

Pironti e Zarur: ChatGPT e o “compliance by design”

06/06/2023

No imaginário humano, o risco de uma criação se voltar contra o criador é alvo de diversas fábulas, contos e filmes, alcançando seu ápice no clássico gótico *Frankenstein*. Nesta obra literária de 1818, a escritora Mary Shelley conta a história do jovem cientista Victor Frankenstein que, por meio de métodos controversos, dá a vida a uma criatura feita de partes de corpos humanos. No entanto, o cientista se arrepende amargamente de sua criação, após ter sua vida e de seus entes queridos ameaçados pelo monstro cadavérico.



Este mesmo arquétipo literário se vê remontado não

somente na ficção, mas em diversos eventos históricos ao decorrer dos séculos, desde a amargura de Santos Dumont ao ver sua maior invenção, o avião motorizado, utilizada como um objeto de guerra, até Alfred Nobel, o inventor da dinamite que, arrependido de sua criação, doou a maior parte de sua fortuna para criação da famosa condecoração, o Prêmio Nobel. Estes fatos demonstram o medo no avanço descontrolado da ciência, principalmente no receio de que as novas criações potencializem a crueldade humana ou gerem novas ameaças, antes impensadas.

Novamente, este mesmo paradigma vê-se atualizado pela criação de Sam Altman, CEO da OpenAi, empresa responsável pela criação do modelo de linguagem de inteligência artificial, o *ChatGPT*. Lançada em novembro de 2022, a ferramenta conquistou uma difícil tarefa, impressionar um mundo já saturado por novidades e novas tecnologias.

Com a finalidade de entender e reproduzir a linguagem natural humana de forma automatizada, a tecnologia permite a conversação entre homem e máquina de forma orgânica e interativa, abrindo margens para as mais diversas aplicações, até mesmo em ambientes que ainda não haviam sido impactados pelo uso deliberado das inteligências artificiais, como o meio literário e publicitário.

Apesar de tais tecnologias, por si só, não representarem grande novidade, o impacto causado pelo *ChatGPT* se deu pela junção massiva de cada uma destas ferramentas, possibilitando o funcionamento à habilidade de geração de textos coerentes e semânticos, que podem ser produzidos com poucos comandos, ou em termos técnicos, *prompts* ou *inputs*.

Ainda, outro grande diferencial se deu pela capacidade da tecnologia manter um diálogo contínuo, apresentando respostas com base nas linhas anteriores da conversa, isto é, suas respostas tomarão como base não somente seu banco de dados e a pergunta realizada, mas sim todo o histórico do atual diálogo mantido entre o usuário e a plataforma.

Com tamanho poder de processamento de linguagem natural, esta ferramenta apresenta diversas ou, até mesmo, ilimitadas aplicações, uma vez que ainda não é possível definir seu potencial final. Dentre as atuais aplicabilidades do *Chat GPT*, se destaca a utilização em diversos setores da indústria da tecnologia, auxiliando na criação de linhas de código, até no ensino de programação à iniciantes na área. A ferramenta pode ser aplicada para criação de textos acadêmicos, científicos e até roteiros de filmes e publicações em redes sociais. No âmbito jurídico, o *Chat* já é aplicado na pesquisa, análise e resumo de jurisprudências, revisão de contratos e identificação de possíveis riscos de *compliance*.

Apesar do inegável impacto econômico e social causado pelo *ChatGPT*, como qualquer ferramenta de tecnologia, a plataforma apresenta algumas limitações que precisam ser observadas, antes do uso desmedido da ferramenta. Ao questionar diretamente o Chat sobre este tópico, a ferramenta relatou que, em virtude da necessidade de uma grande quantidade de dados para seu treinamento, suas respostas podem apresentar traços de viés ou até mesmo preconceitos.

Este fato se dá em razão de que o material base do conhecimento da ferramenta se limita a setembro de 2021, apresentando diversos posicionamentos e orientações políticas, religiosas e sociais, podendo enviesar a resposta apresentada pelo Chat. Tal situação pode ser observada em diversos casos práticos, por exemplo, em que o modelo de linguagem foi considerado tendencioso ao apresentar um comportamento contra a pauta conservadora americana o ex-presidente, Donald Trump [1]. A problemática da apresentação de respostas parciais ou partidárias pode ser agravado pelo fato de que, por vezes, a ferramenta apresenta erros em seus resultados, ou seja, respostas com fatos equivocados ou inexistentes.

Este risco necessita de especial atenção, visto que, ao apresentar respostas às mais diversas perguntas de forma simples, natural e didática, a ferramenta promove uma grande credibilidade, em virtude da maneira certa que suas exposições são entregues aos usuários. Deste modo, ao escrever com tamanha confiança, mesmo as respostas mais equivocadas podem soar como verdades indubitáveis, como no caso exposto pela *BBC Brasil* [2], em que, após perguntado, teria respondido que o Brasil já foi ganhador de, pelo menos, cinco prêmios Oscar, apresentando os filmes e anos em que foram lançados.

Mesmo que o país nunca tenha sido vitorioso na premiação de melhor longa-metragem, ao dispor de tantas informações específicas, as chances de que os usuários confiem nestas afirmações é grande. Tal situação deve ser observada, em particular, por aqueles que utilizam

da ferramenta para sua atividade profissional. Em análise ao cenário jurídico brasileiro, a utilização do *ChatGPT* pode resultar na aplicação de jurisprudências inexistentes ou citações de doutrinas nunca escritas que, apesar de muito bem escritas, não apresentam fundamento na realidade.

Apesar de suas limitações, a ferramenta ganha especial destaque quando utilizada para o fim ao qual foi criada, isto é, a geração de textos de forma natural. Neste sentido, ao ser aplicada no resumo de textos, reescrita de frases, construção de argumentações, estruturação de petições e análise de informações, o *ChatGPT* expõe seu real valor prático. Assim, pode-se concluir que, como qualquer ferramenta, o usuário deve conhecer a forma de extrair os melhores resultados, bem como aplicar estratégias para evitar erros e falhas na ferramenta utilizada.

Feita esta breve contextualização, é de grande utilidade analisar as aplicabilidades do *ChatGPT* em matérias correlatas ao direito, como o *compliance*. Em suas próprias palavras, quando perguntado sobre as utilidades da ferramenta no *compliance*, o Chat declarou que poderia ser aplicado nas seguintes áreas:

- *Análise de Dados*: A ferramenta apresenta a capacidade de analisar grandes volumes de informações, podendo identificar padrões, tendências e constatações relevantes para uma empresa, como riscos de corrupção, fraudes e desconformidades às normativas internas;
- *Elaboração de Políticas e Procedimentos*: Como modelo de linguagem natural, o ChatGPT pode estruturar documentos de *compliance*, como políticas e procedimentos, tomando como base às melhores práticas e regulamentos aplicáveis;
- *Treinamentos*: A ferramenta pode ser utilizada na estruturação, revisão e até aplicação de treinamentos e capacitações de *compliance*, como a elaboração de uma plataforma de "Chatbot", em que o ChatGPT pode fornecer uma série de respostas as perguntas dos colaboradores, tomando como base às políticas e práticas internas adotadas pela empresa; e
- *Comunicações Internas*: Ao utilizar sua capacidade analítica, o ChatGPT pode monitorar as comunicações internas de uma empresa, analisando e-mails, mensagens e registros, visando identificar palavras-chave e comportamentos padrão relacionados a práticas indevidas ou ilícitas; e
- *Previsão de Riscos*: Ao ser apresentada a uma base de dados que contenha o histórico de riscos de *compliance* enfrentados pelo setor atuante de uma empresa, o ChatGPT pode prever os possíveis cenários de risco, bem como apresentar as possíveis melhores saídas e planos de ação para prevenção ou mitigação destes eventos incertos, auxiliando o corpo diretivo destas instituições na tomada de decisões.

Como pôde ser observado, a ferramenta apresenta maior eficácia quando utilizada para a identificação de padrões, estruturação e organização de dados e a previsão de possíveis cenários, tomando como base o histórico apresentado à plataforma. Assim, ao ser treinada de forma correta, esta inteligência artificial aumenta exponencialmente a eficiência e precisão da análise de bases de dados que, mesmo que já possam ser avaliadas automaticamente, agora passam a ser analisadas, organizadas e classificadas de acordo com os padrões e critérios

elencados pelo usuário.

Ainda, com a automação de análises complexas, a tendência de erros humanos e a necessidade da contratação de vários profissionais especializados é minorada, reduzindo os custos operacionais das empresas aderentes à tecnologia. Esta mesma capacidade analítica do *ChatGPT* permite que eventos de risco ou falhas internas sejam identificados de forma mais rápida e eficiente, aumentando o tempo hábil para correção destes problemas, prevenindo perdas financeiras ou danos reputacionais.

Com a aplicação constante da ferramenta, sua base de dados será cada vez mais robusta e personalizada à realidade da empresa, permitindo que o Chat forneça informações e dados referentes à gestão do *compliance*, apresentando possíveis melhorias específicas ao Sistema de Integridade da instituição e, caso seja alimentada com as novas regulamentações e práticas, a ferramenta poderá indicar a melhor forma de aplicar adequações regulatórias.

Apesar de tantas funções, a própria plataforma é categórica em afirmar que sua aplicabilidade não pode substituir completamente o trabalho humano nas práticas de *compliance*, mas sim deve ser utilizada como uma ferramenta para o aumento de eficiência, precisão e rapidez nas atividades executadas por esta área.

Neste sentido, é necessário avaliar novamente as limitações operacionais do *ChatGPT*, agora tomando como base o contexto de sua utilização junto ao *compliance*. Dentro deste cenário, uma das maiores barreiras enfrentadas não diz respeito à ferramenta por si só, mas à qualidade e precisão dos dados fornecidos para o *ChatGPT*. Assim, como as demais inteligências artificiais baseadas em *machine learning*, a eficiência das respostas dadas pelo Chat está diretamente relacionada com a dependência da qualidade das bases de dados aos quais foi treinada e exposta.

Deste modo, os agentes de *compliance* deverão estar atentos ao envio de informações imprecisas ou incompletas, uma vez que, ao apresentar tais dados à ferramenta, os resultados estarão comprometidos. O risco neste caso não se atém somente no prejuízo da qualidade das respostas, mas sim no viés apresentado pela plataforma que, ao ser exposta a dados limitados, parciais ou partidários, poderá fornecer orientações imprecisas, preconceituosas ou tendenciosas.

Para maior exploração deste tópico é de grande valia analisar um caso prático hipotético, vejamos:

Uma empresa de tecidos decide promover uma análise dos setores de sua operação que apresentam maiores riscos de assédio moral, possibilitando a aplicação de atividades específicas destinadas aos departamentos que expressem maior vulnerabilidade. Para tanto, a instituição fornece ao *ChatGPT* o histórico de denúncias apuradas pelo *compliance*, apresentando o número de ocorrências, setores envolvidos e medidas disciplinares aplicadas. Após devidamente alimentada com as informações necessárias, a ferramenta de inteligência artificial aponta que o departamento operacional é o que necessita de maior atenção durante os treinamentos e ações.

Em uma primeira análise o caso hipotético não expõe nenhuma problemática aparente, no entanto, o risco de enviesamento pode ser silencioso e quase imperceptível. Na situação descrita acima, o setor realmente problemático da empresa de tecidos é o departamento de vendas, que sofre diversas pressões e casos de assédio moral por parte da liderança.

O "erro" do *ChatGPT* no cenário imaginado ocorreu em virtude da baixa qualidade dos dados fornecidos, uma vez que a empresa divulgou o canal de denúncias com maior força no setor operacional, por imaginar que seria o local mais vulnerável a práticas repressivas de líderes e coordenadores. Em razão desta massiva divulgação, o setor foi o que mais realizou denúncias quando comparado às demais áreas da empresa.

Por consequência, o resultado de que o departamento operacional é o mais problemático foi apenas um reflexo das informações incompletas fornecidas pela empresa. Ou seja, o risco apresentado neste caso não está na ferramenta, mas sim no usuário, podendo direcionar incorretamente a empresa na tomada de decisões.

Assim como a base de dados, as perguntas destinadas ao Chat necessitam de um grande cuidado, ao passo que sua especificação e objetividade guiará por completo a resposta fornecida pela ferramenta. Deste modo, as perguntas, ou *prompts*, deverão ser claras, completas e detalhadas, apresentando o contexto, objetivo, formato e necessidades que deverão ser atendidas na resposta, descrevendo as informações que a plataforma deverá "observar" ao elaborar a resposta.

Mesmo que a empresa possua colaboradores capacitados e familiarizados com este tipo de ferramenta, para que o *ChatGPT* seja integralizado à outras plataformas, como sistemas internos, será necessário o auxílio de um especialista em programação, possibilitando a implementação da tecnologia nas operações da empresa.

Outro desafio que os profissionais de *compliance* deverão estar preparados para lidar é a dificuldade que o *ChatGPT* apresenta na interpretação de contextos culturais, regionalismos, fatores sociais, políticos ou econômicos nas análises realizadas pela ferramenta. Estas limitações podem prejudicar a avaliação de riscos e identificação de inconformidades promovidas pelo Chat, vez que circunstâncias culturais não serão observadas ou, caso sejam observadas, dependerão das informações inseridas pelo usuário.

Em virtude dos dados analisados pela ferramenta ainda estarem restritos aos acontecimentos ocorridos até setembro de 2021, diversas nuances contextuais não serão avaliadas. O que pode ser agravado pelo fato de que os contextos específicos (como, por exemplo, condição financeira, realidade pessoal e familiar de um denunciado por fraude) também não serão tomados como base caso não sejam apresentados pelo usuário, de forma detalhada e imparcial, à plataforma.

Não há como não se falar de tecnologias e análises de informações sem citar os riscos referentes à privacidade e segurança da informação. Ao utilizar ferramentas como o *ChatGPT*, todos os dados coletados e fornecidos à ferramenta deverão ser precedidos de processos que garantam a devida segurança e conformidade aos padrões estabelecidos pela Lei Geral de Proteção de Dados (Lei nº 13.709/2018).

Em um primeiro momento, os cuidados referentes à privacidade e segurança dos dados pessoais e empresariais deverão partir da própria instituição, ou seja, antes da utilização de qualquer ferramenta, a empresa deverá adotar, por padrão, sistemas e procedimentos que assegurem o adequado tratamento aos dados em sua posse, garantindo o respeito necessário em todas as etapas da utilização destas informações, isto é, desde o momento da coleta, até o armazenamento e posterior descarte.

Já em uma segunda etapa, a análise de segurança e privacidade deverá se voltar ao próprio *ChatGPT*, vez que a inclusão de dados pessoais e informações relacionadas à empresa em uma plataforma terceira poderá colocar em risco o tratamento adequado destes dados. A respeito deste ponto, ao ser perguntado sobre a utilização dos dados inseridos em perguntas de usuários, o *ChatGPT* alega que:

"Essas interações são anônimas e os dados são usados apenas para fins de treinamento e desenvolvimento. As informações são processadas em grande escala e misturadas com as de muitos outros usuários para garantir que os dados pessoais não sejam identificáveis e que sua privacidade seja protegida. É importante lembrar que, apesar de eu usar suas perguntas para melhorar meu treinamento, eu não as armazeno nem as compartilho com terceiros."

(transcrição da resposta fornecida pelo *ChatGPT* à pergunta "*Minhas perguntas são utilizadas para o seu treinamento?*").

Apesar da resposta fornecida pela plataforma, a existência da possibilidade de que as informações inseridas por usuários (como planilhas apresentando dados de colaboradores e informações sensíveis de uma empresa) serão realmente utilizadas somente para o treinamento da ferramenta ou, ainda, não serão destinadas à terceiros ou armazenadas em seu banco de dados, deve ser levada em consideração pelas empresas usuárias desta tecnologia.

Em reforço à esta cautela, empresas como a Microsoft [3], proibiram seus colaboradores de utilizarem ferramentas de inteligência artificial para a correção de erros gramaticais automática em e-mails, sob a justificativa de que estes *softwares* podem ter acesso a informações confidenciais transmitidas nos e-mails corrigidos.

No entanto, estes riscos não devem ser impeditivos ao uso de ferramentas como o *ChatGPT*, mas sim fatores de análise e atenção que deverão ser observados no uso destas aplicações, garantindo que seus benefícios possam ser desfrutados, assegurando os fatores necessários para a privacidade e proteção de dados pessoais ou informações confidenciais da empresa.

Ao observar cada um destes pontos, sejam eles positivos ou negativos, a aplicabilidade desta nova tecnologia no dia a dia deverá ser avaliada pela própria empresa, de modo que a ferramenta não seja incorretamente "vulgarizada", mas também não seja utilizada de forma generalizada e sem o devido controle.

Por fim, estamos convictos de que a regulação normativa e o controle existente sobre essas tecnologias deve passar, necessariamente, por um modelo de "*Compliance by Design*", já que o exercício de controle sobre vieses reprováveis e preconceituosos, ou será realizado pela utilização de novos sistemas de detecção que utilizam própria Inteligência Artificial como



suporte, ou por meio de um processo de regulação em que a própria disponibilização da IA possa garantir (*Compliance by Design*), aos seus usuários, a segurança que dela se espera.

[1] <https://www.usatoday.com/story/tech/2023/02/09/woke-chatgpt-conservatives-bias/11215353002/?ref=hackernoon.com>

[2] <https://www.bbc.com/portuguese/geral-64297796>

[3] <https://olhardigital.com.br/2019/06/24/noticias/microsoft-nao-quer-que-funcionarios-usem-tecnologia-de-rivals-no-trabalho/>

Fonte: <https://conjur.jumps.com.br/2023-jun-06/pironti-zarur-chatgpt-compliance-by-design2/>