem curados: os linguistas, aqueles que são especialistas em como a capacidade da linguagem se operacionaliza no processo de aquisição e uso de uma ou mais línguas. Aqueles que são especialistas também em lidar com a complexidade das línguas humanas: suas variações, registros, ambiguidades, vaguezas, propriedades formais e funcionais. Linguistas dominam as ferramentas conceituais e metodológicas para assegurar que um LLM represente, de forma coerente e inclusiva, a pluralidade linguística do Brasil. E, infelizmente, esse grupo de especialistas e todo o conhecimento por eles acumulado têm sido deixados à margem das discussões do PBIA. Tomara que não o

sejam também quando da sua operacionalização.

*Tiago Torrent é professor da Universidade Federal de Juiz de Fora, onde coordena o Laboratório FrameNet Brasil de Linguística Computacional, o Programa de Pós-Graduação em Linguística e a ReINVenTA: Research and Innovation Network for Vision and Text Analysis. Possui doutorado em Linguística pela UFRJ e foi professor visitante do Departamento de Sueco, Multilingualidade e Tecnologia Linguística da Göteborgs Universitet. É bolsista de Produtividade em Pesquisa do CNPq.

Opinião

Texto em que o autor apresenta e defende suas ideias e opiniões, a partir da interpretação de fatos e dados.

Este texto não reflete, necessariamente, a opinião do UOL

Índice remissivo de assuntos

Direitos Autorais
3, 6