



Conselho Nacional de Justiça

Autos: **PROCEDIMENTO DE CONTROLE ADMINISTRATIVO - 0000416-89.2023.2.00.0000**
Requerente: **FABIO DE OLIVEIRA RIBEIRO**
Requerido: **CONSELHO NACIONAL DE JUSTIÇA - CNJ**

PARECER

Trata-se de Procedimento de Controle Administrativo proposto por FÁBIO DE OLIVEIRA RIBEIRO por meio do qual o interessado pretende que o Conselho Nacional de Justiça (a.) proíba os juízes brasileiros de recorrer ao ChatGTP para proferir e/ou fundamentar suas decisões nos casos concretos em que atuam, bem ainda (b.) defina regras que permitam aos juízes utilizar a Open AI apenas para fins lúdicos, preservando validade e eficácia da norma constitucional que garante aos cidadãos brasileiros o direito de ver seus processos julgados apenas pelas autoridades competentes.

A demanda pretende-se aos seguintes e principais argumentos, extraídos das manifestações IDs 5009416, 5114489 e 5165969:

- (a) o uso do ChatGTP por juízes pode ser prejudicial, pois a ferramenta não consegue avaliar adequadamente parâmetros fáticos e suas respostas são falaciosas;
- (b) juízes devem realizar suas próprias pesquisas em textos legais, doutrina e jurisprudência, sem transferir seu poder e dever de julgamento para um recurso tecnológico;
- (c) todo cidadão tem direito ao julgamento por um juiz humano, investido do poder e dever de julgar processos;
- (d) o CNJ deve frear a invasão da inteligência artificial no sistema de justiça, em razão dos riscos de privatização e desnacionalização tecnológica, com danos semelhantes aos causados pelos algoritmos de redes sociais.

O pedido de tutela de urgência foi indeferido (ID 5030732).

Os autos aportaram no Departamento de Tecnologia da Informação e Comunicação (DTI/CNJ) para manifestação técnica (ID 5042116).



Conselho Nacional de Justiça

Diante dos termos da demanda, segue o parecer.

1. Grandes Modelos de Linguagem: quadro geral

A compreensão do alcance da pretensão deduzida pelo interessado pressupõe a definição de uma taxonomia capaz de representar as múltiplas aplicações associadas ao uso dos modelos de inteligência artificial generativos de texto. Nesse sentido, é necessário compreender o que são *AI Foundation Models*, *Large Language Models* (LLMs) e *Generative AI*, bem como sua relação com a aplicação produzida pela empresa OpenAI denominada ChatGPT.

Segundo Bommasani et al.[\[1\]](#), os *foundation models* são modelos de inteligência artificial treinados em dados amplos que podem ser adaptados para uma grande variedade de tarefas. São modelos versáteis e criticamente centrais; porém, são modelos incompletos. Em razão dos recentes avanços técnicos, no sentido da melhoria na capacidade de processamento dos equipamentos utilizados no processo de treinamento, bem como o desenvolvimento da arquitetura baseada em *transformers* e a disponibilidade de grandes massas de dados para treinamento, os modelos básicos evoluíram para grandes modelos de linguagem (*large language models*).

Sobre os *large language models* (LLMs), Mökander, Schuett, Kirk e Floridi[\[2\]](#) lecionam que os grandes modelos são desenvolvidos em dois estágios. Num primeiro momento, os modelos são pré-treinados usando aprendizado auto-supervisionado a partir de grandes massas de dados não estruturados. Em seguida, o comportamento dos modelos é ajustado a partir de um conjunto de dados menor, com anotações ou exemplos específicos de tarefas. Segundo os autores,

Embora os LLMs representem um grande avanço na pesquisa de IA, a ideia de construir máquinas de processamento de texto não é nova. Desde a década de 1950, pesquisadores e praticantes de PNL têm desenvolvido software que pode analisar, manipular e gerar linguagem natural. Até a década de 1980, a maioria dos sistemas NLP usava regras baseadas em lógica e focava em automatizar a análise estrutural da linguagem necessária para permitir a tradução automática e o reconhecimento de fala. Mais recentemente, o advento do aprendizado profundo, os avanços nas arquiteturas neurais, como *transformers*, o crescimento do poder



Conselho Nacional de Justiça

computacional e a disponibilidade de dados de treinamento extraídos da Internet revolucionaram o campo, permitindo a criação de LLMs que podem se aproximar do desempenho humano em algumas métricas. A principal função dos LLMs é produzir a continuação mais provável de uma sequência de texto.

Embora vocacionados à produção de texto, conforme apreciação de Weidinger et al.[\[3\]](#), os LLMs se relevaram sistemas de previsão de sequência com aplicabilidade muito mais geral em comparação com os modelos anteriores, exibindo maior capacidade de aprendizagem a partir de poucas instruções (few-shot) ou mesmo nenhuma instrução (zero-shot). Já Chen, Andiappan, Jenkin e Ovchinnikov[\[4\]](#) destacam os principais LLMs da atualidade, anotando que

Large language models (LLMs), como PaLM, LaMDA, RoBERTa e Generative Pre-trained Transformers (GPTs) são algoritmos massivos de aprendizado de máquina/inteligência artificial (IA) que processam e geram texto. As versões recentes do GPT incluem GPT-3, lançado em maio de 2020, e GPT-3.5 (InstructGPT), lançado em janeiro de 2022; era baseado no GPT-3, mas incluía um ajuste fino adicional com feedback humano.

Por sua vez, Inie, Falk e Tanimoto[\[5\]](#) destacam o conceito de inteligência artificial generativa (*generative AI*), no sentido do conjunto de tecnologias que geram automaticamente conteúdo visual ou escrito com base em prompts de texto. A aplicação denominada ChatGPT, desenvolvido pela empresa OpenIA, é um exemplo desta categoria.

Especificamente sobre o ChatGPT, Chen, Andiappan, Jenkin e Ovchinnikov[\[6\]](#) ensinam que

O ChatGPT não é um LLM per se, mas sim um aplicativo de IA construído “em cima” do GPT-3.5 com orientação ainda mais humana, lançado em novembro de 2022. Ele rapidamente impressionou o mundo com conversas articuladas, amplo conhecimento geral e notável capacidade de resolução de problemas de gerenciamento de operações.

A aplicação ChatGPT, lançada em novembro de 2022, atingiu a marca de 100 milhões de usuários em dois meses de operação[\[7\]](#), desencadeando forte movimento de desenvolvimento de



Conselho Nacional de Justiça

aplicações baseadas em LLMs. Porém, é importante compreender que, como adverte Floridi[8], os modelos de inteligência artificial baseados em LLMs não raciocinam ou entendem, não são um passo para nenhuma IA de ficção científica e nada têm a ver com os processos cognitivos presentes no mundo animal e, acima de tudo, no cérebro e na mente humanos; mesmo assim os LLMs são capazes de fazer estatisticamente operações semelhantes àquelas que os humanos fazem semanticamente, a partir da operação com a estrutura formal e não com o significado dos textos com os quais lidam.

Embora aplicações baseadas em GPT não operem segundo o paradigma do pensamento humano, fato é que os resultados produzidos pelos grandes modelos são expressivos. Bommarito e Katz[9] utilizaram o modelo GPT 3.5 disponibilizado pela desenvolvedora OpenAI para resolução de questões do Multistate Bar Examination (MBE), prova de múltipla escolha que consiste na primeira etapa do Uniform Bar Examination (UBE), que por sua vez é o conjunto de exames adotados como pré-requisito para que graduados em direito nos Estados Unidos possam exercer a advocacia. Segundo os autores, em experimento publicado em dezembro/2022 o GPT 3.5 superou significativamente as expectativas de desempenho, mesmo sendo utilizado em configurações zero-shot e com esforço mínimo de modelagem e otimização, tendo ultrapassado a taxa de linha de base de adivinhação aleatória e alcançado taxa de aprovação em duas disciplinas de avaliação, uma das quais apresentou paridade com os examinandos humanos.

Já em estudo publicado em março/2023, Katz, Bommarito, Gao e Arredondo[10] noticiam que o modelo GPT 4, também disponibilizado pela desenvolvedora OpenAI, alcançou performance suficiente para ser aprovado em todos os componentes do Uniform Bar Examination (UBE). Segundo os autores, o GPT 4 superou a performance dos examinandos humanos em cinco das sete disciplinas do Multistate Bar Examination (MBE), bem como alcançou a nota de 4.2 de 6.0 nas provas discursivas do Multistate Essay Examination (MEE) e no Multistate Performance Test (MPT). No geral, o modelo pontuou 297 pontos, suficiente para aprovação no exame realizado em qualquer dos Estados norte-americanos. Diante dos resultados alcançados, os autores anotam que grandes modelos de linguagem podem atender ao padrão aplicado a advogados humanos em quase todas as jurisdições nos Estados Unidos, lidando com tarefas complexas que exigem profundo conhecimento jurídico, compreensão de leitura e capacidade de escrita.



Conselho Nacional de Justiça

No Brasil, em experimento realizado em fevereiro/2023 com uso da aplicação ChatGPT, o modelo alcançou pontuação de 48 pontos dos 80 pontos possíveis na primeira fase da prova do exame da Ordem dos Advogados do Brasil[11]. O autor do experimento, Daniel Marques, comentou em tom profético: “os advogados não serão substituídos pelo ChatGPT. Eles serão substituídos pelos advogados capazes de usar a ferramenta”.

2. Uso de modelos de inteligência artificial generativos de texto no Judiciário brasileiro

Admitindo que o atual estágio de desenvolvimento dos modelos de inteligência artificial generativos denota um marco relevante no processo de transformação digital da sociedade, no qual a disseminação do uso dessas soluções impõe-se como um fato, é necessário responder ao seguinte questionamento: é razoável e operacionalmente viável proibir o uso de modelos generativos no Judiciário?

Sob uma perspectiva formal, é plenamente possível que o Conselho Nacional de Justiça, no exercício de seu poder-dever de controle a atuação administrativa do Poder Judiciário estabelecido no artigo 103-B, parágrafo 4º da Constituição da República[12], edite ato normativo com o objetivo de regular ou mesmo proibir o uso de soluções de inteligência artificial por magistrados brasileiros. Esse espaço regulatório já foi, inclusive, ocupado pelo Conselho Nacional de Justiça por meio da edição da Resolução CNJ n. 332, de 21 de agosto de 2020[13], ato que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário.

A Resolução CNJ n. 332/2020 não proíbe o uso de soluções baseadas em inteligência artificial pelo Poder Judiciário. Adotando postura realista, a norma ocupa-se de disciplinar padrões éticos a serem observados no desenvolvimento e uso dos modelos, bem ainda questões relativas à transparência e auditabilidade. Segundo a norma, “a utilização de modelos de Inteligência Artificial deve buscar garantir a segurança jurídica e colaborar para que o Poder Judiciário respeite a igualdade de tratamento aos casos absolutamente iguais” (artigo 5º), sendo que “as decisões judiciais apoiadas em ferramentas de Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo” (artigo 7º).

Como se observa, o regulamento vigente estabelece como premissas fundantes que o uso de modelos de inteligência artificial pelo Poder Judiciário deve contribuir para a segurança jurídica



Conselho Nacional de Justiça

e para a promoção da igualdade. Ademais, soluções tecnológicas baseadas em inteligência artificial são soluções de apoio à atividade jurisdicional, o que significa dizer que a tecnologia não pode ser utilizada para substituir a atividade humana e, notadamente, a atividade decisória propriamente dita; ao contrário, seu funcionamento desse estar sob o controle do usuário, de modo a assegurar sua autonomia e possibilidade de revisão de eventuais propostas de atos decisórios (artigo 17).

Diante do conteúdo da Resolução CNJ n. 332/2020, bem ainda ponderado macrodesafio de promoção da inovação tecnológica previsto na Resolução CNJ n. 325/2020^[14] e o objetivo estratégico de “desenvolver ações e disseminar práticas que estimulem a inovação tecnológica para a ampliação da transformação Digital e da inteligência artificial no Poder Judiciário” estabelecido no Plano Estratégico do Conselho Nacional de Justiça (Objetivo Estratégico n. 9)^[15], a eventual edição de norma restritiva do uso de soluções de inteligência artificial pelo Poder Judiciário caminha no sentido contrário dos direcionadores estratégicos vigentes.

Sob outro enfoque, ainda que se possa restringir o uso institucional de modelos de inteligência artificial generativa, no sentido de impor-se aos órgãos submetidos à autoridade fiscalizatória do Conselho Nacional de Justiça obrigação de não desenvolvimento de iniciativas corporativas de uso da tecnologia, o uso das soluções por magistrados e servidores do Poder Judiciário pode dar-se em contextos que escapam a qualquer estratégia de controle e fiscalização.

Isso porque o uso não institucional das soluções de inteligência artificial generativa pode dar-se por meio de acesso a múltiplas aplicações disponíveis na internet, a exemplo do ChatGPT, Google Bard e Bing Chat. Nessa linha e a título ilustrativo, o repositório Futurepedia^[16] lista mais de 4.700 (quatro mil e setecentos) serviços e aplicações, com variadas finalidades, orientados a disponibilizar o consumo de soluções baseadas em modelos generativos.

Nesse contexto, eventual norma proibitiva do uso de modelos de inteligência artificial generativos no âmbito do Poder Judiciário, dada a multiplicidade de modelos e serviços disponíveis, pode resultar em texto normativo inexecutável ou, quando menos, desprovido de eficácia prática, na medida em que o controle e fiscalização do cumprimento de norma com tal conteúdo importa em desafio operacional que demanda recursos humanos e tecnológicos não disponíveis.

Consideradas as variáveis fáticas e normativas apresentadas, é possível afirmar que não é razoável e tão pouco operacionalmente viável proibir o uso de modelos de inteligência artificial



Conselho Nacional de Justiça

generativa no Judiciário brasileiro. Contudo, importante ponderar que eventual não edição de norma restritiva ou proibitiva das referidas soluções não deve ser interpretado como liberação irrefletida de seu uso, na medida em que a tecnologia possui limitações significativas.

3. Riscos mapeados

Riscos são inerentes a qualquer atividade humana, com variações quanto ao grau de probabilidade de sua ocorrência. Segundo o Tribunal de Contas da União[17], risco consiste na “possibilidade de que um evento afete negativamente o alcance dos objetivos”. Essas variáveis devem ser identificadas e dimensionadas por meio da análise de riscos, bem como tratadas por meio da gestão de riscos. O tratamento de um risco pode implicar na descontinuidade da atividade, no compartilhamento de responsabilidades, na adoção de medidas de prevenção ou contenção, ou ainda na simples aceitação do risco e monitoramento da situação potencialmente danosa.

A disseminação do uso de modelos de inteligência artificial generativa em variados contextos tem motivado estudos para identificação das limitações e riscos associados à tecnologia. Relatório disponibilizado pela consultoria Gartner Inc., elaborado por Burke et al.[18], informa que o texto gerado pela aplicação ChatGPT do fornecedor OpenAI depende, primeiramente, da qualidade do prompt e da qualidade e quantidade de dados de treinamento associados ao tema da interação. Segundo os consultores, o modelo pode apresentar respostas coerentes e intrinsecamente erradas - comportamento nominado como “alucinações” -, bem como respostas tendenciosas, obsoletas ou incompletas, em razão de limites associados ao conjunto de dados utilizados no treinamento.

Especificamente sobre o tema das alucinações, Buchanan e Shapoval[19] relatam que, em uma sequência de iterações sobre temas relacionados às ciências sociais, a aplicação ChatGPT foi provocada a apresentar respostas com citação da fonte consultada, incluindo autor, ano e local da publicação. Os autores classificam como real a citação que permite localizar o texto referido segundo informado pela aplicação, bem como correta a citação pertinente para a resposta apresentada. Como resultado, os autores obtiveram 57 respostas sobre temas variados, sendo que a aplicação apresentou 22% de citações irreais e 28% de citações incorretas em respostas sobre temas



Conselho Nacional de Justiça

gerais, bem como 54% de citações irreais e 71% de citações incorretas em respostas sobre temas específicos.

Já Weidinger et al.[\[20\]](#) catalogam um conjunto de riscos associados ao uso de *large language models* (LLMs), tipificados em riscos de alocação, quando recursos e oportunidades são distribuídos injustamente entre grupos sociais, e riscos de representação, quando modelos reproduzem estereótipos e promovem discriminação. Segundo os autores, os principais riscos podem ser tabulados como: (a.) discriminação e exclusão: tratamento injusto de grupos marginalizados; (b.) riscos de informação: danos à privacidade e comprometimento de informações sensíveis; (c.) desinformação: disseminação de informações falsas ou incorretas; (d.) uso malicioso: modelos utilizados intencionalmente por indivíduos como instrumento para crimes como fraude ou para campanhas de desinformação; (e.) danos à interação humano-computador: o comportamento dos modelos pode desencadear excesso de confiança dos usuários e favorecer danos psicológicos, violações de privacidade e promoção de estereótipos; (f.) automação e acesso: modelos podem ser utilizados para sustentação de aplicativos que beneficiam de forma desproporcional determinados grupos.

Sob outra perspectiva, Chen, Andiappan, Jenkin e Ovchinnikov[\[21\]](#) utilizaram a aplicação ChatGPT em uma sequência de iterações para investigação de eventuais vieses comportamentais, nas linhas da análise de riscos, avaliação de resultados e heurísticas decisórias. Segundo os autores, os resultados do experimento relevaram evidências de (a.) viés de conjunção, no sentido da incorreta estimativa da ocorrência de eventos conjuntos, (b.) viés de ponderação de probabilidades, no sentido da incorreta estimação de resultados potenciais associados a probabilidades declaradas; (c.) excesso de confiança, no sentido da superestimação da probabilidade de apresentação de respostas corretas; (d.) enquadramento, no sentido de que o enquadramento de uma decisão como ganho ou perda pode mudar o resultado da interação; (e.) aversão ao arrependimento, no sentido da influência desproporcional da ponderação antecipada de uma decisão equivocada na tomada de decisão; (f.) dependência de referência, no sentido da incorreta ponderação dos ganhos e perdas associados a uma situação específica; (g.) viés de confirmação, no sentido da busca de evidências que possam confirmar uma hipótese ou crença anterior, ignorando evidências que possam refutá-la.

Como se pode observar, o uso de soluções baseadas em *large language models* (LLMs) é atividade associada a riscos que podem ser sumarizados em: (a.) riscos relativos à privacidade e



Conselho Nacional de Justiça

informações sensíveis; (b.) riscos relativos a informações falsas, incorretas ou incompletas; (c.) riscos associados a vieses discriminatórios; (d.) riscos associados a vieses comportamentais. Considerando que as soluções de inteligência artificial em uso no Poder Judiciário devem estar sob o controle do usuário, no sentido de ser preservada sua autonomia e garantida a possibilidade de revisão de eventuais propostas de atos decisórios (Resolução CNJ n. 332/2020, artigo 17), ainda que se atribua alta probabilidade de materialização dos riscos mapeados, a existência de uma instância de controle humano reduz a severidade dos riscos para níveis que autorizam a não proibição do uso das soluções.

Todavia, os riscos mapeados não podem ser simplesmente aceitos e monitorados; é necessário a implementação de medidas de mitigação e compartilhamento de responsabilidades. Nesse sentido, sugerem-se as seguintes medidas de tratamento de riscos:

- (a) proibição do uso de soluções de inteligência artificial baseadas em *large language models* (LLMs) que importem em automação de atividades decisórias;
- (b) determinação para que implementações de uso de LLMs assegurem ao usuário humano a revisão e opção de escolha quanto ao aproveitamento de insumo fornecido pelo modelo;
- (c) determinação para que, em cenários de uso corporativo, os órgãos do Poder Judiciário adotem providências para preservação de dados pessoais e informações sensíveis;
- (d) determinação para que, em cenários de uso corporativo ou individual, os usuários sejam submetidos à capacitação formal, a fim de serem esclarecidos quanto aos riscos e estratégias de prevenção, incluindo construção de contextos e engenharia de prompt.

Sob uma perspectiva regulatória ampliada, e na linha das providências propostas por Mökander, Schuett, Kirk e Floridi^[22], sugere-se ainda a constituição de grupo de trabalho com o objetivo de avaliar a necessidade de atualização da Resolução CNJ n. 332/2020, esforço orientado pelos seguintes objetivos mínimos:



Conselho Nacional de Justiça

- (a) definição de modelo de governança para gestão do processo de desenvolvimento, sustentação e uso das soluções de inteligência artificial no Poder Judiciário brasileiro, orientado pela transparência e auditabilidade;
- (b) definição de processo de trabalho referencial para auditoria de modelos e soluções de inteligência artificial, sob as perspectivas da segurança da informação, performance, robustez, confiabilidade, vieses, correlação entre entradas e saídas e conformidade legal e ética;
- (c) definição de casos de uso permitidos, regulados e proibidos.

4. Conclusão

Diante das considerações expostas, o Departamento de Tecnologia da Informação e Comunicação do Conselho Nacional de Justiça **manifesta-se** pela não proibição do uso de modelos de inteligência artificial baseadas em *large language models* (LLMs) no âmbito do Poder Judiciário brasileiro.

Por outro lado, o DTI/CNJ **sugere** a adoção das seguintes medidas de tratamento de riscos:

- a) proibição do uso de soluções de inteligência artificial baseadas em *large language models* (LLMs) que importem em automação de atividades decisórias;
- b) determinação para que implementações de uso de LLMs assegurem ao usuário humano a revisão e opção de escolha quanto ao aproveitamento de insumo fornecido pelo modelo;
- c) determinação para que, em cenários de uso corporativo, os órgãos do Poder Judiciário adotem providências para preservação de dados pessoais e informações sensíveis;
- d) determinação para que, em cenários de uso individual, os usuários sejam submetidos à capacitação formal, a fim de serem esclarecidos quanto aos riscos e estratégias de prevenção, incluindo construção de contextos e engenharia de prompt.



Conselho Nacional de Justiça

Por fim, o DTI/CNJ **sugere** ainda a constituição de grupo de trabalho com o objetivo de avaliar a necessidade de atualização da Resolução CNJ n. 332/2020, pautado pelos seguintes objetivos mínimos:

- a) definição de modelo de governança para gestão do processo de desenvolvimento, sustentação e uso das soluções de inteligência artificial no Poder Judiciário brasileiro, orientado pela transparência e auditabilidade;
- b) definição de processo de trabalho referencial para auditoria de modelos e soluções de inteligência artificial, sob as perspectivas da segurança da informação, performance, robustez, confiabilidade, vieses, correlação entre entradas e saídas e conformidade legal e ética;
- c) definição de casos de uso permitidos, regulados e proibidos.

Sendo estas as considerações técnicas possíveis no atual estágio de processamento do feito, submeto o presente parecer à apreciação de Vossa Excelência.

Respeitosamente,

Brasília/DF, 09 de agosto de 2023.

João Thiago de França Guerra

Juiz Auxiliar da Presidência

Supervisor do Departamento de Tecnologia da Informação e Comunicação

Adriano da Silva Araújo

Juiz Auxiliar da Presidência

Supervisor do Departamento de Tecnologia da Informação e Comunicação

Thiago de Andrade Vieira

Diretor do Departamento de Tecnologia da Informação e Comunicação



Conselho Nacional de Justiça

-
- [1] BOMMASANI, Rishi; HUDSON, Drew A.; ADELI, Ehsan; ALTMAN, Russ; ARORA, Simran; VON ARX, Sydney ; BERNSTEIN, Michael S.; BOHG, Jeannette; BOSSELUT, Antoine; BRUNSKILL, Emma; BRYNJOLFSSON, Erik; BUCH, S.; CARD, Dallas; CASTELLON, Rodrigo; CHATTERJI, Niladri S.; CHEN, Annie S.; CREEL, Kathleen A.; DAVIS, Jared; DEMSZKY, Dora; DONAHUE, Chris; DOUMBOUYA, Moussa; DURMUS, Esin; ERMON, Stefano; ETCHEMENDY, John; ETHAYARAJH, Kawin; FEI-FEI, Li; FINN, Chelsea; GALE, Trevor; GILLESPIE, Lauren E.; GOEL, Karan; GOODMAN, Noah D.; GROSSMAN, Shelby; GUHA, Neel; HASHIMOTO, Tatsunori; HENDERSON, Peter; HEWITT, John; HO, Daniel E.; HONG, Jenny; HSU, Kyle; HUANG, Jing; ICARD, Thomas F.; JAIN, Saahil; JURAFSKY, Dan; KALLURI, Pratyusha; KARAMCHETI, Siddharth; KEELING, Geoff; KHANI, Fereshte; KHATTAB, O.; KOH, Pang Wei; KRASS, Mark S.; KRISHNA, Ranjay; KUDITIPUDI, Rohith; KUMAR, Ananya; LADHAK, Faisal; LEE, Mina; LEE, Tony; LESKOVEC, Jure; LEVENT, Isabelle; LI, Xiang Lisa; LI, Xuechen; MA, Tengyu; MALIK, Ali; MANNING, Christopher D.; MIRCHANDANI, Suvir; MITCHELL, Eric; MUNYIKWA, Zanele; NAIR, Suraj; NARAYAN, Avanika; NARAYANAN, Deepak; NEWMAN, Benjamin; NIE, Allen; NIEBLES, Juan Carlos; NILFOROSHAN, Hamed; NYARKO, J. F.; OGUT, Giray; ORR, Laurel J.; PAPADIMITRIOU, Isabel; PARK, Joon Sung; PIECH, Chris; PORTELANCE, Eva; POTTS, Christopher; RAGHUNATHAN, Aditi; REICH, Robert; REN, Hongyu; RONG, Frieda; ROOHANI, Yusuf H.; RUIZ, Camilo; RYAN, Jack; R'E, Christopher; SADIGH, Dorsa; SAGAWA, Shiori; SANTHANAM, Keshav; SHIH, Andy; SRINIVASAN, Krishna Parasuram; TAMKIN, Alex; TAORI, Rohan; THOMAS, Armin W.; TRAMÈR, Florian; WANG, Rose E.; WANG, William; WU, Bohan; WU, Jiajun; WU, Yuhuai; XIE, Sang Michael; YASUNAGA, Michihiro; YOU, Jiaxuan; ZAHARIA, Matei A.; ZHANG, Michael; ZHANG, Tianyi; ZHANG, Xikun; ZHANG, Yuhui; ZHENG, Lucia; ZHOU, Kaitlyn; LIANG, Percy. On the Opportunities and Risks of Foundation Models. ArXiv abs/2108.07258, 2021. Disponível em <https://arxiv.org/abs/2108.07258>. Acesso em 08 ago. 2023.
- [2] MOKANDER, Jakob; SCHUETT, Jonas; KIRK, Hannah Rose; FLORIDI, Luciano. Auditing large language models: a three-layered approach. ArXiv abs/2302.08500, 2023. Disponível em <https://arxiv.org/abs/2302.08500>. Acesso em 08 ago. 2023.
- [3] WEIDINGER, Laura, MELLOR, John F. J., RAUH, Maribeth, GRIFFIN, Conor, UESATO, Jonathan, HUANG, Po-Sen, CHENG, Myra, GLAESE, Mia, BALLE, Borja, KASIRZADEH, Atoosa, KENTON, Zachary, BROWN, Sande Minnich, HAWKINS, William T., STEPLETON, Tom, BILES, Courtney, BIRHANE, Abeba, HAAS, Julia, RIMELL, Laura, HENDRICKS, Lisa Anne, ISAAC, William S., LEGASSICK, Sean, IRVING, Geoffrey, GABRIEL, Iason. Ethical and social risks of harm from Language Models. ArXiv abs/2112.04359, 2021. Disponível em <https://arxiv.org/abs/2112.04359>. Acesso em 08 ago. 2023.
- [4] CHEN, Yang; ANDIAPPAN, Meena; JENKIN, Tracy; OVCHINNIKOV, Anton. A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do? SSRN, 2023. Disponível em <https://ssrn.com/abstract=4380365>. Acesso em 08 ago. 2023.
- [5] INIE, Nanna; FALK, Jeanette; TANIMOTO, Steve. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 2023. Disponível em <https://arxiv.org/abs/2303.08931>. Acesso em 08 ago. 2023.
- [6] CHEN, Yang; ANDIAPPAN, Meena; JENKIN, Tracy; OVCHINNIKOV, Anton. A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do? SSRN, 2023. Disponível em <https://ssrn.com/abstract=4380365>. Acesso em 08 ago. 2023.



Conselho Nacional de Justiça

- [7] HU, Krystal. ChatGPT sets record for fastest-growing user base. Reuters, 02 fev. 2023. Disponível em <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. Acesso em 08 ago. 2023.
- [8] FLORIDI, Luciano. AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. *Philos. Technol.* 36, 15, 2023. Disponível em <https://doi.org/10.1007/s13347-023-00621-y>. Acesso em 08 ago. 2023.
- [9] BOMMARITO, Michael James; KATZ, Daniel Martin. GPT Takes the Bar Exam. SSRN, 29 dez. 2022. Disponível em <https://ssrn.com/abstract=4314839>. Acesso em 08 ago. 2023.
- [10] KATZ, Daniel Martin; BOMMARITO, Michael James; GAO, Shang; ARREDONDO, Pablo. GPT-4 Passes the Bar Exam. SSRN, 15 mar. 2023. Disponível em <https://ssrn.com/abstract=4389233>. Acesso em 08 ago. 2023.
- [11] ChatGPT é ‘aprovado’ em prova da primeira fase da OAB. *Estadão*, 21 fev. 2023. Disponível em <https://www.estadao.com.br/link/cultura-digital/chatgpt-e-aprovado-em-prova-da-primeira-fase-da-oab/>. Acesso em 08 ago. 2023.
- [12] BRASIL. Constituição da República Federativa do Brasil de 1988. Disponível em https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 08 ago. 2023.
- [13] CONSELHO NACIONAL DE JUSTIÇA. Resolução n. 332, de 21 de agosto de 2020. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em 08 ago. 2023.
- [14] CONSELHO NACIONAL DE JUSTIÇA. Resolução n. 325, de 29 de junho de 2020. Dispõe sobre a Estratégia Nacional do Poder Judiciário 2021-2026 e dá outras providências. Disponível em <https://atos.cnj.jus.br/atos/detalhar/3365>. Acesso em 08 ago. 2023.
- [15] CONSELHO NACIONAL DE JUSTIÇA. Plano Estratégico do Conselho Nacional de Justiça. Disponível em <https://www.cnj.jus.br/wp-content/uploads/2023/07/plano-estrategico-5a-versao-2023.pdf>. Acesso em 08 ago. 2023.
- [16] FUTUREPEDIA, the largest AI tools directory. Disponível em <https://www.futurepedia.io/>. Acesso em 08 ago. 2023.
- [17] TRIBUNAL DE CONTAS DA UNIÃO. Manual de gestão de riscos do TCU. Brasília: TCU, Secretaria de Planejamento, Governança e Gestão (Seplan), 2020, p. 15. Disponível em [file:///C:/Users/joao.guerra/Downloads/manual de gestao de riscos 2aEdicao Final.pdf](file:///C:/Users/joao.guerra/Downloads/manual%20de%20gestao%20de%20riscos%202aEdicao%20Final.pdf). Acesso em 08 ago. 2023.
- [18] BURKE, Brian; ELLIOT, Bern; CHANDRASEKARAN, Arun; SICULAR, Svetlana; ANDREWS, Whit. Quick Answer: What Is ChatGPT? Gartner Inc., 30 jan. 2023. Disponível em <https://www.gartner.com/document/4031299?ref=solrAll&refval=374490217&>. Acesso em 08 ago. 2023.
- [19] BUCHANAN, Joy; SHAPOVAL, Olga. GPT-3.5 Hallucinates Nonexistent Citations: Evidence from Economics. SSRN, 03 jun. 2023). Disponível em <https://ssrn.com/abstract=4467968>. Acesso em 08 ago. 2023.



Conselho Nacional de Justiça

- [20] WEIDINGER, Laura, MELLOR, John F. J., RAUH, Maribeth, GRIFFIN, Conor, UESATO, Jonathan, HUANG, Po-Sen, CHENG, Myra, GLAESE, Mia, BALLE, Borja, KASIRZADEH, Atoosa, KENTON, Zachary, BROWN, Sande Minnich, HAWKINS, William T., STEPLETON, Tom, BILES, Courtney, BIRHANE, Abeba, HAAS, Julia, RIMELL, Laura, HENDRICKS, Lisa Anne, ISAAC, William S., LEGASSICK, Sean, IRVING, Geoffrey, GABRIEL, Iason. Ethical and social risks of harm from Language Models. ArXiv abs/2112.04359, 2021. Disponível em <https://arxiv.org/abs/2112.04359>. Acesso em 08 ago. 2023.
- [21] CHEN, Yang; ANDIAPPAN, Meena; JENKIN, Tracy; OVCHINNIKOV, Anton. A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do? SSRN, 2023. Disponível em <https://ssrn.com/abstract=4380365>. Acesso em 08 ago. 2023.
- [22] MOKANDER, Jakob; SCHUETT, Jonas; KIRK, Hannah Rose; FLORIDI, Luciano. Auditing large language models: a three-layered approach. ArXiv abs/2302.08500, 2023. Disponível em <https://arxiv.org/abs/2302.08500>. Acesso em 08 ago. 2023.