Algoritmos de IA infringem direitos autorais? – Parte 1

CARLOS ALBERTO DOS SANTOS
Professor aposentado do Instituto de Física da UFRGS e professor visitante da UFERSA

Nos últimos meses, tem havido um intenso debate na imprensa internacional, com imediata repercussão na imprensa nacional, a respeito da suposta infringência dos direitos autorais por parte de algoritmos de inteligência artificial (IA). Há ações judiciais em vários setores, incluindo a indústria fonográfica, mas quero aqui me deter nos chatbots, entre os quais o ChatGPT é o que tem tido maior visibilidade. Por absoluta falta de conhecimento jurídico, não posso entrar nos detalhes legais que os advogados ressaltam em seus processos, razão pela qual fixarei meu olhar exclusivamente na alegada capacidade do ChatGPT produzir documentos plagiados de material disponível na internet. De fato, o algoritmo tem essa capacidade, mas seus programadores têm conseguido evitar a produção de documentos idênticos àqueles que ele consulta para gerar seu texto. Se você ainda não sabe como funciona o ChatGPT, veja esse artigo na TN (http://www.tribunadonorte.com.br/noticia/sera-paradoxal-a-invena-a-o-do-chatgpt/555121).

Em um artigo publicado em maio último (https://pike.psu.edu/publications/www23.pdf), pesquisadores da Penn State University (EUA) afirmaram que o ChatGPT faz plágio muito mais sofisticado que simplesmente copiar e colar. No artigo eles adotam a seguinte definição de três tipos de plágio: (1) Plágio literal: cópias exatas de palavras ou frases sem transformação; (2) Plágio de paráfrase: substituição de sinônimos, reordenamento de palavras e/ou retrotradução; (3) Plágio de ideias: representação de conteúdo central de forma alongada.

> Assim como os verificadores de plágio falham na sua missão, fiz alguns testes que mostram que o GPTZero falha na identificação da origem de textos. Mostrarei no próximo artigo o procedimento que usei para verificar essa falha do GPTZero."

Ora, ora, ora, quem produz conhecimento, não importa a área, sabe muito bem que, como disse Isaac Newton, tem que se apoiar em ombros de gigantes. Foi assim que ele justificou sua extraordinária e relevante produção científica. Portanto, "plagiando" Newton podemos dizer que apenas os dois primeiros tipos constituem plágio, e mesmo assim podem ser admissíveis quando a fonte é citada. É o que se chama de citação direta (primeiro caso), e citação indireta (segundo caso). Claro que um trabalho feito inteiramente de citações diretas ou indiretas não é aceitável. Essas citações, que eventualmente são indispensáveis, deve ser usadas parcimoniosamente. Finalmente, se não existisse o terceiro tipo, o conhecimento não avançaria.

Atualmente, a forma simples e rápida de detectar o plágio mais grosseiro, do tipo 1, consiste em copiar trechos da obra e colar no Google. Se o trecho for colocado "entre aspas", o Google identifica os locais na Internet onde ele se encontra ipsis litteris. Para detectar formas mais sutis de plágio existem muitas ferramentas disponíveis na Internet, de um modo geral muito deficientes. Como os chatbots não fazem plágio grosseiro, eles apenas produzem "plágios do tipo 3", que eu não considero plágio, a moda recente são ferramentas que supostamente identificam se determinado texto foi escrito por um ser humano ou por um algoritmo de IA. Entre esses algoritmos, talvez o mais proeminente, no momento, seja o GPTZero. O algoritmo não afirma que o texto analisado foi produzido por um ser humano ou por uma ferramenta de IA. Ele informa que "é provável que seu texto seja escrito inteiramente por IA", ou que "é provável que seu texto seja escrito inteiramente por um humano". O nível de probabilidade é calculado pelo GPTZero por meio de uma variável denominada perplexidade, um conceito usado na teoria da informação, e que mede a incerteza de prever a próxima palavra em uma sequência. Portanto, quanto maior for o índice de perplexidade, maior a probabilidade de o texto ter sido escrito por um humano, porque os textos de humanos são mais imprevisíveis do que aqueles gerados por um algoritmo de IA.

Assim como os verificadores de plágio falham na sua missão, fiz alguns testes que mostram que o GPTZero falha na identificação da origem de textos. Mostrarei no próximo artigo o procedimento que usei para verificar essa falha do GPTZero, que adianto aqui resumidamente. Usei textos (em português e também traduzidos para o inglês) que publiquei sobre a descoberta dos raios-X e textos produzidos pelo ChatGPT sobre o mesmo assunto. Os textos em português foram interpretados pelo GPTZero como sendo produzidos por humanos, enquanto as versões em inglês tiveram resultados diferentes. O do algoritmo foi interpretado como sendo de IA, e o meu como produzido por humano.