Campos e Badaró: O uso de IAs generativas no setor público

No final de 2022, a empresa OpenAI lançou o *Chat GPT*, modelo de linguagem natural em larga escala projetado para gerar respostas sobre diversos assuntos e em diferentes contextos.

A ferramenta é treinada com base um imenso conjunto de dados de páginas da web e, a partir de inferências e probabilidades, é capaz de responder perguntas, traduzir e resumir textos, gerar códigos, criar poemas, dentre outros, de forma muito semelhante às produções humanas, constituindo o grupo das chamadas inteligências artificiais (IAs) generativas — aquelas que podem criar novos dados com base em padrões e estruturas aprendidas a partir de dados já existentes.

Tamanha novidade foi acompanhada por sua enorme popularização — vide a quantidade recorde de acessos em apenas alguns meses — bem como de receios e controvérsias sobre seu uso em diversos setores.

Fruto dessas preocupações foi a recente atuação da autoridade de proteção de dados italiana, que impôs uma limitação temporária ao processamento de dados, pela OpenAI, de usuários residentes no país e determinou que a empresa comunicasse as iniciativas realizadas para cumprir as previsões do Regulamento Geral de Proteção de Dados europeu.

A subsequente decisão da empresa de bloquear o acesso aos serviços levou o *European Data Protection Board* a lançar uma força-tarefa dedicada exclusivamente ao *ChatGPT*, com foco na cooperação e na troca de informações sobre ações conduzidas pelas autoridades nacionais de proteção de dados. Além disso, as IAs generativas têm sido objeto de ações judiciais, nos Estados Unidos e na Europa, que discutem possíveis violações de direitos autorais e a formulação de alegações difamatórias pela ferramenta.

No Brasil, o uso do *ChatGPT* será, em breve, objeto de análise pelo Conselho Nacional do Ministério Público (CNMP), a partir de uma representação que pautou o uso de ferramentas de inteligência artificial especificamente por promotores e procuradores na elaboração de petições.

A discussão foi levantada pelo advogado Fábio de Oliveira Ribeiro, que também recorreu ao Conselho Nacional de Justiça (CNJ) para impedir que juízes utilizassem a inteligência artificial em decisões. O advogado chamou a atenção, nos últimos dias, por ter sido condenado por litigância de má-fé pelo Tribunal Superior Eleitoral (TSE) ao protocolar uma petição redigida pelo *ChatGPT* para tentar ser admitido como amicus curiae em processo que envolve o ex-presidente Jair Bolsonaro. O autor da peça teria se aproveitado da grande visibilidade do caso para chamar atenção para os perigos do uso de ferramentas de IA no Direito.

Freepik


Freepik

Desde o lançamento do *ChatGPT*, diversas outras notícias têm tratado de seu uso na área jurídica. Em fevereiro deste ano, por exemplo, um juiz colombiano redigiu a primeira sentença com o uso de inteligência artificial no país, em um caso sobre o direito à saúde de uma criança autista, detalhando, na sentença, as perguntas realizadas ao bot e as respostas obtidas.

Na Índia, um juiz seguiu o mesmo caminho para decidir acerca da fiança em um processo criminal e, com base em respostas oferecidas pelo bot, decidiu por sua não concessão. Nos Estados Unidos, um advogado admitiu ter utilizado o *ChatGPT* na defesa de seu cliente, a fim de encontrar julgados semelhantes para fundamentar seu pedido em um processo contra uma companhia aérea. Nesse caso específico, as respostas oferecidas pela IA eram compostas por jurisprudências fictícias, o que foi percebido tanto pelo juiz do caso quanto pela parte contrária ao tentarem encontrar registros dos processos mencionados. Esses e outros casos trazem diversas implicações éticas e jurídicas, sobre as quais este artigo pretende fazer algumas considerações.

De início, podem ser apontados os problemas de confiabilidade nas ferramentas de IA generativa. No caso do *ChatGPT*, como a própria empresa destaca em sua página, a ferramenta pode apresentar respostas incorretas, falsas ou imprecisas, apesar de construções semânticas aparentemente corretas — é o caso das chamadas "alucinações" (*hallucinations*), isto é, percepções irreais que parecem reais, como ocorrido na lista de precedentes utilizada pelo advogado estadunidense.

Isso porque a qualidade do *output* depende de vários fatores, dentre eles o conjunto de dados em que se baseia, os inputs do usuário e outros aspectos envolvendo os métodos de treinamento. Assim, as alucinações podem ser fruto tanto da insuficiência dos dados de treinamento, especialmente em ramos muito especializados, como o jurídico, quanto da qualidade dos comandos fornecidos pelo usuário.

No mesmo sentido, considerando a grande quantidade e diversidade dos dados de treinamento, o design e o desenvolvimento dos *chatbots* podem resultar em ferramentas que absorvam diferentes vieses, que podem incluir questões culturais e linguísticas, raciais e de gênero, vieses cognitivos, vieses de confirmação, dentre outros, com riscos de reforço de estereótipos e preconceitos já presentes na sociedade. Esse tópico é especialmente relevante em áreas sensíveis, como a do direito, que tem o condão de tomar decisões que afetam diretamente a vida das pessoas envolvidas.

A busca pela melhor solução jurídica em um determinado caso concreto passa, como sabido, por análises

da situação fática apresentada e do complexo arcabouço legal vigente no país, mas também de questões contextuais ou potenciais exceções a regras. Além disso, requer, muitas vezes, o uso de técnicas interpretativas e métodos de ponderação específicos, dos quais tecnologias como o *ChatGPT* não conseguem dar conta.

É fundamental, nesse cenário, compreender o exato funcionamento dessa e de outras ferramentas semelhantes que, apesar do potencial para auxiliar em tarefas jurídicas, não têm (ou não deveriam ter) o condão de substituir o fator humano. Por serem grandes redes neurais que meramente preveem o próximo token em uma sequência com base em uma lógica de probabilidade, os outputs formulados requerem sempre a avaliação posterior de sua adequação e pertinência diante dos elementos fáticos e legais que se apresentam em determinado caso concreto.

Um outro ponto a ser considerado diz respeito à proteção de dados dos usuários. Em uma perspectiva mais ampla, o *ChatGPT* traz preocupações quanto ao cumprimento de diversas previsões legais não só da LGPD brasileira, mas também de outras legislações de proteção de dados, como já tratado em outras ocasiões.

No contexto dos usos aqui abordados, porém, destaca-se a questão do armazenamento e do compartilhamento dos dados, bem como dos inputs fornecidos à ferramenta, que podem incluir dados pessoais (inclusive dados sensíveis) de terceiros — isto é, daqueles envolvidos nos casos concretos para os quais profissionais do Direito pretendem buscar auxílio. Isso porque os dados compartilhados com a ferramenta podem ser usados para seu treinamento e aprimoramento, colocando em risco o sigilo das informações, inclusive daquelas cujos processos correm em segredo de justiça.

Em 25 de abril, a OpenAI fez uma publicação informando a disponibilização de uma nova configuração, na qual é possível desativar o histórico do bate-papo com o *ChatGPT* e, nesses casos, as informações ali inseridas não serão utilizadas para melhoria do sistema ou para a oferta de serviços.

Ainda serão, no entanto, retidas por 30 dias para fins de revisão caso necessário; após o período, serão permanentemente deletadas. A alteração se mostra importante, mas ainda insuficiente para garantir a devida proteção das variadas informações constantes de processos judiciais, especialmente considerando-se o episódio de março deste ano, quando o *ChatGPT* sofreu a primeira grande violação de dados pessoais, quando foram expostas informações pessoais de parte dos assinantes da versão plus.

Uma alternativa seria a utilização do *ChatGPT* por meio do consumo de suas APIs (application programming interface), cuja política de uso de dados prevê a não utilização de dados fornecidos pelos usuários finais para treinar os modelos como regra, sendo o compartilhamento uma opção disponibilizada ao usuário.

Diante dos desafios colocados pela introdução de IAs generativas em diversos setores, reguladores em diferentes países têm se movimentado para estabelecer um ambiente que impulsione a inovação ao mesmo tempo em que proteja os direitos dos utilizadores dessas ferramentas. Na Europa, está em fase de discussão a proposta do *AI Act*, que busca fortalecer regras voltadas à qualidade dos dados, à responsabilidade e à transparência dos sistemas, com uma abordagem que prevê obrigações proporcionais ao nível de risco apresentado pela IA.

Na última versão do texto, foram incluídas previsões específicas aos provedores de IAs generativas, como no caso do ChatGPT. No Reino Unido, ao contrário, de acordo com o *white paper* publicado em março, será adotada uma abordagem flexível, com o fortalecimento e a capacitação de autoridades reguladoras já existentes para facilitar o uso seguro das IAs no país.

No Brasil, está em tramitação o PL 2.338/23, que propõe a criação de um marco regulatório para a inteligência artificial no contexto nacional, seguindo a linha proposta pela União Europeia.

No caso específico do uso por procuradores e promotores do Ministério Público — conforme será analisado pelo conselheiro Rodrigo Badaró — deve-se ter em mente o potencial dessas tecnologias para auxiliar em diversas tarefas, especialmente as rotineiras, trazendo mais eficiência e agilidade ao exercício das funções.

No entanto, diante das preocupações aqui apontadas, é necessário que sejam pensadas formas de se utilizar a ferramenta de maneira responsável e em observância aos direitos de terceiros, buscando evitar, como resultados, o tratamento desigual de indivíduos, decisões injustas/incorretas ou o descumprimento dos deveres legais impostos aos membros do MP, conforme previsão do art. 43 da Lei n. 8.625/1993.

Sugere-se, assim, a realização de análise de risco sobre as formas pelas quais ferramentas como o *ChatGPT* podem ser utilizadas nesse contexto institucional específico e, a partir disso, a criação de políticas que apresentem diretrizes e orientações para o uso, levando em consideração, em especial: a natureza acessória das IAs generativas, que deverá ser utilizada apenas para atividades-meio, de modo que não haja transferência do poder decisório à ferramenta; a constante necessidade de supervisão humana dos *outputs*, em atenção a informações incorretas, prejudiciais ou tendenciosas; a importância da anonimização dos dados inseridos no sistema, como forma de resguardar dados pessoais e outros que possam identificar os processos e a necessidade de se optar pelo não compartilhamento dos dados com a plataforma para fins de melhoria do sistema.

Outro aspecto importante diz respeito à transparência e *accountability*. Nesse sentido, sugere-se que se estabeleça a necessidade de indicação, quando for o caso, acerca de qual tecnologia foi utilizada e em que medida. Um exemplo seria apontar, em um documento jurídico, quais trechos contaram com o auxílio de IA. Além disso, é importante que se adote um sistema de certificação no qual os profissionais atestem não ter havido compartilhamento de informações confidenciais e dados pessoais de terceiros.

Iniciativas nesse sentido já vêm sendo adotadas, a título ilustrativo, por renomadas universidades do mundo: Harvard e Oxford já reconheceram o potencial complementar — e não substitutivo — do *ChatGPT* para o aprendizado, ao passo que universidades como Cambrigde anunciaram que irão admitir a utilização da ferramenta para determinadas atividades, desde que haja transparência quanto a seu uso.

Embora reconheça-se não serem soluções definitivas, tais medidas parecem ser um importante ponto de partida para uma melhor governança do uso de uma tecnologia que não aparenta ser uma tendência passageira.

Meta Fields