



GOVERNANÇA *de* INTELIGÊNCIA ARTIFICIAL GENERATIVA *nas* UNIVERSIDADES

Juliano Maranhão
Juliana Krueger Pela
Francisco Marino
Amalia Batocchio

lawgorithm*}

GOVERNANÇA DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NAS UNIVERSIDADES

Relatório da Comissão de Inteligência Artificial Generativa da Faculdade de
Direito da Universidade de São Paulo

Juliano Maranhão

Juliana Krueger Pela

Francisco Marino

Amália Batocchio

Em cooperação com a Associação Lawgorithm de Pesquisa em Inteligência Artificial, no âmbito do Termo de Cooperação Acadêmica com a Faculdade de Direito da Universidade de São Paulo

Sumário

SUMÁRIO EXECUTIVO	3
CAPÍTULO 1: ÉTICA E GOVERNANÇA NO USO DE INTELIGÊNCIAS ARTIFICIAIS GENERATIVAS PARA FINS EDUCACIONAIS.....	5
1. Introdução.....	5
2. Inteligência Artificial Generativa na educação	10
2.1. Usos de IAG nas atividades acadêmicas	10
2.2. Governança de IAG nas editoras científicas	14
2.3. Governança do uso de IAG nas universidades.....	18
2.4. A preocupação com integridade acadêmica	25
2.5. Outras preocupações e medidas de governança no uso de IAG.....	32
CAPÍTULO 2: O USO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA PELOS(AS) ALUNOS DA FACULDADE DE DIREITO DA USP (FDUSP).....	36
1. Metodologia.....	36
2. Análise dos resultados	38
3. Conclusões e principais achados.....	45
Referências Bibliográficas.....	47
ANEXOS	56
ANEXO I - Governança da Inteligência Artificial como gestão de riscos	56
ANEXO II - Introdução aos Sistemas de Inteligência Artificial Generativa	62
ANEXO III - Riscos e Governança de sistemas de IA Generativa de textos.....	69

SUMÁRIO EXECUTIVO

O presente relatório trata de procedimentos de governança para o uso ético de sistemas de inteligência artificial generativa (“IAG”) para fins acadêmicos e educacionais, especialmente aqueles aplicáveis ao ensino superior.

Busca-se discutir as potencialidades e riscos associadas aos usos da IAG, bem como possíveis medidas que podem ser adotadas para mitigar seus riscos e, ao final, apresenta-se pesquisa realizada entre os alunos de graduação e de pós-graduação da Faculdade de Direito da Universidade de São Paulo (“FDUSP”) realizada no ano de 2023.

As IAGs são ferramentas que possibilitam a geração de conteúdo novo (seja este de texto, áudio, imagens e vídeo) em interfaces intuitivas que permitem interação em linguagem natural. As IAGs são desenvolvidas por meio de recursos intensivos de dados, capacidade computacional e expertise.

O conteúdo gerado pela IAGs é, em grande medida e ao menos até o presente momento, indistinguível do conteúdo gerado por humanos sem o uso de ferramentas de detecção. Ocorre que os resultados gerados por IA podem conter erros, imprecisões e até mesmo “alucinações” (referência a fatos ou documentos inexistentes) que demandariam revisão pelo usuário. Além disso, há outras preocupações como a falta de transparência no uso, possível violação autoral, a geração de conteúdo discriminatório ou abusivo, riscos à privacidade e proteção de dados, com a incorporação de conteúdo sensível em requisições de texto pelo usuário (*prompts*), com possível incorporação do mesmo por sistemas de treinamento contínuo.

Além disso, a popularidade das IAGs torna implausível que o seu uso seja proibido de modo eficaz. De fato, a impossibilidade de detecção da aplicação de ferramentas de IAG, a dúvida quanto à licitude de seu uso para fins acadêmicos e possíveis retaliações, fazem com que seu emprego não seja, em muitos casos, transparente. Isso tudo, aliado aos riscos que já apontamos acima, tornam mais urgentes a adoção de medidas para levantamento dos usos realizados pelos alunos e docentes, culminando no estabelecimento de políticas de governança próprias para essa tecnologia.

Em relação à aplicação de IAGs para fins educacionais, as principais preocupações éticas estão ligadas ao viés de automação (excesso de deferência aos resultados gerados pela máquina) com (i) eventual reprodução de vieses; (ii) problemas relacionados à privacidade e à proteção de dados pessoais; (iii) qualidade dos resultados; (iv) infrações a direitos autorais.

A forma de mitigar tais riscos está na adoção de medidas de governança, por meio de (i) treinamento e divulgação para esclarecimento quanto aos usos adequados e inadequados da ferramenta; (ii) revisão do conteúdo gerado, de forma a garantir sua acurácia e mitigar a presença de vieses discriminatórios; (iii) transparência em seu uso; e (iv) evitar requisições que contenham dados pessoais e permitam a avaliação de detecção de possíveis violações a direitos autorais.

Já em relação a medidas institucionais cuja adoção é recomendada pela literatura e experiência internacional, estão: (i) necessidade de transparência quanto ao uso da IAG; (ii) necessidade de revisão do conteúdo produzido pela ferramenta de IAG; e (iii) não inclusão, dentre os *prompts*, de dados pessoais ou informações sigilosas.

Tais medidas de governança são abordadas pelas mais prestigiadas universidades do mundo e estão de acordo com documentos elaborados pela UNESCO.

Por fim, quanto ao levantamento sobre o uso de IAGs por alunos da FDUSP, observa-se que o uso pelos alunos é significativo. Cerca de 70% dos alunos que responderam afirmam já ter feito uso de alguma ferramenta de inteligência artificial para fins relacionados às atividades acadêmicas.

De forma salutar, mais de 85% dos alunos apontam que revisam o conteúdo gerado pela ferramenta. Os respondentes apontam, como principais dificuldades e limitações ao uso das IAGs, a falta de confiança nos textos gerados, a preocupação com a licitude e ética no seu uso. A percepção de limitações da ferramenta é positiva, por demonstrar a tendência a revisar o conteúdo gerado.

Contudo, dúvidas quanto à licitude em se usar de ferramentas de IAG para trabalhos juntos à faculdade podem levar à falta de transparência. Isso, de fato, se verifica, uma vez que 71% dos alunos apontaram que não informam no trabalho ou atividade acadêmica que a IAG foi utilizada.

Essa falta de transparência é preocupante, por dificultar processos internos de revisão e correção dos trabalhos e atividades acadêmicas. Esse fato, aliado a dúvidas quanto à legitimidade e à legalidade do uso, apontam para a necessidade de ações pela faculdade para promover a conscientização e a criação de regras e diretrizes que promovam a transparência e o uso ético das IAGs.

CAPÍTULO 1: ÉTICA E GOVERNANÇA NO USO DE INTELIGÊNCIAS ARTIFICIAIS GENERATIVAS PARA FINS EDUCACIONAIS

1. Introdução

Nas últimas décadas, especialmente com o grande desenvolvimento de avanço dos modelos de aprendizado de máquina, a inteligência artificial (“IA”) tem sido utilizada em diversas áreas, como no setor de serviços¹ e da administração pública², no setor agrícola³ e na medicina,⁴ na educação,⁵ proteção ao meio ambiente,⁶ segurança pública,⁷ e gestão de transportes.⁸

No campo do Direito, a IA encontra uma série de aplicações, seja nos escritórios de advocacia ou nos tribunais. Alguns usos que podemos mencionar são a classificação e acompanhamento de processos, síntese e extração de informações relevantes de documentos, buscas inteligentes de conteúdo, em particular jurisprudência, predição de decisões e análise de desempenho dos

¹ STANFORD UNIVERSITY – Human Centered Artificial Intelligence (HAI). **AI Index Report**. 7ª edição, 2024. Disponível em: <<https://aiindex.stanford.edu/report/>>. Acesso em 03 jul. 2024.

² BANCO INTERAMERICANO DE DESENVOLVIMENTO. **Contratações de inovação**: guia de alternativas jurídicas e de boas práticas para contratações de inovação. 1. ed. Washington, D.C.: BID, 2023. Disponível em: <<https://publications.iadb.org/pt/contratacoes-de-inovacao-guia-de-alternativas-juridicas-e-de-boas-praticas-para-contratacoes-de>>. Acesso em: 15 ago. 2024.

³ DELL’ACQUA, F., MCFOWLAND III, E., MOLLICK, E., LIFSHITZ-ASSAF, H., KELLOGG, K.C., RAJENDRAN, S., KRAYER, L., CANDELON, F. and LAKHANI, K.R. (2023). **Navigating the Jagged Technological Frontier**: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Working Paper. Disponível em: <https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf> Acesso em 12 de jul. 2024.

⁴ MARANHÃO, JULIANO; ALMADA, M. **Inteligência artificial no setor no setor de saúde**: ética e proteção de dados. In: Analluza Bolivar Dallari, Gustavo Ferraz de Campos Monaco. (Org.). LGDP na saúde. 1ed.São Paulo: RT, 2021, v. 1, p. 357-370.

⁵ UNESCO. **Artificial Intelligence and Education**: Guidance for Policy-makers. Paris: UNESCO, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000376709>. Acesso em: 15 ago. 2024. MOLINA, E.; COBO, C.; PINEDA, J.; ROVNER, H. (2024). AI revolution in education: What you need to know. In Digital Innovations in Education. World Bank.

⁶ CHEN, Lin et al. **Artificial intelligence-based solutions for climate change**: a review. *Environmental Chemistry Letters*, v. 21, n. 5, p. 2525-2557, 2023.

⁷ RIGANO, Christopher, “Using Artificial Intelligence to Address Criminal Justice Needs,” *NIJ Journal* 280, January 2019, <<https://www.nij.gov/journals/280/Pages/using-artificialintelligence-to-address-criminal-justice-needs.aspx>>.

⁸ IYER, S.L. AI enabled applications towards intelligent transportation. *Transportation Engineering*, v.5, 2021.

tribunais e geradores automáticos de minutas documentos (petições, despachos etc.).

A IA será um facilitador fundamental da Quarta Revolução Industrial, marcada por tecnologias como impressão 3D, veículos autônomos, biotecnologia, nanotecnologia, computação quântica, robótica e Internet das Coisas, todas sustentadas pela IA. Para além, a IA já é onipresente no trabalho moderno, desde a manufatura ao setor bancário. Isso se reflete nos empregos de alta qualificação, que passam a exigir habilidades criativas e analíticas únicas. Assim, profissionais deverão desenvolver novas habilidades técnicas e sociais para ingressar no mercado de trabalho em um mundo dominado pela IA.⁹

Dentre as ferramentas de inteligência artificial, as Inteligências Artificiais Generativas (IAGs) vêm sendo apontadas como a grande revolução no campo da IA. Relatórios do setor sugerem que a IAG possa aumentar o produto interno bruto (PIB) global em 7% - quase 7 trilhões de dólares – e aumentar a produtividade em 1,5 pontos percentuais no prazo de 10 anos.¹⁰

A IAG popularizou-se a partir de 2022, especialmente aquelas que geram conteúdo novo de texto (e.g. ChatGPT, Llama, BingAI, Gemini, Claude), de imagens e de vídeo (e.g. Dall-e, StyleGAN, Midjourney, Stable Diffusion), e de áudio (e.g. Suno, Udio, Sonauto), cuja interface de fácil utilização e interação é acessível a usuários leigos.

Além disso, com o fácil acesso e a popularização das ferramentas de IAG disponíveis, seu uso deixou de ser uma decisão centralizada nas organizações,

⁹ UNESCO. **Artificial Intelligence and Education: Guidance for Policy-makers**. Paris: UNESCO, 2021. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000376709>>. Acesso em: 15 ago. 2024. p. 12

¹⁰ GOLDMAN SACHS. **Generative AI could raise global GDP by 7%**. 5 de abril de 2023. Disponível em: <<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>>. Acesso em 9 de jul. 2024.

nem sempre havendo transparência por parte dos usuários quanto à sua utilização.¹¹

A IA possui potencial para transformar de forma significativa a educação, fornecendo oportunidades para aumento da personalização, eficiência e qualidade da aprendizagem.¹² Contudo, ensinar e aprender sobre IA preparam a comunidade acadêmica para os desafios e oportunidades. O desafio relevante aqui é promover o desenvolvimento de competências e habilidades críticas como resolução de problemas, pensamento crítico, criatividade e tomada de decisão.¹³

Por outro lado, a IA, por suas próprias características, traz preocupações inerentes como: (i) risco de erro, cujas principais causas são as limitações de acurácia estatística, falhas no *design* ou em treinamento, bem como a incompletude e incerteza dos dados para observação do domínio de aplicação; (ii) risco de discriminação (de gênero, raça, social etc.), na medida que os dados de treinamento podem incorporar vieses sociais estruturais (presentes nos dados disponíveis usados para treinamento ou no design do modelo) ou vieses cognitivos dos desenvolvedores humanos, (iii) risco à privacidade e proteção de dados, na medida em que os dados de treinamento incorporam dados pessoais e o poder de inferência da IA é capaz de extrair informações para finalidades diversas da coleta e (iv) risco de transparência, tendo em vista a dificuldade em identificar critérios determinantes para tomada de decisão, predição ou recomendação automatizada ou interpretar causalmente as correlações

¹¹ ZAO-SANDERS, Marc. **100 Applications of Generative AI**: How people are really using the technology in the wild. Disponível em: <<https://learn.filtered.com/thoughts/ai-now-report>>. Acesso em 15 de jul. 2024.

¹² CEIBAL. Building Artificial Intelligence for Education. 2024. Disponível em: <<https://ceibal.edu.uy/wp-content/uploads/2024/05/Building-Artificial-Intelligence-for-education.pdf>>. Acesso em 20 de set. 2024. p. 10.

¹³ CEIBAL. Building Artificial Intelligence for Education. 2024. Disponível em: <<https://ceibal.edu.uy/wp-content/uploads/2024/05/Building-Artificial-Intelligence-for-education.pdf>>. Acesso em 20 de set. 2024. p. 19.

estatísticas encontradas, principalmente em modelos complexos de aprendizado de máquina, como redes neurais.¹⁴

Portanto, não se ignora que o emprego das IAGs para fins acadêmicos importa reconhecer diversos desafios, alguns ainda desconhecidos, que surgirão à medida em que se fizer uso da ferramenta.

Nesse sentido, a colaboração é essencial para assegurar que a tecnologia seja usada de forma ética, de forma a promover equilíbrio entre o potencial da IAG e a experiência e julgamentos humanos.¹⁵

O presente relatório está dividido da seguinte forma. Para aqueles que não estão familiarizados com temas relacionados aos riscos e à governança de sistemas de Inteligência Artificial, elaboramos três anexos. O ANEXO I traz uma introdução à Governança de Sistemas de Inteligência artificial como Gestão de Riscos. O ANEXO II faz uma introdução aos sistemas de IA Generativa, em particular os geradores de textos. Já o ANEXO III traz uma introdução aos riscos associados ao uso de sistemas de IA Generativa de textos, indicando as medidas adequadas de governança atualmente indicadas, conforme o atual estado da tecnologia.

É possível, porém, seguir a leitura, principalmente para aqueles já com alguma familiaridade com o tema, caminhando imediatamente para a Seção 2 deste Capítulo 1, em que abordaremos medidas de governança para uso de IAs Generativas que vem sendo praticadas nas dez melhores universidades do mundo, de acordo com o ranking Times Higher Education e que adotaram diretrizes ou *guidelines* para seu uso. Na sequência, a Seção 1 do Capítulo 2 descreve a metodologia da pesquisa por adesão sobre o uso de IAGs na Faculdade de Direito da USP. Na Seção 2, analisamos os dados provenientes do

¹⁴ MARANHÃO, JULIANO; NAVAS, JOÃO. **Certificação como instrumento de regulação da Inteligência Artificial no AI Act**. In Comentários ao EU AI Act. Rony Vainzof, Andriei Gutierrez, Gustavo Godinho e Alexandra Krastins (Coords.) Publicação prevista para 2024.

¹⁵ CEIBAL. Building Artificial Intelligence for Education. 2024. Disponível em: <<https://ceibal.edu.uy/wp-content/uploads/2024/05/Building-Artificial-Intelligence-for-education.pdf>>. Acesso em 20 de set. 2024. p. 10.

levantamento e, por fim, tecemos considerações com recomendações básicas sobre o uso de IAGs na Faculdade de Direito da USP.

2. Inteligência Artificial Generativa na educação

2.1. Usos de IAG nas atividades acadêmicas

A IAG pode auxiliar estudantes em várias tarefas, como na escrita de trabalhos e artigos, criação de resumos e paráfrases, criação de mapas de conceitos, seleção de exercícios ou sistematização de debates, correção gramatical, traduções, geração de vídeos sintéticos com professores abordando oralmente conteúdo a ser lido etc. A ferramenta pode também ser útil para professores, ao colaborar em trabalhos escritos, propostas de exercícios e atividades didáticas, conduzir pesquisas e auxiliar em tarefas administrativas, como redação de relatórios.¹⁶ Por outro lado, surgem preocupações como integridade acadêmica e o fato de que habilidades antes valorizadas no ambiente escolar – como ler e escrever – podem ser afetadas.

O uso de IA para atividades acadêmicas possui as mesmas limitações e pode trazer os mesmos riscos que já apresentamos acima, além de trazer algumas preocupações adicionais.

Há a constatação de que, em grande medida, é impossível impedir ou proibir o uso de ferramentas de IAG. Além disso, uma vez à medida de ferramentas como o ChatGPT são incorporadas às esferas profissionais de competências, às quais certamente os alunos estarão inseridos, pode ser mais interessante incorporar a ferramenta do que esperar que os alunos evitem completamente a IA em seus estudos.¹⁷⁻¹⁸

Tratando-se de trabalhos e tarefas acadêmicas, as capacidades da IAG se sobrepõem em um grau incomparável a qualquer tecnologia anterior. É nesse

¹⁶ ATLAS, S. (2023). **ChatGPT for higher education and professional development**: A guide to conversational AI. Retrieved March 12, 2023. Disponível em: <https://digitalcommons.uri.edu/cba_facpubs/548/>. Acesso em 15 de jul. 2024. p. 24

¹⁷ WILLIAMS, C. (2023). **Hype, or the future of learning and teaching?** 3 Limits to AI's ability to write student essays. London School of Economics internet blog. Disponível em: <<https://kar.kent.ac.uk/99505/>>. Acesso em 11 de jul. de 2024. p. 3

¹⁸ BAIDOO-ANU, David; OWUSU ANSAH, Leticia. **Education in the Era of Generative Artificial Intelligence (AI)**: Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning (January 25, 2023). Disponível em: <<https://ssrn.com/abstract=4337484>>. Acesso em 10 de jul. 2024. p. 59

sentido que diversas instituições de ensino superior desaconselham a instituição de proibições gerais das IAGs.¹⁹

A IAG pode ser usada para diversos fins acadêmicos. A Universidade de Harvard²⁰ aponta alguns exemplos:

(i) **Criação de novos conteúdos:** a IAG pode criar planos de estudos, planos de aula e pesquisas a partir de instruções simples. Pode gerar novas ideias ou ajudar o usuário a liberar sua criatividade. Aqui, será relevante a escrita (ou como é comumente chamado, a “engenharia”) de *prompts*.

(ii) **Geração ou correção de códigos de computador:** IAGs podem gerar código de computador, identificar e corrigir *bugs* e melhorar a qualidade e a eficiência do código usando *prompts* de linguagem natural.

(iii) **Criar fórmulas de planilhas:** ferramentas de IAG podem criar fórmulas para Excel e outros programas de criação de planilha.

(iv) **Pesquisa:** ferramentas de IAG podem filtrar rapidamente informações de muitas fontes diferentes em razão dos dados em que foram treinadas, e cada vez mais mecanismos de pesquisa populares como Google e Bing estão incorporando resultados gerados por IA em seus produtos.

(v) **Resumo de documentos:** ferramentas de AIG podem analisar e gerar resumos concisos de documentos longos. Isso faz com que sejam úteis em tarefas como destacar as principais ideias de artigos de pesquisa, resumir e-mails, relatórios e notas de reuniões ou identificar as principais descobertas de dados de planilhas.

(vi) **Sintetizar informações:** Ferramentas de IAG podem analisar conjuntos de dados complexos e sintetizar as suas informações. Um bom exemplo desse uso é a análise de respostas de uma pesquisa, de forma a obter

¹⁹ UNIVERSIDADE DE OXFORD. **Beyond ChatGPT: State of AI in Academic Practice.** Outubro de 2023. Disponível em: <https://www.cctl.ox.ac.uk/sites/default/files/cctl/documents/media/beyond_chatgpt_-_state_of_ai_for_autumn_2023_correct.pdf>. Acesso em 6 de jul. 2024.

²⁰ UNIVERSIDADE DE HARVARD. **Ideas for experimenting with Generative AI: Use cases and things to keep in mind.** Disponível em: <<https://huit.harvard.edu/news/ai-use-cases>>. Acesso em: 6 de ago. 2024.

um resumo geral do que as pessoas estão dizendo ou escolher as respostas mais comuns, palavras-chave e frases.

(vii) **Traduções:** ferramentas de AIG podem traduzir longas passagens de textos.

(viii) **Revisão ou aprimoramento de seu trabalho:** a IAG pode ser útil na edição e aprimoramento da escrita. É possível pedir à IA que sugira melhorias, procure erros, avalie o fluxo e a legibilidade, bem como faça sugestões de conteúdo.

(ix) **Aumento de produtividade:** As ferramentas de IAG podem resumir notas de reuniões e destacar conclusões importantes.

(x) **Melhorar acessibilidade:** ferramentas de IAG podem ajudar a tornar o conteúdo mais acessível, identificando problemas em seus documentos, sugerindo melhorias como texto alternativo e descrições de áudio para imagens ou convertendo fala em texto para legendas.

A lista acima é meramente exemplificativa. Ferramentas de IAG podem ser usadas para criar estruturas de artigos científicos e criar imagens por meio de instruções de textos para sua inclusão em trabalhos.²¹ Além disso, atualmente existem ferramentas de IAG que prometem ser um “assistente” na escrita de trabalhos científicos, ajudando desde a escrita do trabalho até as citações e edição.²²

Por envolver diversas competências estudantis, o uso de IAGs para fins acadêmicos traz preocupação quanto à autonomia e ao desenvolvimento de habilidades importantes, como análise crítica, pensamento criativo e pesquisa rigorosa.²³ Ainda, como já apontado, os resultados das IAGs contêm limitações, inconsistências e potenciais vieses. Uma dessas preocupações é justamente a

²¹ JISC. **AI in tertiary education:** A summary of the current state of play. Third edition - September 2023. Disponível em: <<https://repository.jisc.ac.uk/9232/1/ai-in-tertiary-education-a-summary-of-the-current-state-of-play-september-2023.pdf>>. Acesso em 1 de out. 2024.

²² Diversos são os exemplos de plataformas criadas cujo objetivo é auxiliar na redação de trabalho acadêmicos. Alguns exemplos que podemos citar são: Jenni AI, Samwell.ai, Research Rabbit, entre outros.

²³ UNIVERSIDADE DE CAMBRIDGE. Using Generative AI: General Principles for use of Generative AI. Disponível em: <<https://blendedlearning.cam.ac.uk/guidance-support/ai-and-education/using-generative-ai>>. Acesso em 1 de out. 2024

criação de uma dependência excessiva dessas ferramentas, de forma que elas venham a limitar o desenvolvimento do pensamento crítico e das habilidades de resolução de problemas.²⁴

Parte da educação universitária consiste em adquirir certas habilidades como a assimilação de conteúdo, construção de argumentos com base em evidências, e a capacidade de expressar-se de forma clara e coerente.²⁵ Assim, alunos que fazem uso de ferramentas de IAG devem utilizá-las como auxílio para adquirir essas habilidades, interagindo com os sistemas e não apenas delegando tarefas.

Diversos estudos propõem repensar a forma pela qual as avaliações são feitas, com objetivo de exigir que os alunos demonstrem resultados de aprendizagem de ordem superior. Ou seja, avaliações que proporcionem desafios acadêmicos robustos para os alunos em uma variedade de tarefas e condições.²⁶ Além disso, os alunos aprendem interagindo ativamente com o material do curso. E essa interação também poderá se dar por meio das próprias atividades que as ferramentas de IAG “automatizam”, como resumir um argumento, sintetizar informações ou redigir um esboço inicial.²⁷

No entanto, como a IAG é novidade no ambiente acadêmico, os estudantes não têm clareza sobre como utilizá-la. Esclarecer as expectativas em relação ao uso da IAG inclui explicar se e como os alunos têm permissão para usar essas ferramentas, e como devem ser transparentes com o uso da IAG em seus trabalhos. Embora se possa proibir o uso de IAGs, não se pode evitá-la. A solicitação para que os alunos não usem as ferramentas de IAG, ou usem-nas

²⁴ UNIVERSIDADE DE CAMBRIDGE. Using Generative AI: General Principles for use of Generative AI. Disponível em: <<https://blendedlearning.cam.ac.uk/guidance-support/ai-and-education/using-generative-ai>>. Acesso em 1 de out. 2024

²⁵ UNIVERSIDADE DE OXFORD. **Use of generative AI tools to support learning**. Disponível em: <<https://www.ox.ac.uk/students/academic/guidance/skills/ai-study>>. Acesso em 2 de out. 2024

²⁶ UNIVERSIDADE DE OXFORD. **Four lessons from ChatGPT: Challenges and opportunities for educators**. Disponível em: <<https://www.ctl.ox.ac.uk/article/four-lessons-from-chatgpt-challenges-and-opportunities-for-educators>>. Acesso em 27 de set. 2024.

²⁷ UNIVERSIDADE DE PRINCETON. **Rights, Rules, Responsibilities: Academic Regulations**. Disponível em: <<https://rrr.princeton.edu/students-and-university/24-academic-regulations>>. Acesso em 6 de ago. 2024.

apenas em situações específicas, pressupõe que aceitem essa restrição, por meio da compreensão dos riscos da ferramenta e da importância da restrição para seu desenvolvimento intelectual.²⁸

2.2. Governança de IAG nas editoras científicas

Preocupadas com o uso de IAGs na elaboração de artigos submetidos a publicações científicas, uma série de editoras de periódicos atualizaram suas regras de publicação e divulgaram princípios para lidar com o uso dessa nova tecnologia.

Trazemos aqui o exemplo de três grandes editoras acadêmicas (Elsevier, Springer e Oxford University Press), de forma a apresentar as principais preocupações e orientações adotadas, elaborados com o objetivo de fornecer maior transparência e orientação aos autores, leitores, revisores, editores em relação à IA Generativa e às tecnologias assistidas por IA.

As questões suscitadas pelas editoras dizem respeito à atribuição e reconhecimento de autoria, à responsabilidade quanto aos resultados gerados e à transparência pelos autores quanto ao uso de IAG. Além disso, há debate sobre quais ferramentas e tarefas podem ser usadas livremente e quais tarefas trazem preocupação, merecendo regramentos próprios.

A Elsevier, por exemplo, destaca que sua política é específica para o uso IA Generativa e ferramentas assistidas por IA capazes conteúdo original para publicação. Assim, a política não se aplicaria a ferramentas como verificadores ortográficos ou gramaticais, nem gerenciadores de referências que permitem aos autores coletar, organizar, anotar e usar referências a artigos acadêmicos (como Mendeley, EndNote, Zotero e outros). Essas ferramentas podem ser utilizadas pelos autores sem necessidade de divulgação pelos autores.²⁹

²⁸ UNIVERSIDADE DE NOVA YORK (NYU). **Teaching with Generative AI**. Disponível em: <<https://www.nyu.edu/faculty/teaching-and-learning-resources/teaching-with-generative-tools.html>>. Acesso em: 28 de set. 2024.

²⁹ ELSEVIER. **The use of generative AI and AI-assisted technologies in writing for Elsevier**. Disponível em: <<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>>. Acesso em 14 nov. 2024.

Também é permitido o uso de ferramentas de IA Generativa que permitem melhorar a legibilidade e a linguagem do trabalho, que não substituam tarefas essenciais de autoria, como se dá com a produção de conhecimentos científicos, pedagógicos ou médicos, bem como tirar conclusões científicas ou fornecer recomendações clínicas.³⁰

Também aponta a Elsevier que sua política não impede o uso de IA Generativa e de ferramentas assistidas por IA em projetos formais de pesquisa ou métodos de pesquisa. Contudo, sendo este o caso, a editora aponta a necessidade de que o uso da IA seja descrito como parte da metodologia do trabalho.³¹

No mesmo sentido a Springer entende que ferramentas de IA Generativa usadas para fins de promover melhorias no texto, como facilitar a leitura e o estilo, ou para garantir que os textos estejam livres de erros de gramática, ortografia, pontuação e tom, não precisam ser apontados ou declarados no trabalho. Isso porque tais melhorias assistidas, ainda que resultem na alteração da redação e formatação dos textos, não incluem trabalho editorial generativo e de criação autônoma de conteúdo.³²

Quanto à análise dos principais pontos suscitados pelas editoras em seus guias, a Oxford University Press traz relevante questão, abordada também no ANEXO III, referente ao plágio realizado pela própria ferramenta de IAG. Assim, a OUP destaca que os autores não devem inserir como *prompt* (texto de requisição) ou fazer o upload de seus textos em qualquer ferramenta, o que inclui até mesmo os rascunhos de manuscritos. Isso porque a IA generativa pode reter e ser

³⁰ ELSEVIER. **The use of generative AI and AI-assisted technologies in writing for Elsevier.** Disponível em: <<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>>. Acesso em 14 nov. 2024.

³¹ ELSEVIER. **The use of generative AI and AI-assisted technologies in writing for Elsevier.** Disponível em: <<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>>. Acesso em 14 nov. 2024.

³² SPRINGER. **Artificial Intelligence (AI).** Disponível em: <<https://www.springer.com/gp/editorial-policies/artificial-intelligence--ai-/25428500?srsId=AfmBOorboLOdr98567gYbe1x5HHdH2aSyjmQ70e91hGYDvs1B9th21fP>>. Acesso em 13 nov. 2024.

treinada com base nas informações inseridas e, como consequência, levar a plágio daquele conteúdo.³³

Em relação aos três temas principais de preocupação, quais sejam, autoria, responsabilidade e transparência, as editoras trazem os seguintes parâmetros:

(i) **Autoria:**

Para a OUP, a IAG não pode ser qualificada como “autor”, não devendo, portanto, ser usada para realizar as principais tarefas autorais, como criação de argumentos e *insights* científicos, escrita de análises ou conclusões.³⁴ A Springer também destaca que não é possível colocar ferramentas de IAG na condição de autor, uma vez que a atribuição de autoria acarreta responsabilidades pelos trabalhos.³⁵

No mesmo sentido, para a Elsevier, a autoria implica responsabilidades e tarefas que só podem ser atribuídas e executadas por seres humanos. Cada (co-)autor é responsável por garantir que questões relacionadas à precisão ou à integridade de qualquer parte do trabalho sejam adequadamente investigadas e resolvidas, de modo que a autoria pressupõe a capacidade de aprovar a versão final do trabalho e concordar com sua submissão. Os autores também são responsáveis por garantir que o trabalho seja original, que os autores declarados se qualifiquem para a autoria e que o trabalho não infrinja direitos de terceiros.³⁶

³³ OXFORD UNIVERSITY PRESS. **Author use of generative Artificial Intelligence (AI)**. Disponível em: <<https://academic.oup.com/pages/authoring/books/author-use-of-artificial-intelligence>>. Acesso em 13 nov. 2024.

³⁴ OXFORD UNIVERSITY PRESS. **Author use of generative Artificial Intelligence (AI)**. Disponível em: <<https://academic.oup.com/pages/authoring/books/author-use-of-artificial-intelligence>>. Acesso em 13 nov. 2024.

³⁵ SPRINGER. **Artificial Intelligence (AI)**. Disponível em: <<https://www.springer.com/gp/editorial-policies/artificial-intelligence-ai-/25428500?srsId=AfmBOorboLOdr98567gYbe1x5HHdH2aSyjmQ70e91hGYDvs1B9th21fP>>. Acesso em 13 nov. 2024.

³⁶ ELSEVIER. **The use of generative AI and AI-assisted technologies in writing for Elsevier**. Disponível em: <<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>>. Acesso em 14 nov. 2024.

(ii) Responsabilidade

De uma forma bastante próxima do quanto apontado com relação às atribuições da autoria, as editoras consideram os autores responsáveis pela precisão, integridade e originalidade de seus trabalhos, o que inclui qualquer conteúdo gerado por IA que ele possa conter.

A OUP aponta que compete aos próprios autores verificar a exatidão e adequação de seus trabalhos, sendo responsáveis por qualquer violação ética e potenciais preconceitos ou conflitos de interesse publicados. Também a Elsevier destaca em seu guia que os autores são pessoalmente responsáveis por seu trabalho, apontando que o uso da IAG deve ser feito com supervisão e controle humano, devendo todo o trabalho ser revisado e editado, ante o risco de que a ferramenta gere resultados incorretos, incompletos ou tendenciosos.³⁷

(iii) Transparência

Para a OUP, o uso de IAG, seja na criação ou preparação de conteúdo, deve ser divulgado ao editor e devidamente citado no texto e/ou em notas de rodapé ou finais. Além disso, a OUP exige que autores recebam permissão por escrito da editora para entregar conteúdo gerado por IA, o que inclui a coleta e análise de dados ou a produção de elementos gráficos do texto, como parte da submissão de seus trabalhos. A editora também aponta que os autores serão obrigados a substituir conteúdo gerado por IA por conteúdo gerado por humanos, caso a OUP entenda necessário.³⁸

A Elsevier também prevê para seus autores a obrigação de divulgar em seus manuscritos o uso de IAG e tecnologias de IA. A editora pede que autores que usaram IAG insiram uma declaração no final de seu manuscrito, imediatamente acima das referências ou bibliografia intitulada “Declaração de IA

³⁷ ELSEVIER. **The use of generative AI and AI-assisted technologies in writing for Elsevier.** Disponível em: <<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>>. Acesso em 14 nov. 2024.

³⁸ OXFORD UNIVERSITY PRESS. **Author use of generative Artificial Intelligence (AI).** Disponível em: <<https://academic.oup.com/pages/authoring/books/author-use-of-artificial-intelligence>>. Acesso em 13 nov. 2024.

generativa e tecnologias assistidas por IA no processo de escrita”, em que deve ser especificada a ferramenta que foi utilizada e o motivo de sua utilização, bem como declarando que o conteúdo teria sido revisado e editado.³⁹

2.3. Governança do uso de IAG nas universidades

Uma série de universidades tem divulgado diretrizes sobre o uso de IAG tanto por docentes quanto discentes. Em geral, não se proíbe o uso dessas ferramentas, mas incentiva-se tanto a transparência, quanto a adoção de programas de treinamento e de docentes e discentes, quanto às potencialidades das novas ferramentas, suas limitações e os riscos em seu emprego, de modo a enfrentar adequadamente os desafios trazidos.⁴⁰

A Universidade de Cambridge, por exemplo, informa que criará guia de treinamento interno, bem como investirá em treinamento sobre escrita de *prompts* e verificação de fatos.⁴¹

No Reino Unido, o “Russel Group”, que reúne 24 universidades, adotou um conjunto de princípios, de forma a ajudar universidades a garantir que estudantes e funcionários sejam “alfabetizados em IA” (no original, “AI literate”), de forma a permitir que eles possam capitalizar as oportunidades que os avanços tecnológicos proporcionam para o ensino e a aprendizagem.⁴²

Os cinco princípios são os seguintes:

1. As universidades apoiarão estudantes e funcionários para que estes se tornem “letrados em IA”.

³⁹ ELSEVIER. **The use of generative AI and AI-assisted technologies in writing for Elsevier.** Disponível em: <<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>>. Acesso em 14 nov. 2024.

⁴⁰ Um exemplo é a Universidade de Stanford, que disponibilizou um guia de ensino. UNIVERSIDADE DE STANFORD. **Artificial Intelligence Teaching Guide.** Disponível em <<https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide>>. Acesso em 1 de out. 2024.

⁴¹ UNIVERSIDADE DE CAMBRIDGE. **How we use generative AI tools.** Disponível em: <<https://www.communications.cam.ac.uk/generative-ai-tool-guidelines>>. Acesso em 6 de ago. 2024.

⁴² RUSSEL GROUP. **New principles on use of AI in education.** 4 de julho de 2023. Disponível em: <<https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/>>. Acesso em 2 de ago. 2024.

2. Os funcionários devem estar preparados para apoiar os alunos na utilização eficaz e adequada de ferramentas de IAG em sua experiência de aprendizagem.
3. As universidades adaptarão o ensino e formas de avaliação para incorporar o uso ético da IAG e apoiarão a igualdade de acesso.
4. As universidades garantirão que o rigor acadêmico e a integridade sejam mantidos.
5. As universidades trabalharão em colaboração para partilhar as melhores práticas à medida que a tecnologia e a sua aplicação na educação evoluem.

É cada vez mais frequente a publicação de guias por universidades e institutos de pesquisa para indicar usos apropriados e traçar limites para uso das ferramentas de IAG. De forma a apontar quais foram suas principais conclusões, analisamos os guias de melhores práticas publicados pelas 10 melhores universidades do mundo de acordo com a Times Higher Education.⁴³

Dois pontos são consistentemente destacados em todos os guias analisados: (i) reconhecimento de que o uso legítimo pode ser benéfico para o aprendizado e trabalho acadêmico, e (ii) foco na integridade acadêmica.

De fato, diversas faculdades se manifestam de forma abertamente positiva, incentivando ao uso de ferramentas de IAG. A Universidade de Chicago, por exemplo, reconhece que a IAG oferece muitos recursos e eficiências que podem aprimorar o trabalho acadêmico.⁴⁴ A Universidade de Harvard confirma que a apoia a experimentação responsável com ferramentas de IAG⁴⁵, e Oxford declara seu desejo de apoiar alunos e professores a usar a IA de forma ética e

⁴³ TIMES HIGHER EDUCATION. **World University Rankings 2024**. Disponível em: <<https://www.timeshighereducation.com/world-university-rankings/2024/world-ranking>>. Acesso em 6 de ago. 2024.

⁴⁴ UNIVERSIDADE DE CHICAGO. **Generative Artificial Intelligence (AI) Guidance**. Disponível em: <<https://its.uchicago.edu/generative-ai-guidance/>>. Acesso em 17 de set. 2024.

⁴⁵ UNIVERSIDADE DE HARVARD. **Initial guidelines for the use of Generative AI tools at Harvard**. Disponível em: <<https://huit.harvard.edu/ai/guidelines#block-boxes-1694711840>>. Acesso em 6 de ago. 2024.

apropriada em seu trabalho e a considerar a alfabetização em IA como uma habilidade fundamental.⁴⁶

Contudo, os guias variam quanto ao detalhamento das preocupações e soluções apresentadas. A grande maioria se resume a algumas páginas de sites apontando os riscos, instruções simples, e consequências para o descumprimento.

Recomenda-se, em geral, que as regras sejam simples e consideravelmente abertas, pelas seguintes razões:

- (i) a tecnologia está em constante evolução, não sendo possível determinar, neste momento, todas as preocupações que podem vir a surgir pelo uso da IAG. Para citar alguns exemplos, a Universidade de Columbia⁴⁷ e a Universidade de Cambridge⁴⁸ apontam que as regras podem vir a mudar com o tempo. A Universidade de Harvard destaca que irá desenvolver e atualizar suas políticas conforme entenderem melhor as implicações de usar ferramentas de IAG.⁴⁹ Também a Universidade de Stanford destaca que a legalidade e implicações éticas de como a IA vem se desenvolvendo e é usada ainda está em evolução e, portanto, os alunos devem estar cientes disso ao analisarem os presentes recursos e políticas disponibilizados pela instituição sobre o assunto.⁵⁰

⁴⁶ UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <<https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024.

⁴⁷ UNIVERSIDADE DE STANFORD. Generative AI Policy Guidance. Guidance adopted on February 16, 2023. Disponível em: <<https://communitystandards.stanford.edu/generative-ai-policy-guidance>>. Acesso em 6 de ago. 2024.

⁴⁸ UNIVERSIDADE DE CAMBRIDGE. **How we use generative AI tools**. Disponível em: <<https://www.communications.cam.ac.uk/generative-ai-tool-guidelines>>. Acesso em 6 de ago. 2024.

⁴⁹ UNIVERSIDADE DE HARVARD. **Initial guidelines for the use of Generative AI tools at Harvard**. Disponível em: <<https://huit.harvard.edu/ai/guidelines#block-boxes-1694711840>>. Acesso em 6 de ago. 2024.

⁵⁰ UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford**: Enabling innovation through AI best practices. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

- (ii) cursos específicos possuem preocupações diversas quanto ao uso de ferramentas de IAG. As preocupações da faculdade de medicina podem variar das preocupações da faculdade de engenharia. Assim, apontar apenas regras gerais dá às faculdades, ou mesmo a matérias ou cursos individualmente considerados, a liberdade para criar políticas específicas, deixando aos guias apenas as regras consideradas inegociáveis, como a obrigação de ser transparente quanto ao uso das ferramentas de IAG e a forma pela qual elas foram usadas.⁵¹
- (iii) Deve-se respeitar a autonomia universitária dos docentes, de modo que possam optar a dar maiores detalhes sobre como a IAG pode ser usada para diferentes tarefas ou mesmos em trabalhos específicos.^{52 53}
- (iv) apesar da novidade trazida pelas ferramentas da IAG, riscos à integridade acadêmica e preocupações quanto a aprendizado e didática envolvendo novas tecnologias não são novos ou únicos.⁵⁴ Para enfrentar essas velhas e novas preocupações, políticas de honestidade acadêmica devem ser consideradas e podem perfeitamente endereçar esses desafios.⁵⁵

⁵¹ UNIVERSIDADE DE COLUMBIA – Columbia Business Scholl. **Generative AI Policy**. Disponível em: <<https://students.business.columbia.edu/office-of-student-affairs/academic-advising-and-student-success/academic-integrity/generative-ai-policy#:~:text=The%20use%20of%20generative%20Artificial,the%20course%20instructor%20granting%20permission>>. Acesso em 2 de out. 2024.

⁵² UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <<https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024

⁵³ No original: "First and foremost, the decision to allow, limit, or prohibit generative AI in a course or in undergraduate independent work will remain our faculty's. Faculty members have the discretion to set their own generative AI policy for their courses. We encourage each instructor to establish a clear policy and post it on their syllabus." UNIVERSIDADE DE PRINCETON. **Generative AI in Teaching and Learning Fall 2023, Faculty and Student Memo**. Disponível em: <<https://odoc.princeton.edu/about/official-deans-communications/2023/generative-ai-teaching-and-learning-fall-2023-faculty-and>>. Acesso em 6 de ago. 2024.

⁵⁴ UNIVERSIDADE DE OXFORD. Four lessons from ChatGPT: Challenges and opportunities for educators. Disponível em: <<https://www.ctl.ox.ac.uk/article/four-lessons-from-chatgpt-challenges-and-opportunities-for-educators>>. Acesso em 27 de set. 2024.

⁵⁵ No original: *For all the anxiety about academic integrity, generative AI presents no new philosophical issues. Students are not to take credit for work they did not do, whether that work is copied from elsewhere, or created for them from scratch. AI use looks like cut-and-paste, because it is automated, but AI use works like existing ghostwriting services, often called 'contract cheating'.*

A tabela abaixo sintetiza o levantamento de medidas de governança, em relação à privacidade, adotadas pelas dez melhores universidades indicadas pelo Times Higher Education para o ano de 2024, mostrando o consenso com relação à adoção de medidas de governança para proteger a privacidade, para assegurar a adequada revisão do conteúdo gerado e transparência entre discentes e docentes.⁵⁶ Há, também, medidas disciplinares em relação ao uso não transparente ou para atividades em que o uso da ferramenta não é autorizado. Embora haja um certo consenso em relação à adoção de medidas para dissuadir o uso não autorizado ou não transparente, há divergência sobre como o uso de ferramentas de IAG deve ser tratado. Algumas inserem o “mau” uso de IAG na mesma categoria de medidas disciplinares para plágio, outras o encaram como formas de auxílio por terceiro, ou como formas de violação à integridade acadêmica.

Tabela 1: Principais preocupações suscitadas nos guias das dez Melhores Universidades Indicadas Pelo Times Higher Education em 2024

UNIVERSIDADE	PREOCUPAÇÕES SUSCITADAS			PENALIDADE PELO USO NÃO AUTORIZADO
	PROTEÇÃO À PRIVACIDADE	NECESSIDADE DE REVISÃO	NECESSIDADE DE TRANSPARÊNCIA	
OXFORD ⁵⁷⁻⁵⁸⁻⁵⁹⁻⁶⁰	SIM	SIM	SIM	SIM. SIMILAR AO PLÁGIO

UNIVERSIDADE DE NOVA YORK (NYU). **Teaching with Generative AI**. Disponível em: <<https://www.nyu.edu/faculty/teaching-and-learning-resources/teaching-with-generative-tools.html>>. Acesso em: 28 de set. 2024.

⁵⁶ TIMES HIGHER EDUCATION. **World University Rankings 2024**. Disponível em: <<https://www.timeshighereducation.com/world-university-rankings/2024/world-ranking>>. Acesso em 6 de ago. 2024.

⁵⁷ UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <<https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024

⁵⁸ UNIVERSIDADE DE OXFORD. **Use Generative AI services such as ChatGPT safely**. Disponível em: <<https://infosec.web.ox.ac.uk/use-generative-ai-services-such-as-chatgpt-safely>>. Acesso em 6 de ago. 2024.

⁵⁹ UNIVERSIDADE DE OXFORD. **Use of generative AI tools to support learning**. Disponível em: <<https://www.ox.ac.uk/students/academic/guidance/skills/ai-study>>. Acesso em 2 de out. 2024

⁶⁰ UNIVERSIDADE DE OXFORD. **An introduction to the use of generative AI tools in teaching**: ten ideas on the use of AI to support students in their learning, together with suggested responses for teaching. Disponível em: <<https://www.ctl.ox.ac.uk/ai-tools-in-teaching>> Acesso em 2 de out. 2024

UNIVERSIDADE	PREOCUPAÇÕES SUSCITADAS			PENALIDADE PELO USO NÃO AUTORIZADO
	PROTEÇÃO À PRIVACIDADE	NECESSIDADE DE REVISÃO	NECESSIDADE DE TRANSPARÊNCIA	
STANFORD ⁶¹⁻⁶²	SIM	SIM	SIM	SIM. SIMILAR AO AUXÍLIO POR TERCEIROS
MIT ⁶³⁻⁶⁴	SIM	SIM	SIM	SIM. SUJEITO ÀS REGRAS DE INTEGRIDADE ACADÊMICA/PLÁGIO
HARVARD ⁶⁵⁻⁶⁶⁻⁶⁷	SIM	SIM	NÃO	NÃO MENCIONA
CAMBRIDGE ⁶⁸⁻⁶⁹⁻⁷⁰	SIM	SIM	SIM	SIM, SIMILAR AO PLÁGIO
PRINCETON ⁷¹⁻⁷²⁻⁷³	SIM	SIM	SIM	SIM, SIMILAR À DESONESTIDADE ACADÊMICA

⁶¹ UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford**: Enabling innovation through AI best practices. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

⁶² UNIVERSIDADE DE STANFORD. **Generative AI Policy Guidance**. Guidance adopted on February 16, 2023. Disponível em: <<https://communitystandards.stanford.edu/generative-ai-policy-guidance>>. Acesso em 6 de ago. 2024.

⁶³ MIT - MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Generative AI & Your Course**. Disponível em: <<https://tll.mit.edu/teaching-resources/course-design/gen-ai-your-course/>>. Acesso em 6 de ago. 2024.

⁶⁴ MIT - MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Generative AI for Teaching & Learning**: Explore our resources to unlock AI's potential for enhancing teaching and learning at MIT Sloan. Disponível em: <<https://mitsloanedtech.mit.edu/ai/>>. Acesso em 6 de out. 2024.

⁶⁵ UNIVERSIDADE DE HARVARD. **Initial guidelines for the use of Generative AI tools at Harvard**. Disponível em: <<https://huit.harvard.edu/ai/guidelines#block-boxes-1694711840>>. Acesso em 6 de ago. 2024.

⁶⁶ UNIVERSIDADE DE HARVARD. **Generative AI Tool Comparison**. Disponível em: <<https://huit.harvard.edu/ai/tools>>. Acesso em 6 de ago. 2024.

⁶⁷ UNIVERSIDADE DE HARVARD. **Ideas for experimenting with Generative AI: Use cases and things to keep in mind**. Disponível em: <<https://huit.harvard.edu/news/ai-use-cases>>. Acesso em: 6 de ago. 2024.

⁶⁸ UNIVERSIDADE DE CAMBRIDGE. **How we use generative AI tools**. Disponível em: <<https://www.communications.cam.ac.uk/generative-ai-tool-guidelines>>. Acesso em 6 de ago. 2024.

⁶⁹ UNIVERSIDADE DE CAMBRIDGE. **Plagiarism and Academic Misconduct**. Disponível em: <<https://www.plagiarism.admin.cam.ac.uk/what-academic-misconduct/artificial-intelligence>>. Acesso em 1 de out. 2024

⁷⁰ UNIVERSIDADE DE CAMBRIDGE. **Using Generative AI: General Principles for use of Generative AI**. Disponível em: <<https://blendedlearning.cam.ac.uk/guidance-support/ai-and-education/using-generative-ai>>. Acesso em 1 de out. 2024

⁷¹ UNIVERSIDADE DE PRINCETON. **Generative AI in Teaching and Learning Fall 2023, Faculty and Student Memo**. Disponível em: <<https://odoc.princeton.edu/about/official-deans-communications/2023/generative-ai-teaching-and-learning-fall-2023-faculty-and>>. Acesso em 6 de ago. 2024.

⁷² UNIVERSIDADE DE PRINCETON. **Rights, Rules, Responsibilities: Academic Regulations**. Disponível em: <<https://rrr.princeton.edu/students-and-university/24-academic-regulations>>. Acesso em 6 de ago. 2024.

⁷³ UNIVERSIDADE DE PRINCETON. **Generative AI**. Disponível em: <<https://mcgraw.princeton.edu/generative-ai>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE	PREOCUPAÇÕES SUSCITADAS			PENALIDADE PELO USO NÃO AUTORIZADO
	PROTEÇÃO À PRIVACIDADE	NECESSIDADE DE REVISÃO	NECESSIDADE DE TRANSPARÊNCIA	
CALTECH ⁷⁴	SIM	SIM	SIM	SIM. SIMILAR AO PLÁGIO E DESONESTIDADE ACADÊMICA
IMPERIAL COLLEGE LONDON ⁷⁵⁻⁷⁶	SIM	SIM	SIM	SIM. SUJEITO ÀS REGRAS DE INTEGRIDADE ACADÊMICA
BERKELEY ⁷⁷⁻⁷⁸	SIM	SIM	SIM	SIM. SIMILAR AO PLÁGIO
YALE ⁷⁹⁻⁸⁰	SIM	SIM	SIM	SIM. SIMILAR AO PLÁGIO

Os documentos destacam o papel fundamental dos educadores para incorporar a visão de que IAGs podem ser úteis como assistentes para personalizar experiências de aprendizagem. Implementar a IA de forma responsável e em parceria com a experiência humana é apontado como vetor para estimular uma educação altamente eficaz e profundamente centrada no ser humano.⁸¹ Para tanto, recomenda-se que as instituições de ensino invistam no treinamento de seu corpo estudantil quanto ao uso responsável de ferramentas

⁷⁴ CALTECH - CALIFORNIA INSTITUTE OF TECHNOLOGY. **Guidance on the Use of Generative AI and Large Language Model Tools**. 22 de outubro de 2023. Disponível em: <<https://www.caltech.edu/campus-life-events/campus-announcements/guidance-on-the-use-of-generative-ai-and-large-language-model-tools>>. Acesso em 6 de ago. 2024.

⁷⁵ IMPERIAL COLLEGE LONDON. **AI & Education Guidance Hub**. Disponível em: <<https://www.imperial.ac.uk/about/education/resources/ai-education-hub/>>. Acesso em 6 de ago. 2024.

⁷⁶ IMPERIAL COLLEGE LONDON. **Introduction to AI**. Disponível em: <<https://www.imperial.ac.uk/about/education/resources/ai-education-hub/introduction-to-ai/>>. Disponível em: 1 de out. 2024.

⁷⁷ UNIVERSIDADE DE BERKELEY, CALIFORNIA. **Appropriate Use of Generative AI Tools**. Disponível em: <<https://oercs.berkeley.edu/privacy/privacy-resources/appropriate-use-generative-ai-tools>>. Acesso em 6 de ago. 2024.

⁷⁸ UNIVERSIDADE DE BERKELEY, CALIFORNIA. **Understanding AI Writing Tools and Their Uses for Teaching and Learning at UC Berkeley**. Disponível em: <<https://teaching.berkeley.edu/understanding-ai-writing-tools-and-their-uses-teaching-and-learning-uc-berkeley>>. Acesso em 6 de out. 2024.

⁷⁹ UNIVERSIDADE DE YALE. **Guidelines for the Use of Generative AI Tools**. 20 de setembro de 2023. Disponível em: <<https://provost.yale.edu/news/guidelines-use-generative-ai-tools>>. Acesso em 6 de ago. 2024.

⁸⁰ UNIVERSIDADE DE YALE. **AI Guidance for Teachers**. Disponível em: <<https://poorvucenter.yale.edu/AIguidance>>. Acesso em 6 de ago. 2024.

⁸¹ MOLINA, E.; COBO, C.; PINEDA, J.; ROVNER, H. (2024). **AI revolution in education**: What you need to know. In Digital Innovations in Education. World Bank. p. 7

de IA, incluindo suas implicações éticas e a importância da integridade acadêmica.⁸²

2.4. A preocupação com integridade acadêmica

A principal preocupação em relação aos guias e diretrizes de governança está na proteção e promoção da integridade acadêmica.

O uso indiscriminado de ferramentas de IA Generativa pode comprometer a aquisição de habilidades relevantes para os alunos, como competências de escrita, investigação, pensamento crítico e resolução de problemas.

Além disso, pode haver a geração de resultados pelas ferramentas que possa ser classificado como plágio, a depender dos *prompts* feitos pelos usuários que possam resultar no fenômeno de “regurgitação”. A regurgitação pode ocorrer quando o *prompt* especifica determinado conteúdo que provavelmente foi utilizado na base de dados. Embora o sistema gere os resultados a partir de parâmetros extraídos da base de dados e produza conteúdo efetivamente inédito, que se distingue totalmente de “colagens” de diferentes documentos, alguns trechos do texto ou *output* gerado podem resultar bastante próximos de documentos que compuseram a base de treinamento do sistema, a partir de indicações ou referências ao texto no conteúdo da requisição de geração de texto (*prompt*).

Por outro lado, o plágio pode ser caracterizado não só quando há reprodução *verbatim* do documento original, como também quando o conteúdo veiculado trazer tese ou expressão de ideias que mereceriam o reconhecimento de autoria original, sem citação no documento que o veicula. Em outras palavras, o resultado gerado por uma ferramenta de IAG pode conter plágio, na condição de paráfrase do trabalho autoral. Nesse sentido, as acusações de plágio estariam ligadas à falta de referências precisas às fontes e à atribuição da autoria da pesquisa em geral.⁸³

⁸² MOLINA, E.; COBO, C.; PINEDA, J.; ROVNER, H. (2024). **AI revolution in education**: What you need to know. In Digital Innovations in Education. World Bank. p. 20

⁸³ Testes apontam que a ferramenta de IA Generativa do Google, Gemini (antigamente conhecida por Bard), por exemplo, gera um resultado que contém entre 5% e 45% de plágio.

Assim, embora a IAG gere conteúdo novo, pode falhar em fazer citação de fonte relevante, onde determinadas expressões originais de ideia mereceriam o reconhecimento, dentro da ética acadêmica.

Não é simples para alunos de graduação, ou mesmo para pesquisadores e professores, identificar tais resultados da ferramenta, o que pode fazer com que sejam incorporados no conteúdo final, sem as devidas referências bibliográficas. Para tanto, é importante o treinamento para conscientização e instruções para formulação dos *prompts* de modo a mitigar “regurgitações” ou reprodução de conteúdos autorais que mereceriam citação, além de revisão dos textos gerados. Existem, atualmente, softwares que buscam identificar plágios em textos, que podem ser empregados, mas que podem resultar tanto em falsos negativos quanto em falsos positivos.

A seguir, trazemos alguns exemplos de plataformas disponíveis para detecção de plágio

- plagium.com: é uma ferramenta de detecção de plágio baseada em GPT projetada especificamente para identificar texto gerado por IA que foi copiado ou parafraseado
- semihuman.ai: ferramenta de detecção de texto gerado por IA que, além disso, funciona como ferramenta “humanizadora” de textos gerados por IA. Garante reescrever o texto produzido por IAGs, de forma a torná-lo livre de plágio e indetectável outras ferramentas de detecção.

PLAGIARISMCHECK.ORG. **Google AI chatbot Gemini caught plagiarizing: PlagiarismCheck's research.** 18 de novembro de 2024. Disponível em: <<https://plagiarismcheck.org/blog/google-ai-chatbot-bard-generated-to-45-of-plagiarism-plagiarismchecks-research/>>. Acesso em 18 nov. 2024. Quanto ao ChatGPT, a mesma conclusão também se observa, ainda que com resultados variáveis. Há dados de que 60% dos resultados do ChatGPT contêm plágio, sendo que 45,7% desses resultados continham textos idênticos. AUSTIN, Doug. **Nearly 60 Percent of GPT 3.5 Outputs Contained Plagiarized Content:** Artificial Intelligence Trends. Disponível em: <<https://ediscoverytoday.com/2024/02/26/nearly-60-percent-of-gpt-3-5-outputs-contained-plagiarized-content-artificial-intelligence-trends/>>. Acesso em 18 nov. 2024. Contudo, outros estudos apontam que o ChatGPT parece ser mais eficiente na produção de textos com originalidade e menor chance de conter plágio, ainda que em alguns testes, o resultado tenha apresentado porcentagens altas de conteúdo plagiado. KHALIL, Mohammad; ER, Erkan. **Will ChatGPT get you caught? Rethinking of Plagiarism Detection.** Disponível em: <https://arxiv.org/pdf/2302.04335>. Acesso em 18 nov. 2024.

- bypassGPT: ferramenta de “humanização” de textos produzidos por IAG.

Há quem aponte, ademais, que as ferramentas de detecção de plágio deverão se tornar obsoletas, na medida que o campo acadêmico passe a receber cada vez mais conteúdo gerado ou assistido por IA.⁸⁴

A integridade acadêmica também é afetada quando uma tarefa, destinada à avaliação do aluno ou do pesquisador, perante bancas, emprega ferramentas de IAG de modo não transparente (sem informar o avaliador sobre o uso e aspectos relevantes do uso que afetaram o resultado final). Com a evolução dos métodos didáticos a partir da nova tecnologia, acredita-se que as atividades propostas e avaliação de conteúdo de pesquisa por editoras acadêmicas já levem em consideração o uso de IAG na produção literária, porém deve sempre haver limites e expectativas quanto ao empenho do pesquisador humano e participação relevante no resultado final. Para tanto, é fundamental haver transparência quanto ao uso.

Como aponta a Universidade de Oxford, ferramentas de IAG não cumprem os critérios para serem listados como “autores” ou “coautores”, uma vez que não podem assumir a responsabilidade pelo conteúdo e pela integridade do resultado de uma investigação. Assim, apenas os pesquisadores são responsáveis pela originalidade, validade, confiabilidade e integridade dos resultados de suas pesquisas, incluindo aquelas pesquisas que se usam de ferramentas de IAG.⁸⁵⁻⁸⁶ Assim, qualquer utilização da ferramenta, por exemplo, na coleta, análise e

⁸⁴ DWIVEDI, Yogesh K.; et. al. **Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy**, *International Journal of Information Management*. Volume 71, 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0268401223000233>>. Acesso em 12 de nov. 2024. P. 27

⁸⁵ UNIVERSIDADE DE OXFORD. **Publication and Authorship**. Disponível em: <<https://researchsupport.admin.ox.ac.uk/governance/integrity/publication>>. Acesso em 10 de out. 2024.

⁸⁶ UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <<https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024.

interpretação dos dados, deve ser citada nas secções de métodos ou agradecimentos do resultado, conforme apropriado.⁸⁷.

A transparência ajudará os leitores a entender mais sobre o papel da IA em seu trabalho e a construir confiança em sua integridade académica,⁸⁸ além de reduzir mal-entendidos quanto às fontes de informação, limitando potencialmente alegações de desonestidade académica ou plágio.⁸⁹

Também existem ferramentas de detecção de uso de IAG em texto. No mercado, elas ainda não são unanimidade quanto à sua efetividade. Softwares de detecção destacam seções duvidosas e, normalmente, fornecem um valor percentual de quanto do texto é gerado por computador ou qual a probabilidade de ser gerado por computador.⁹⁰ Assim, abordagens para detectar textos escritos por IAG envolvem o desenvolvimento de métodos que analisam e avaliam características estatísticas ou topológicas do resultado fornecido pela ferramenta. Além disso, há abordagens que se baseiam na detecção de padrões, frases e estilos específicas dos textos processados.⁹¹ De forma bastante interessante, um estudante da Universidade de Princeton desenvolveu um aplicativo chamado “GPTZero” para detectar conteúdos produzidos por IA Generativa tomando como base, essencialmente, o grau de aleatoriedade do

⁸⁷ UNIVERSIDADE DE OXFORD. **Publication and Authorship**. Disponível em: <<https://researchsupport.admin.ox.ac.uk/governance/integrity/publication>>. Acesso em 10 de out. 2024.

⁸⁸ UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford**: Enabling innovation through AI best practices. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

⁸⁹ CALTECH - CALIFORNIA INSTITUTE OF TECHNOLOGY. **Guidance on the Use of Generative AI and Large Language Model Tools**. 22 de outubro de 2023. Disponível em: <<https://www.caltech.edu/campus-life-events/campus-announcements/guidance-on-the-use-of-generative-ai-and-large-language-model-tools>>. Acesso em 6 de ago. 2024.

⁹⁰ GIMPEL, H.; HALL, K.; DECKER, S.; EYMANN, T.; LÄMMERMANN, L.; MÄDCHE, A.; RÖGLINGER, R.; RUINER, C.; SCHOCH, M.; SCHOOP, M.; URBACH, N.; VANDIRK, S. (2023). **Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers**. University of Hohenheim. Disponível em <https://digital.uni-hohenheim.de/fileadmin/einrichtungen/digital/Generative_AI_and_ChatGPT_in_Higher_Education.pdf>. Acesso em 10 de jul. 2024. p. 36

⁹¹ ALEMANHA. Federal Office for Information Security. **Generative AI Models: Opportunities and Risks for Industry and Authorities**. Disponível em: <https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4> Acesso em 20 de jul. de 2024. p. 22

texto, de forma a avaliar se a escrita é complexa e variada como a de um humano.⁹²

Alguns exemplos de ferramentas de detecção de plágio para uso de IA Generativa:⁹³

- Copyleaks (<https://copyleaks.com>): ferramenta habilitada para IA, projetada para usar algoritmos de correspondência de texto;
- Plagiarism Checker X (<https://plagiarismcheckerx.com>): ferramenta desenvolvida especialmente para que alunos possam checar seus trabalhos antes de entregá-los para avaliação;
- Scribbr (<https://www.scribbr.co.uk/plagiarism-checker>): ferramenta gratuita que usa algoritmos;
- Turnitin (<https://www.turnitin.com>): ferramenta de detecção de plágio amplamente utilizada para comparar trabalhos com extenso banco de dados de trabalhos acadêmicos, artigos e páginas da Internet. A empresa Turnitin também possui diversas outras pequenas empresas de detecção de plágio;⁹⁴
- Unicheck (<https://unicheck.com>): Projetado para usar algoritmos de IA para verificar os trabalhos por meio de um extenso banco de dados de fontes acadêmicas e online.

As ferramentas atualmente disponíveis podem ser visualizadas no quadro abaixo:

⁹² DWIVEDI, Yogesh K.; et. al. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy, *International Journal of Information Management*. Volume 71, 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0268401223000233>>. Acesso em 12 de nov. 2024. P. 27

⁹³ HOLMES, Wayne. *The Unintended Consequences of Artificial Intelligence and Education*. (2023) Education International: Brussels, Belgium. Disponível em: <<https://discovery.ucl.ac.uk/id/eprint/10179267/1/Holmes%20-%202023%20-%20The%20Unintended%20Consequences%20of%20Artificial%20Intellig.pdf>>. Acesso em 15 out. 2024. p. 40

⁹⁴ A ferramenta Turnitin é disponibilizada pela USP como parte das ações de valorização da integridade acadêmica (<https://www.abcd.usp.br/informa/82940/>).

Tabela 2 – Exemplos de ferramentas de Inteligência Artificial

Ferramentas de detecção de plágio	Ferramentas de detecção de IA Generativa	Ferramentas que usam IA Generativa para dificultar detecção de plágio	Ferramentas que usam IA Generativa para dificultar detecção de uso de IA Generativa
plagium.com copyleaks Scribbr Plagiarism Checker X Expertchat AI Unicheck Grammarly	plagium.com copyleaks Scribbr semihuman.ai Turnitin undetectable AI jusdone.ai humbot.ai Expertchat AI	Grammarly plagiarismremover.net Expertchat AI	semihuman.ai bypassGPT undetectable AI writehuman.ai humbot.ai

É preciso apontar os riscos de acurácia ligados a essas ferramentas.

Tais softwares de detecção de plágio ou de detecção de uso de ferramentas de IA Generativa podem trazer falsos positivos, com consequências negativas para o aluno e para a confiança entre pares e entre discentes e docentes no ambiente universitário.

Assim, não se exclui o risco de que conteúdo legítimo possa ser indicado como plágio ou como texto automatizado, levando a acusações injustas,⁹⁵ o que demanda medidas de governança abalizadas, com oportunidades de defesa e revisão, diálogo, antes da aplicação de medidas disciplinares.

Tendo isso em mente, a UNESCO recomenda que a estratégia institucional imediata quanto à defesa da integridade acadêmica está nos seres humanos, ao repensar tarefas e avaliações dadas aos alunos, incentivando a transparência

⁹⁵ HOLMES, Wayne. **The Unintended Consequences of Artificial Intelligence and Education**. (2023) Education International: Brussels, Belgium. Disponível em: <<https://discovery.ucl.ac.uk/id/eprint/10179267/1/Holmes%20-%202023%20-%20The%20Unintended%20Consequences%20of%20Artificial%20Intellig.pdf>>. Acesso em 15 out. 2024. p. 40

quando ao uso das ferramentas e o uso de tarefas em avaliações que exijam habilidades humanas não bem reproduzidas por IA ou ainda, que envolvam o uso explícito de sistemas de IA, complementado pela supervisão ou aportes humanos.⁹⁶

Embora o plágio (falta de reconhecimento autoral no texto) e a falta de transparência quanto a uso de IAG sejam problemas de ordem distinta quanto a integridade acadêmica, algumas universidades sugerem que medidas similares quanto a sua governança. Isso porque autores não devem declarar, como suas, aquelas ideias ou textos que foram gerados por ferramentas de IAG⁹⁷, o que seria semelhante a cópia de frases e parágrafos inteiros de outra fonte.⁹⁸

Outras universidades incluem o uso não transparente de ferramentas de IAG como “auxílio de outra pessoa”, apontando que, em algumas situações, o uso da ferramenta não deveria ser permitido, como se daria com o emprego da IAG para realizar, de forma substancial, uma tarefa ou avaliação.⁹⁹

Também a Universidade de Nova York (NYU) adota o entendimento de que o uso de IAGs se enquadraria na categoria de submeter um trabalho que foi realizado por outra pessoa, substancialmente ou no todo, como próprio.¹⁰⁰

⁹⁶ UNESCO. **Guidance for generative AI in education and research**. Paris: UNESCO, 2023. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000386693>>. Acesso em 13 out. 2024. p. 28

⁹⁷ UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <<https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024

⁹⁸ UNIVERSIDADE DE BERKELEY, CALIFORNIA. **Understanding AI Writing Tools and Their Uses for Teaching and Learning at UC Berkeley**. Disponível em: <<https://teaching.berkeley.edu/understanding-ai-writing-tools-and-their-uses-teaching-and-learning-uc-berkeley>>. Acesso em 6 de out. 2024.

⁹⁹ UNIVERSIDADE DE STANFORD. **Generative AI Policy Guidance**. Guidance adopted on February 16, 2023. Disponível em: <<https://communitystandards.stanford.edu/generative-ai-policy-guidance>>. Acesso em 6 de ago. 2024.

¹⁰⁰ UNIVERSIDADE DE NOVA YORK (NYU). **Teaching with Generative AI**. Disponível em: <<https://www.nyu.edu/faculty/teaching-and-learning-resources/teaching-with-generative-tools.html>>. Acesso em: 28 de set. 2024.

Ainda, outras faculdades colocam a prática em uma categoria mais genérica de “má conduta acadêmica” ou “integridade acadêmica”.¹⁰¹⁻¹⁰²⁻¹⁰³⁻¹⁰⁴

Em todos os casos, as universidades analisadas não criaram penalidades específicas para violações de integridade acadêmica quando há mau uso das ferramentas de IAG, englobando-as às condutas e penalidade já previstas em seus códigos de ética acadêmica.

2.5. Outras preocupações e medidas de governança no uso de IAG

Além da preocupação com a integridade acadêmica, as preocupações e limitações inerentes às ferramentas de IAG que já apontamos acima também se aplicam à educação, e devem ser enfrentadas pelas instituições.

Não por outra razão, as universidades que analisamos apontam medidas para mitigá-las.

(i) Necessidade de revisão do conteúdo produzido pela IAG

Conforme já destacamos, pela própria natureza das IAGs, seus resultados podem ter incorreções ou serem completamente fictícios. Assim, compete aos usuários a responsabilidade pela acurácia daquilo que escrevem.¹⁰⁵⁻¹⁰⁶ Em

¹⁰¹ UNIVERSIDADE DE CAMBRIDGE. **Plagiarism and Academic Misconduct**. Disponível em: <<https://www.plagiarism.admin.cam.ac.uk/what-academic-misconduct/artificial-intelligence>>. Acesso em 1 de out. 2024

¹⁰² UNIVERSIDADE DE PRINCETON. **Generative AI in Teaching and Learning Fall 2023, Faculty and Student Memo**. Disponível em: <<https://odoc.princeton.edu/about/official-deans-communications/2023/generative-ai-teaching-and-learning-fall-2023-faculty-and>>. Acesso em 6 de ago. 2024.

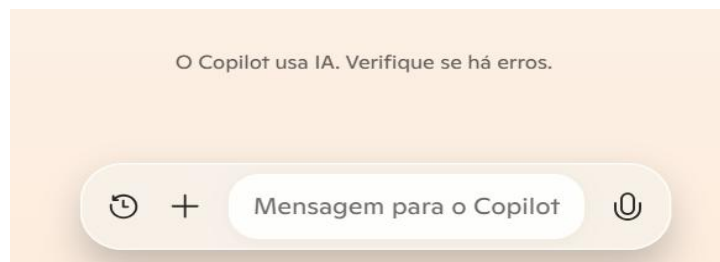
¹⁰³ CALTECH - CALIFORNIA INSTITUTE OF TECHNOLOGY. **Guidance on the Use of Generative AI and Large Language Model Tools**. 22 de outubro de 2023. Disponível em: <<https://www.caltech.edu/campus-life-events/campus-announcements/guidance-on-the-use-of-generative-ai-and-large-language-model-tools>>. Acesso em 6 de ago. 2024.

¹⁰⁴ UNIVERSIDADE DE YALE. **Guidelines for the Use of Generative AI Tools**. 20 de setembro de 2023. Disponível em: <<https://provost.yale.edu/news/guidelines-use-generative-ai-tools>>. Acesso em 6 de ago. 2024.

¹⁰⁵ UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <<https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024

¹⁰⁶ UNIVERSIDADE DE HARVARD. **Initial guidelines for the use of Generative AI tools at Harvard**. Disponível em: <<https://huit.harvard.edu/ai/guidelines#block-boxes-1694711840>>. Acesso em 6 de ago. 2024.

resumo, (ainda) não é possível confiar completamente no resultado produzido pela IAG.¹⁰⁷



Mensagem inicial do Copilot, plataforma de IAG da Microsoft

Considerando as limitações das ferramentas de IAG, as universidades destacam que se deve manter certo ceticismo quanto às IAGs, uma vez que estas podem e frequentemente fazem as coisas erradas.

Como abordado no ANEXO II, o conteúdo gerado por IA pode ser impreciso, enganoso ou totalmente fabricado (o que se costuma chamar de “alucinações”). Um chatbot pode fornecer uma resposta plausível e que parece confiável. No entanto, a resposta pode ser completamente falsa ou significativamente distorcida.¹⁰⁸

E apesar das melhorias na tecnologia, os resultados da IA generativa continuam a ser imprevisíveis. Devido à sua natureza probabilística, a IAG produz sempre resultados diferentes para os mesmos *prompts*, o que a torna diferente de softwares tradicionais, contrariando as expectativas que os usuários depositam nela.¹⁰⁹

Além disso, há o risco de que as ferramentas de IAG produzam conteúdos preconceituosos ou com vieses (políticos, por exemplo). IAGs não produzem

¹⁰⁷ UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford:** Enabling innovation through AI best practices. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

¹⁰⁸ UNIVERSIDADE DE HARVARD. **Ideas for experimenting with Generative AI: Use cases and things to keep in mind.** Disponível em: <<https://huit.harvard.edu/news/ai-use-cases>>. Acesso em: 6 de ago. 2024.

¹⁰⁹ UNIVERSIDADE DE OXFORD. **Beyond ChatGPT:** State of AI in Academic Practice. Outubro de 2023. Disponível em: <https://www.ctl.ox.ac.uk/sites/default/files/ctl/documents/media/beyond_chatgpt_-_state_of_ai_for_autumn_2023_correct.pdf>. Acesso em 6 de jul. de 2024

respostas neutras, uma vez que as fontes de informação usadas para treinar a ferramenta são criadas por pessoas e, como consequência, contêm preconceitos e estereótipos.¹¹⁰ Assim, também quanto a este ponto é necessário proceder com cautela ao utilizar os resultados das ferramentas de IAG sem uma análise aprofundada e reflexão crítica.¹¹¹⁻¹¹²

(ii) Confidencialidade / proteção de dados

Para endereçar a preocupação com confidencialidade e proteção de dados, as universidades alertam os alunos sobre o próprio funcionamento das ferramentas de IAG.

Assim, uma vez que quaisquer dados colocados em sistemas de IA de terceiros são transmitidos e armazenados em servidores externos sobre os quais as instituições não têm controle direto, as universidades sugerem medidas como: (i) ser cauteloso com aquilo que será inserido na IAG, evitando incluir informações confidenciais ou sensíveis que são inseridas nas IAGs, (ii) desabilitar opções relacionadas ao histórico salvo para evitar que as informações sejam registradas, ou rastreadas, para fins de treinamento da ferramenta.¹¹³⁻¹¹⁴

Além disso, para aquelas ferramentas que possibilitam ajustar as opções de compartilhamento de dados, é sugerido que os alunos adotem a configuração que impeça a utilização de informações e de prompts escritos pelo usuário para

¹¹⁰ UNIVERSIDADE DE CAMBRIDGE. **How we use AI tools**. Disponível em: <<https://www.communications.cam.ac.uk/generative-ai-tool-guidelines>>. Acesso em 6 de ago. 2024.

¹¹¹ UNIVERSIDADE DE CAMBRIDGE. **Using Generative AI: General Principles for use of Generative AI**. Disponível em: <<https://blendedlearning.cam.ac.uk/guidance-support/ai-and-education/using-generative-ai>>. Acesso em 1 de out. 2024

¹¹² CALTECH - CALIFORNIA INSTITUTE OF TECHNOLOGY. **Guidance on the Use of Generative AI and Large Language Model Tools**. 22 de outubro de 2023. Disponível em: <<https://www.caltech.edu/campus-life-events/campus-announcements/guidance-on-the-use-of-generative-ai-and-large-language-model-tools>>. Acesso em 6 de ago. 2024.

¹¹³ UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford: Enabling innovation through AI best practices**. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

¹¹⁴ UNIVERSIDADE DE PRINCETON. **Generative AI**. Disponível em: <<https://mcgraw.princeton.edu/generative-ai>>. Acesso em 6 de ago. 2024.

o treinamento dos modelos.¹¹⁵⁻¹¹⁶ Outra medida sugerida é usar apenas aquelas ferramentas aprovadas pela instituição.¹¹⁷⁻¹¹⁸

Medida de governança por vezes indicada é a contratação da ferramenta pela universidade ou unidade de ensino para sua disponibilização aos discentes e docentes, em interfaces próprias, o que tende a trazer maior transparência entre os pares quanto ao seu uso, além de permitir o treinamento da IAG para conteúdos típicos gerados pela atividade da IES e a sistematização e benchmarking de prompts bem sucedidos para geração de conteúdos apropriados para as atividades da IES. A medida pode, também, permitir a introdução de restrições técnicas à ferramenta para detectar e evitar a introdução de dados pessoais nos prompts (por meio de avisos aos usuários, por exemplo). Além disso filtros adicionais quanto ao conteúdo gerado podem ser utilizados para alertar quanto a conteúdos discriminatórios ou abusivos.

Digno de nota, entretanto, que a aquisição de licenças para essas ferramentas apresenta custo significativo. Esse custo deve diminuir à medida em que os vários sistemas que vêm sendo lançados levem à consolidação das melhores ferramentas, tornando a aquisição mais acessível.

¹¹⁵ MIT - MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Generative AI for Teaching & Learning:** Explore our resources to unlock AI's potential for enhancing teaching and learning at MIT Sloan. Disponível em: <<https://mitsloanedtech.mit.edu/ai/>>. Acesso em 6 de out. 2024.

¹¹⁶ UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford:** Enabling innovation through AI best practices. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

¹¹⁷ MIT - MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Generative AI for Teaching & Learning:** Explore our resources to unlock AI's potential for enhancing teaching and learning at MIT Sloan. Disponível em: <<https://mitsloanedtech.mit.edu/ai/>>. Acesso em 6 de out. 2024.

¹¹⁸ UNIVERSIDADE DE HARVARD. **Initial guidelines for the use of Generative AI tools at Harvard.** Disponível em: <<https://huit.harvard.edu/ai/guidelines#block-boxes-1694711840>>. Acesso em 6 de ago. 2024.

CAPÍTULO 2: O USO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA PELOS(AS) ALUNOS DA FACULDADE DE DIREITO DA USP (FDUSP)

1. Metodologia

Considerando os impactos que IAGs podem acarretar na FDUSP, a estratégia de pesquisa adotada foi quantitativa, por adesão, com aplicação de quesitos, via formulário enviado aos alunos de graduação e pós graduação por meio de e-mail institucional, entre 21 de novembro e 12 de dezembro de 2023, com 94 respondentes, sendo 32 da pós-graduação e 62 da graduação.¹¹⁹

Tal metodologia se mostra adequada como aproximação sobre como as IAGs vem sendo usadas e qual a percepção dos próprios alunos sobre a ferramenta e o interesse no seu uso, com o objetivo de compreender a realidade e auxiliar os trabalhos de revisão da regulação da IA na FDUSP. O fato de os respondentes terem sido em pequeno número, no universo do total de alunos da instituição, aliado à ausência de dados sobre o seu perfil demográfico, dificulta generalizações em relação ao total de alunos da faculdade e pode trazer um viés, em que alunos mais afeitos à tecnologia tenham se interessado em participar do levantamento. Apesar disso, os indicadores sobre os tipos de uso, dificuldades encontradas e grau de transparência, trazem informações relevantes para orientar a política da FDUSP em relação ao uso de IAG.

A captura da percepção subjetiva dos alunos quanto à tecnologia é fator relevante para a implementação bem-sucedida das ferramentas digitais em e-Governance, assim como o treinamento e familiaridade quanto ao uso de IAGs.¹²⁰

Assim, o presente item apresenta a forma pela qual foi realizada a pesquisa junto aos alunos de graduação e pós-graduação da FDUSP, sendo que o item subsequente desenvolverá a análise crítica dos resultados.

¹¹⁹ GATTI, Bernadette. A. Estudos quantitativos em educação. Educação e Pesquisa, São Paulo, v. 30, n.1, p. 11-30, jan./abr. 2004. Disponível em: <https://www.scielo.br/j/ep/a/XBpXkMkBSsbBCrCLWjzyWyB/?lang=pt&format=pdf>. Acesso em 18 ago.2024.

¹²⁰ AHN, Michael J., CHEN, Yu-Che. Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. Government Information Quarterly, Volume 39, Issue 2, 2022

Conforme destacado no Capítulo 1, as principais preocupações éticas e de governança ligadas ao uso de IAGs para fins acadêmicos dizem respeito à qualidade dos resultados gerados e à transparência sobre o emprego da ferramenta dentro da faculdade.

Como visto, as ferramentas de IAGs de textos são treinadas para simular conteúdo gerado por humanos e não para retratar fidedignamente fatos, ou seja, não são ferramentas de busca, e sua aplicação em domínios técnicos do conhecimento pode apresentar “alucinações” (referências a precedentes, bibliografia ou fatos inexistentes), bem como imprecisões e equívocos conceituais.

Além disso, tais ferramentas não possuem capacidade cognitiva humana de compreensão do conteúdo ou raciocínio (inferências lógicas sobre o conteúdo), gerando os resultados apenas a partir de correlações estatísticas entre palavras e trechos dos textos fornecidos e gerados de modo recorrente. Resultados inadequados podem também resultar da falta de familiaridade dos usuários e de conhecimento da ferramenta. O entendimento do funcionamento das IAGs e a habilidade no seu uso são relevantes da perspectiva de revisão humana dos textos gerados por máquina. Outro ponto importante sobre a qualidade dos resultados diz respeito ao tipo ou finalidade de uso, tendo em vista que há usos inapropriados das IAGs, em particular o uso para pesquisas ou buscas sobre fatos.

Assim, a preocupação central do levantamento da pesquisa realizada voltou-se a perquirir (i) em que medida tais ferramentas são utilizadas efetivamente pelos alunos para atividades acadêmicas; (ii) para quais tarefas ou finalidades são utilizadas (permitindo avaliar se os usos praticados são apropriados ou não); (iii) a frequência de uso e o grau de familiaridade dos(as) usuários(as) com as IAGs; (iv) as principais vantagens, desvantagens e limitações percebidas com o uso; (v) a transparência sobre o uso da ferramenta; e (vi) a opinião dos(as) alunos(as) sobre a oportunidade do uso (se o mesmo deveria ser estimulado ou proibido e para quais finalidades) bem como sobre a oportunidade do treinamento e instrução sobre a ferramenta.

Para o levantamento das práticas dos usuários, alunos de graduação e pós-graduação, foi encaminhado o questionário do Apêndice A, via Google Forms, cujos *links* são aqui disponibilizados¹²¹ para consulta.

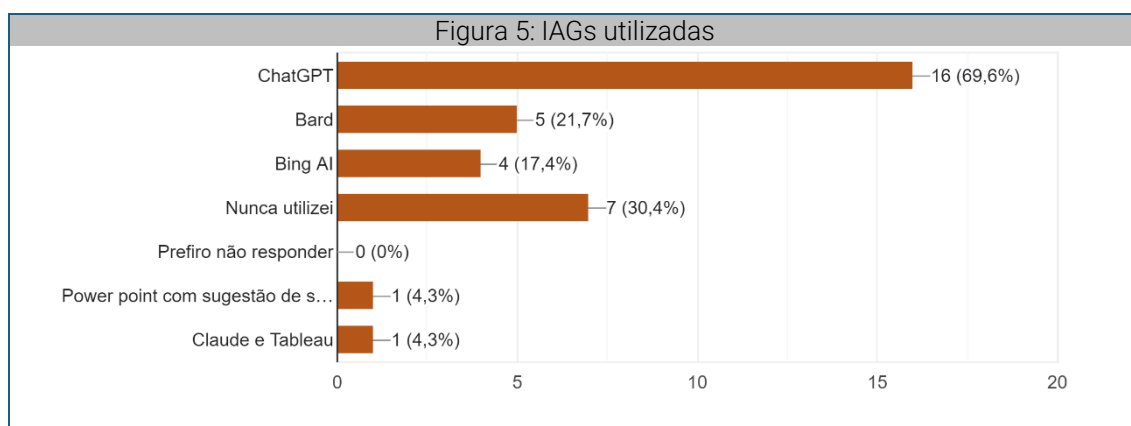
O formulário elaborado para a graduação possuía 21 questões e o formulário da pós-graduação 18 perguntas, a sua maioria em múltipla escolha, de forma a traçar o perfil de uso do respondente com relação a ferramentas de IAG para obter a informação sobre sua utilização, finalidades de uso e oportunidades avaliadas pelo respondente. As perguntas, em sua maioria, eram de múltipla escolha com alternativas predeterminadas.

2. Análise dos resultados

Neste item, passamos à análise das informações do levantamento sobre o uso de IAG na FDUSP.

Número expressivo de alunos, 65,2% dentre aqueles que responderam à pesquisa, informaram já ter utilizado algum tipo de ferramenta de IAG para atividades da faculdade. Dentre as ferramentas utilizadas, vê-se que a maioria, dentre os respondentes, já teve experiência de uso do ChatGPT (69,6%), atual líder de mercado, mas também percentual significativo, experimentou outras ferramentas que vem ganhando popularidade no Brasil, como Bard (hoje, Gemini) (Google AI) e Bing AI, valendo notar que são ferramentas associadas a outros softwares de largo emprego em atividades pessoais e profissionais, como buscadores e pacotes de ferramentas de escritório.

¹²¹ Link para a pesquisa da graduação: <<https://forms.gle/fWmKyPeTuScjL9Z18>>. Link para a pesquisa da pós-graduação: <<https://forms.gle/XgfgUy4qHcmNthv4A>>.



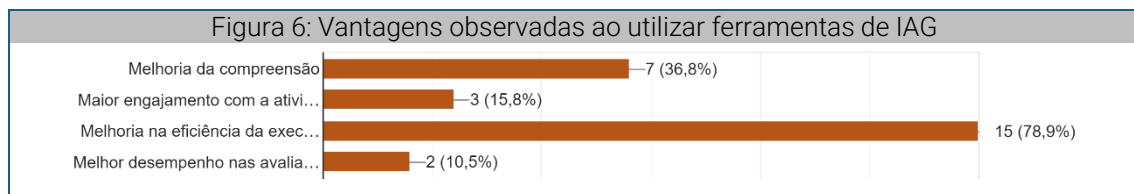
Não só a utilização, mas também a frequência aumenta dentre aqueles que frequentam ambientes acadêmicos, como docentes ou discentes em cursos de graduação ou pós-graduação.

Dado que, como visto no Capítulo 1, há relevante preocupação com a qualidade dos resultados das IAGs (sobretudo em domínios técnicos de aplicação, tais como produção de artigos e trabalhos acadêmicos) e com a inadequação das ferramentas para pesquisas fáticas ou buscas de informação, uma das perguntas chave da pesquisa buscou levantar os tipos de uso pelos respondentes.

Embora a maioria tenha indicado usos adequados, como perguntas diversas ao *chatbot*, como consultas em língua estrangeira ou português, revisões de texto, auxílio na redação de trabalhos, aperfeiçoamento textual, há aqueles que usaram a ferramenta para pesquisa de fontes (16,7%) e jurisprudência (5,6%). O uso da ferramenta como mecanismo de busca não é, em si, um problema, desde que haja a verificação dos resultados pelo usuário, antes do uso do conteúdo gerado em sua atividade discente ou docente. Vale considerar que cerca da metade dos respondentes indicou usar a ferramenta apenas esporadicamente (47,8%). Assim, é possível afirmar que a maioria dos alunos tinha pouca experiência de uso dessas ferramentas.

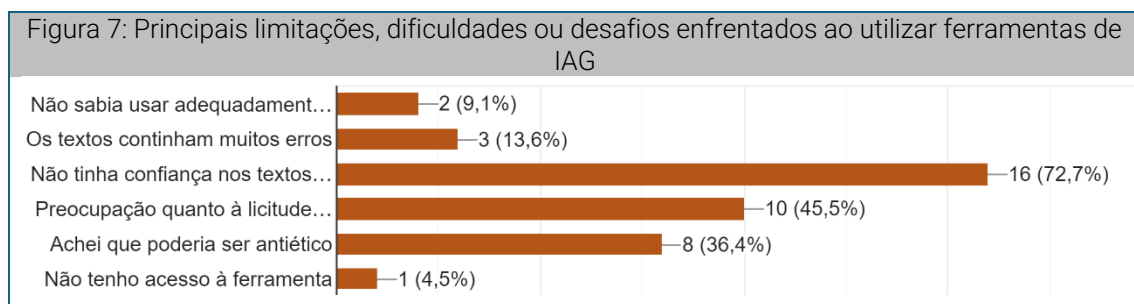
Quanto à pergunta sobre a percepção de vantagens do uso da ferramenta de IAG para fins acadêmicos, percentual significativo dos respondentes indicou

que sim (87%). Dentre as vantagens, 78,9% entendem que há melhoria na execução de tarefas e 36,8% apontaram melhoria de compreensão.



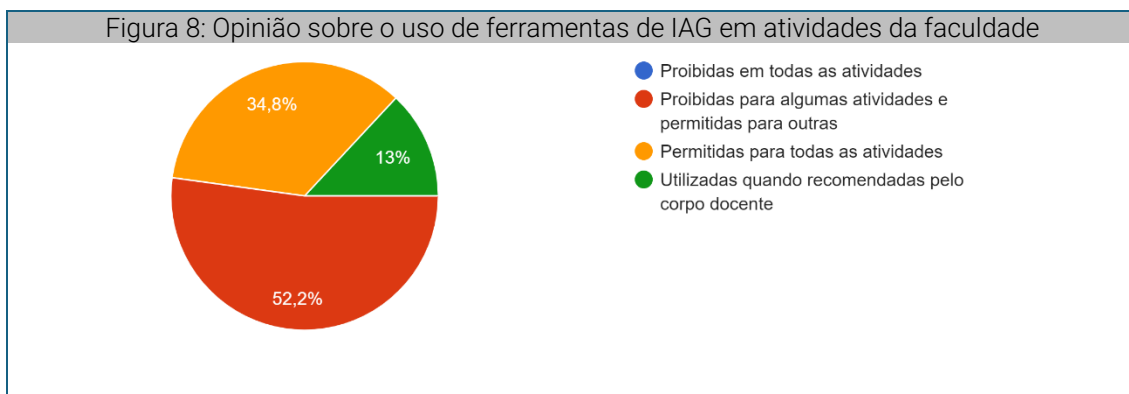
Dentre as limitações ou desvantagens apontadas pelos(as) respondentes, 72,7% indicaram que não tinham confiança nos textos produzidos, o que também é a percepção de 13,6% dos respondentes, que indicaram que os textos continham erros.

Número significativo também aponta não ter certeza da licitude de usar a IAG para trabalhos acadêmicos (45,5%). 9,12% apontam que não sabiam usar adequadamente a ferramenta.

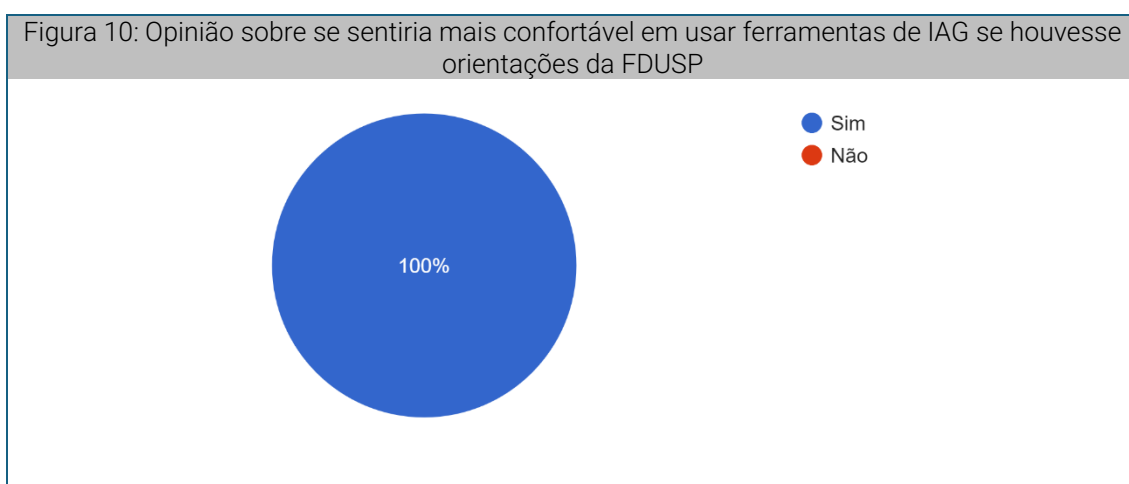
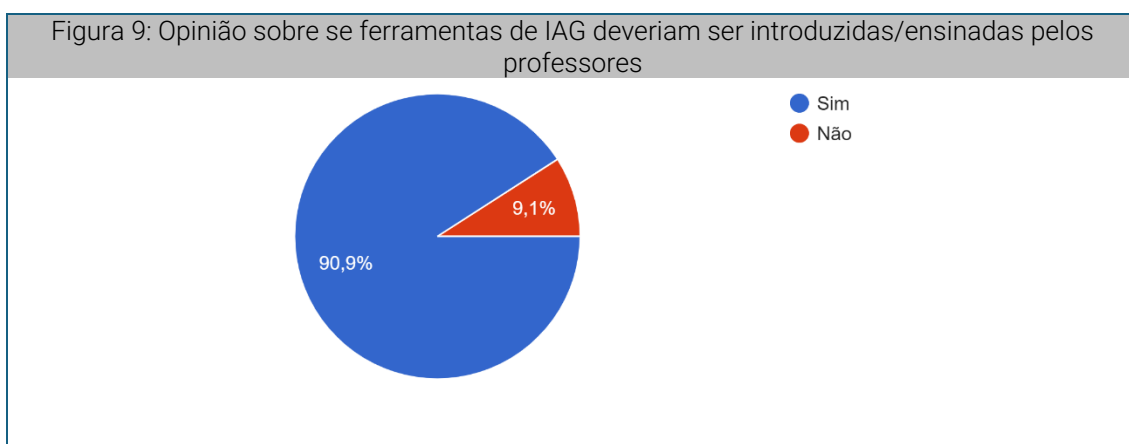


Tais respostas são um elemento importante a corroborar a necessidade de treinamento dos alunos para aprimorar suas habilidades e potencializar os benefícios do uso dessas ferramentas.

Apesar das limitações e dificuldades enfrentadas, os alunos possuem uma visão positiva sobre as IAGs. Nenhum dos alunos respondeu que a ferramenta deveria ser proibida em todas as atividades junto à faculdade, ainda que parcela considerável entenda que deve haver proibição para algumas tarefas.



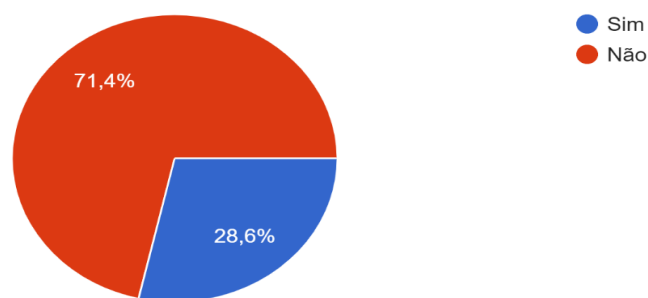
Tal percepção encontra-se refletida nas respostas às perguntas sobre interesse em treinamento e capacitação para uso das ferramentas de IAGs, e se os alunos se sentiriam mais confortáveis em usar a IAG se houvesse orientações da faculdade.



No tocante à transparência, a pesquisa revelou dados importantes. Em primeiro lugar, destaca-se que cerca de 71,4% dos alunos revelaram não informar

ter utilizado a ferramenta em suas atividades acadêmicas. Como destacado no Capítulo 1, tendo em vista que o resultado do uso de IAGs de texto é indiscernível do conteúdo humano, a falta de informação pode passar despercebida e dificultar, assim, o necessário trabalho de correção, principalmente em se tratando de domínio técnico de aplicação, como é o campo jurídico. Essa questão também aponta preocupações quanto à integridade acadêmica, conforme já apontado no Capítulo 1.

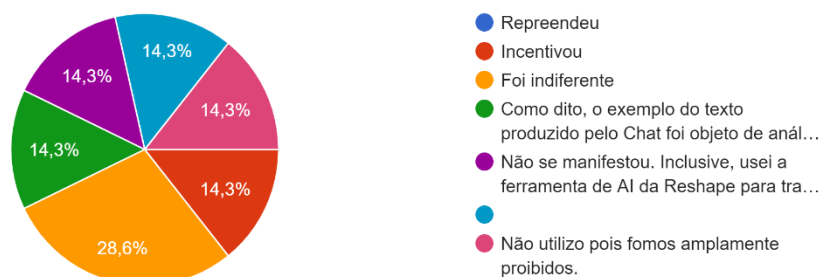
Figura 11: Costuma indicar no trabalho/atividade que a IAG foi utilizada



É importante a orientação sobre o uso e necessidade de transparência, de um lado, para que não se corra o risco de que resultados imprecisos, equivocados ou mesmo com “alucinações” estejam presentes nos trabalhos entregues pelos alunos, bem como conscientizá-los de que, a utilização da ferramenta sem devida indicação representa uma violação de políticas de integridade acadêmica.

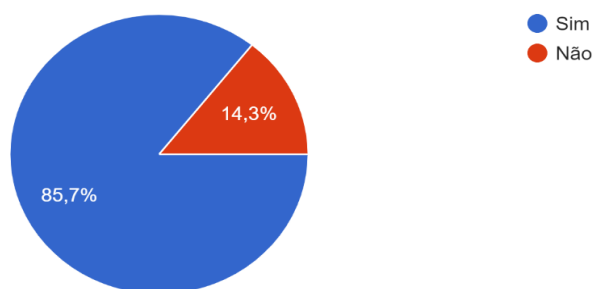
Entre aqueles alunos que responderam que teriam indicado o uso da ferramenta em seus trabalhos, nenhum aluno apontou ter sido repreendido.

Figura 12: Reação do professor quando o aluno indicou no trabalho/atividade que a IAG foi utilizada



Apesar da falta de transparência quanto ao uso, percentual considerável de alunos apontou que costuma revisar o conteúdo gerado pela IA.

Figura 13: Costuma revisar o conteúdo gerado pela ferramenta de IAG



Algumas perguntas específicas foram feitas com os alunos de pós-graduação, de forma a aferir se as ferramentas de IAG eram usadas especificamente para fins de produção das dissertações e teses.

Figura 14: Para quais fins de produção da dissertação ou tese o aluno usa de ferramentas de IAG

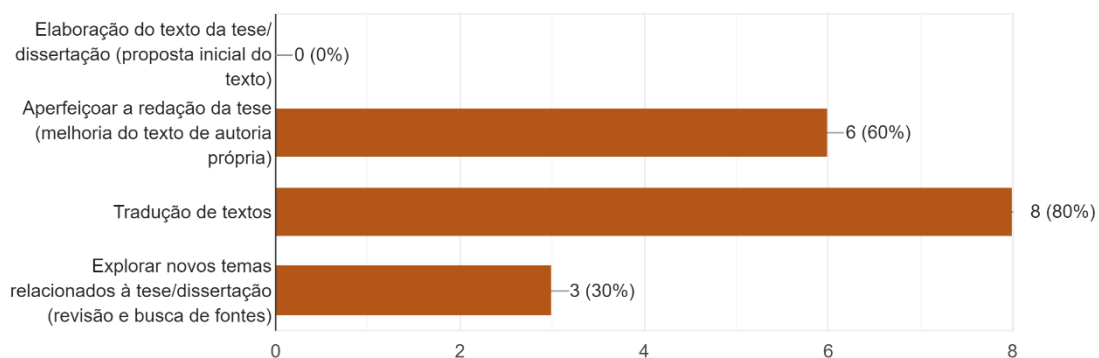
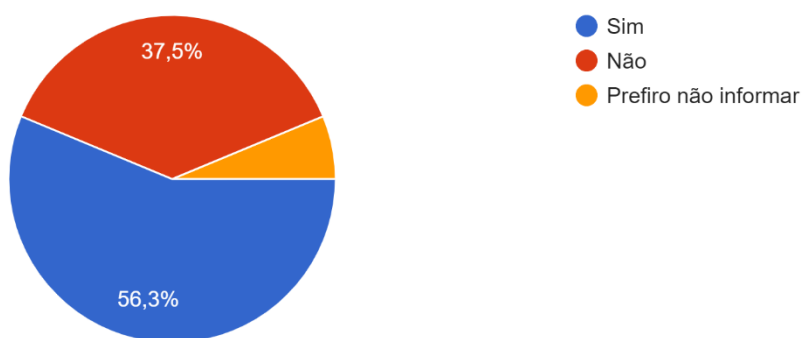


Figura 15: Comunicou o orientador quanto ao uso de ferramentas de IAG na dissertação ou tese o aluno



Os usos realizados pelos alunos de pós-graduação para atividades relacionadas à dissertação ou tese são adequados. Contudo, persiste o problema da falta de transparência junto ao orientador.

3. Conclusões e indicações de melhores práticas

O presente Relatório teve dois objetivos centrais: (i) abordar aspectos de governança sobre o uso de sistemas de inteligência artificial generativa (IAG) por instituições de ensino superior; e (ii) relatar resultados de pesquisa sobre o uso de ferramentas de IAG de texto pelos alunos da FDUSP.

O levantamento sobre políticas quanto ao uso de IAG nas principais universidades internacionais revelou o reconhecimento da inexorabilidade do uso dessas ferramentas nas atividades docentes e no desempenho futuro da profissão, razão pela qual tais instituições têm estimulado o letramento de docentes, discentes e funcionários, para o aprendizado e desenvolvimento de habilidades com o uso desses sistemas, bem como a adaptação das metodologias de ensino e de avaliação. O objetivo é estimular o uso responsável, o que significa potencializar as capacidades dos alunos e professores, tanto na docência quanto na pesquisa, sempre preservando o rigor acadêmico e a integridade.

Para tanto, tais universidades desenvolveram medidas para estimular a transparência quanto ao uso de IAGs, a revisão do conteúdo geral, orientações quanto aos riscos à privacidade e violação autoral, a incorporação dessas ferramentas nos métodos didáticos e de avaliação, a discriminação entre usos autorizados e não autorizados e medidas disciplinares em relação a usos não autorizados. Tais medidas, estímulos e restrições, considerando a autonomia universitária, podem ser consensadas na instituição ou adotadas por cada docente. Algumas criaram centros específicos dentro da instituição para treinamento de alunos e de docentes para letramento e incorporação de habilidades no uso dessas ferramentas, além da reformulação de grades e adaptação do ensino para preparação do aluno a um mercado em processo de transformação pelo uso dessa tecnologia.

O levantamento dos usos da IAG na FDUSP mostrou que, dentre as dificuldades de uso dos alunos, estão a falta de familiaridade com a tecnologia, a necessidade de revisão de resultados e a dúvida quanto à licitude e à ética do seu

emprego. Tais fatores parecem indicar uma fase de experimentação, com interesse crescente pelos discentes, o que é indicado pela elevada manifestação de desejo de receberem orientações e ver tais ferramentas incorporadas ao ensino na FDUSP. A dúvida quanto à ética e licitude pode explicar a razão pela qual há elevada opacidade quanto ao uso da ferramenta pelos discentes. Percentual elevado dos respondentes que usam IAG indicaram não revelar tal uso ao professor e percentual significativo de orientandos na pós-graduação indicaram não revelar o uso ao próprio orientador.

Tais elementos parecem indicar a conveniência de movimento da FDUSP, no sentido de acompanhar essas novas tendências, com a possível criação de centro para desenvolvimento de novos métodos de ensino e pesquisa, envolvendo o uso de sistemas de IA, bem como para a capacitação e o treinamento de discentes e docentes a fim de incorporar de novas habilidades, seja para a atividade de pesquisa ou para o mercado de trabalho.

Seria conveniente, por outro lado, a adoção de parâmetros gerais ou diretrizes, no sentido de estimular o uso responsável desses sistemas por docentes, discentes e funcionários, assegurando a transparência, a proteção de dados pessoais, a integridade acadêmica, o rigor metodológico e a constante evolução para o acompanhamento das melhores práticas de emprego da tecnologia na educação.

Por fim, valendo-se deste relatório como veículo de conscientização, seria recomendável a realização periódica do levantamento do uso por discentes, que pode também ser estendida a docentes, de modo a acompanhar em que medida o uso efetivo da ferramenta dentro da FDUSP vem se adequando às melhores práticas de uso responsável da tecnologia.

Referências Bibliográficas

ABBOT, Ryan. **Rational Robots**: a response to interesting, renegotiated, and relational robots. *Jerusalem Review of Legal Studies*, n. 1 (2022), p. 40-53. <https://doi.org/10.1093/jrls/jlac004>

AHN, Michael J., CHEN, Yu-Che. Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly*, Volume 39, Issue 2, 2022

ALEMANHA. Federal Office for Information Security. **Generative AI Models**: Opportunities and Risks for Industry and Authorities. Disponível em: <https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4> Acesso em 20 de jul. de 2024.

ALMADA, Marco; MARANHÃO, Juliano; SARTOR, Giovanni. **Artificial intelligence**. In: PARCU, Pier Luigi; ROSSI, Maria Alessandra; BOTTA, Marco (Eds.). *Research handbook in competition & technology*. Edward Elgar Publishing, 2024. Draft, 22 September 2023.

ALQAHTANI, Tariq et al. **The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research**. *Research in Social and Administrative Pharmacy*, v. 19, n. 8, p. 1236-1242, 2023.

ATLAS, S. (2023). **ChatGPT for higher education and professional development**: A guide to conversational AI. Retrieved March 12, 2023. Disponível em: <https://digitalcommons.uri.edu/cba_facpubs/548/>. Acesso em 15 de jul. 2024. p. 24

AUSTIN, Doug. **Nearly 60 Percent of GPT 3.5 Outputs Contained Plagiarized Content**: Artificial Intelligence Trends. Disponível em: <<https://ediscoverytoday.com/2024/02/26/nearly-60-percent-of-gpt-3-5-outputs-contained-plagiarized-content-artificial-intelligence-trends/>>. Acesso em 18 nov. 2024.

BAIDOO-ANU, David; OWUSU ANSAH, Leticia. **Education in the Era of Generative Artificial Intelligence (AI)**: Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning (January 25, 2023). Disponível em: <<https://ssrn.com/abstract=4337484>>. Acesso em 10 de jul. 2024. p. 56

BANCO INTERAMERICANO DE DESENVOLVIMENTO. **Contratações de inovação**: guia de alternativas jurídicas e de boas práticas para contratações de inovação. 1. ed. Washington, D.C.: BID, 2023. Disponível em: <<https://publications.iadb.org/pt/contratacoes-de-inovacao-guia-de-alternativas-juridicas-e-de-boas-praticas-para-contratacoes-de>>. Acesso em: 15 ago. 2024.

BANDI, A.; ADAPA, P.V.S.R.; KUCHI, Y.E.V.P.K. **The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges**. *Future Internet* 2023, 15, 260. <https://doi.org/10.3390/fi15080260>

BBC News. **Sony World Photography Award 2023**: Winner refuses award after revealing AI creation. Disponível em: <https://www.bbc.com/news/entertainment-arts-65296763> Acesso em 18 de ago/2024.

BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. 2021. **On the Dangers of Stochastic Parrots**: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. Disponível em: <<https://doi.org/10.1145/3442188.3445922>>. Acesso em 9 de jul. 2024. p. 613

BEVIR, M. (2007). Governance. In BEVIR, M. **Encyclopedia of governance** (pp. 364-381). 2 volume set. Thousand Oaks, California: SAGE Publications, Inc.

BEVIR, M. **A Theory of Governance**. Los Angeles: University of California Press, 2013.

BEVIR, Mark. **Governance: A Very Short Introduction**. Oxford: Oxford University Press, 2012.

BIRKSTEDT, Teemu et al. AI governance: themes, knowledge gaps and future agendas. **Internet Research**, v. 33, n. 7, p. 133-167, 2023.

BONFANTI, Matteo E.; CAVELTY, Myriam Dunn; WENGER, Andreas, **Artificial intelligence and cyber-security**, in: The Routledge Social Science Handbook of AI, [s.l.]: Routledge, 2021.

BOWMAN, Samuel R. **Eight Things to Know about Large Language Models**. *Preprint, submetido em 2 abr. 2023. 10.48550/arXiv.2304.00612*.

BRIGGS, J. KODNANI, D. **The potentially large effects of artificial intelligence on economic growth**, Goldman Sachs Global economic Analyst, 2023.

CALTECH - CALIFORNIA INSTITUTE OF TECHNOLOGY. **Guidance on the Use of Generative AI and Large Language Model Tools**. 22 de outubro de 2023. Disponível em: <<https://www.caltech.edu/campus-life-events/campus-announcements/guidance-on-the-use-of-generative-ai-and-large-language-model-tools>>. Acesso em 6 de ago. 2024.

CEIBAL. Building Artificial Intelligence for Education. 2024. Disponível em: <<https://ceibal.edu.uy/wp-content/uploads/2024/05/Building-Artificial-Intelligence-for-education.pdf>>. Acesso em 20 de set. 2024. p. 10.

CHEN, Lin et al. **Artificial intelligence-based solutions for climate change: a review**. Environmental Chemistry Letters, v. 21, n. 5, p. 2525-2557, 2023.

CITRON, Danielle Keats; PASQUALE, Frank. **The scored society: Due process for automated predictions**. Wash. L. Rev., v. 89, p. 1, 2014.

CLIFFORD, R. D. **Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?** Tulane Law Review.Vol. 71. 1997. p. 1675-1703. https://scholarship.law.umassd.edu/fac_pubs/77/

CMA. **AI Foundation Models: Initial Report**. Competition & Markets Authority, set. 2023. Disponível em: https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf Acesso em 18 de ago. 2024.

COMPETITION & MARKETS AUTHORITY. **AI Foundation Models**: Technical Update Report. London: Department for Science, Innovation and Technology, 2023. Disponível em: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf. Acesso em: 19 ago. 2024.

CRAIG, Carys J.; KERR, Ian R., **The Death of the AI Author**. Osgoode Legal Studies Research Paper (March 25, 2019), (2021) 52(1) Ottawa Law Review 31 (<https://rdo-olr.org/2021/the-death-of-the-ai-author/>), Available at SSRN: <https://ssrn.com/abstract=1784951> or <http://dx.doi.org/10.2139/ssrn.3374951>

CRAWFORD, K. **Atlas of AI: power, politics, and the planetary costs of artificial intelligence**. 2021, Yale University Press.

DALY, Angela et al. AI, governance and ethics: global perspectives. **University of Hong Kong Faculty of Law Research Paper**, n. 2020/051, 2020.

DAVIES, C. R. **An Evolutionary Step in Intellectual Property Rights** – Artificial Intelligence and Intellectual Property. *Computer Law & Security Review*, v. 27, p. 601-630, 2011

DE LIMA LÚCIO, Magda et al. **Sentidos e significados de se planejar estrategicamente nas organizações públicas-planejamento estratégico sociotécnico (PLANES): análise de uma experiência.** *Nau Social*, v. 5, n. 9, 2014.

DELL'ACQUA, F., MCFOWLAND III, E., MOLLIICK, E., LIFSHITZ-ASSAF, H., KELLOGG, K.C., RAJENDRAN, S., KRAYER, L., CANDELON, F. and LAKHANI, K.R. (2023). **Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality.** Working Paper. Disponível em: <https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf> Acesso em 12 de jul. 2024.

DWIVEDI, Yogesh K.; et. al. **Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy,** *International Journal of Information Management*. Volume 71, 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0268401223000233>>. Acesso em 12 de nov. 2024.

EC HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. **Orientações éticas para uma IA de confiança.** 2018. Disponível em: <https://doi.org/10.2759/2686>. Acesso em: 15 jul. 2024.

ESTADOS UNIDOS. Copyright Office. **Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence.** 16 mar. 2023. Disponível em: <https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence>. Acesso em: 16 ago. 2024.

ESTADOS UNIDOS. Copyright Office. **Zarya of the Dawn.** Washington, D.C.: U.S. Copyright Office, 2023.

FEDERAL TRADE COMMISSION; DEPARTMENT OF JUSTICE; CONSUMER FINANCIAL PROTECTION BUREAU. **Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems.** Washington, D.C.: FTC, 2023. Disponível em: https://www.ftc.gov/system/files/ftc_gov/pdf/ai-joint-statement.pdf. Acesso em: 19 ago. 2024.

FLORIDI, Luciano. **AI as agency without intelligence: on ChatGPT, Large Language Models, and other generative models.** *Philosophy & Technology*, v. 36, n. 1, 2023

FREITAG, C.; BERNERS-LEE, M.; WIDDICKS, K.; KNOWLES, B.; BLAIR, G. S.; FRIDAY, A. The real climate and transformative impact of ICT: a critique of estimates, trends, and regulations. *Patterns*, v. 3, n. 8, 2022.

GATTI, Bernadette. A. Estudos quantitativos em educação. *Educação e Pesquisa*, São Paulo, v. 30, n.1, p. 11-30, jan./abr. 2004. Disponível em: <https://www.scielo.br/j/ep/a/XBpXkMkBSsbBCrCLWjzyWyB/?lang=pt&format=pdf>. Acesso em 18 ago.2024.

GHIM, Jong-Lyul; AHN, Sangzin. **Transforming clinical trials: the emerging roles of large language models.** *Translational and Clinical Pharmacology*, v. 31, n. 3, p. 131, 2023.

GIMPEL, H.; HALL, K.; DECKER, S.; EYMANN, T.; LÄMMERMANN, L.; MÄDCHE, A.; RÖGLINGER, R.; RUINER, C.; SCHOCH, M.; SCHOOP, M.; URBACH, N.; VANDIRK, S. (2023). **Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers.** University of Hohenheim. Disponível em <https://digital.uni-hohenheim.de/fileadmin/einrichtungen/digital/Generative_AI_and_ChatGPT_in_Higher_Education.pdf>. Acesso em 10 de jul. 2024.

GINSBURG, J. C.; BUDIARDJO, L. A. **Authors and Machines**. Berkeley Technology Law Journal, v. 34, n. 2, p. 343-456, 2018. Disponível em: <https://scholarship.law.columbia.edu/faculty_scholarship/2323/>.

GOLDMAN SACHS. 2023 **Global Economics Analyst**: The Potentially Large Effects of Artificial Intelligence on Economic Growth. GS Publishing. Disponível em: <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>. Acesso em: 15 ago. 2024.

GOLDMAN SACHS. **Generative AI could raise global GDP by 7%**. 5 de abril de 2023. Disponível em: <<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>>. Acesso em 9 de jul. 2024.

HAGENDORFF, Thilo. **Machine psychology**: Investigating emergent capabilities and behavior in large language models using psychological methods. **arXiv preprint arXiv:2303.13988**, 2023.

HILLIER, Mathew. **Why does ChatGPT generate fake references?** - TECHE. Macquarie University's Learning and Teaching blog, 10 mar. 2023. Disponível em: <<https://teche.mq.edu.au/2023/02/why-does-chatgpt-generate-fake-references/>>. Acesso em: 29 mai. 2023.

HOLMES, Wayne. **The Unintended Consequences of Artificial Intelligence and Education**. (2023) Education International: Brussels, Belgium. Disponível em: <<https://discovery.ucl.ac.uk/id/eprint/10179267/1/Holmes%20-%202023%20-%20The%20Unintended%20Consequences%20of%20Artificial%20Intellig.pdf>>. Acesso em 15 out. 2024.

HUSON, G. I, **Copyright**. Santa Clara High Technology Law Journal, v 35, n. 2, p. 54-78, 2018. <https://digitalcommons.law.scu.edu/chtlj/vol35/iss2/2/>.
IBM. **Confidential Data in Prompt**. Disponível em: <https://www.ibm.com/docs/en/watsonx/saas?topic=atlas-confidential-data-in-prompt>. Acesso em: 15 ago. 2024.

IMPERIAL COLLEGE LONDON. **AI & Education Guidance Hub**. Disponível em: <<https://www.imperial.ac.uk/about/education/resources/ai-education-hub/>>. Acesso em 6 de ago. 2024.

IMPERIAL COLLEGE LONDON. **Introduction to AI**. Disponível em: <<https://www.imperial.ac.uk/about/education/resources/ai-education-hub/introduction-to-ai/>>. Disponível em: 1 de out. 2024.

INTERNATIONAL ENERGY AGENCY. **Electricity 2024, Analysis and forecast to 2026**. 2024. Disponível em: <<https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf>>.

IYER, S.L. **AI enabled applications towards intelligent transportation**. *Transportation Engineering*, v.5, 2021.

JISC. **AI in tertiary education**: A summary of the current state of play. Third edition - September 2023. Disponível em: <<https://repository.jisc.ac.uk/9232/1/ai-in-tertiary-education-a-summary-of-the-current-state-of-play-september-2023.pdf>>. Acesso em 1 de out. 2024.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. **The global landscape of AI ethics guidelines**. *Nature machine intelligence*, v. 1, n. 9, p. 389-399, 2019.

KAMINSKI, Margot. **The right to explanation, explained**. Berkeley Technology Law Journal, v. 34, n. 1, 2019

KASNECI, E.; SEßLER, K.; KÜCHEMANN, S.; BANNERT, M.; DEMENTIEVA, D.; FISCHER, F.; GASSER, U.; GROH, G.; GÜNNEMANN, S.; HÜLLERMEIER, E.; et al. (2023) **ChatGPT for good? On opportunities and challenges of large language models for education**. *Learn Individ Differ* 103(102):274. p. 7

KHALIL, Mohammad; ER, Erkan. **Will ChatGPT get you caught? Rethinking of Plagiarism Detection**. Disponível em: <https://arxiv.org/pdf/2302.04335>. Acesso em 18 nov. 2024.

KUPER, Simon. **This artist is dominating AI-generated art, and he's not happy about it**. *MIT Technology Review*, 16 set. 2022. Disponível em: <https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>. Acesso em: 16 ago. 2024.

LEHR, David; OHM, Paul. **Playing with the Data**: what legal scholars should learn about machine learning. *UC Davis Law Review*, v. 51, n. 2, 2017

LORENZ, P.; PERSET, K.; BERRYHILL, J. **Initial policy considerations for generative artificial intelligence**. *OECD Artificial Intelligence Papers*, No. 1, OECD Publishing, Paris, 2023. Disponível em: <https://doi.org/10.1787/fae2d1e6-en>. Acesso em: 15 ago. 2024.

MÄNTYMÄKI, Matti et al. Defining organizational AI governance. *AI and Ethics*, v. 2, n. 4, p. 603-609, 2022.

MARANHÃO, Juliano e ALMADA, Marco. **Contribuições e Limites da Lei Geral de Proteção de Dados para a Regulação da Inteligência Artificial no Brasil**. *Revista Direito Público*, v. 20, p. 385-413, 2023.

MARANHÃO, JULIANO. **The New York Times versus OpenAI**. *Jota*, 13 jan. 2024. Disponível em: <https://legalgroundsinstitute.com/blog/the-new-york-times-versus-openai/>.

MARANHÃO, JULIANO; ABRUSIO, J. ; ALMADA, M. **Inteligência artificial aplicada ao direito e o direito da inteligência artificial**. *Suprema - Revista de Estudos Constitucionais*, v. 1, p. 154-180, 2021.

MARANHÃO, JULIANO; ALMADA, M. **Inteligência artificial no setor de saúde: ética e proteção de dados**. In: Analluza Bolivar Dallari, Gustavo Ferraz de Campos Monaco. (Org.). *LGDP na saúde*. 1ed.São Paulo: RT, 2021, v. 1, p. 357-370.

MARANHÃO, Juliano; MENEZES, Josie; ALMADA, Marco. **Inteligência artificial e concorrência: navegando em mar aberto**. *Consultor Jurídico*, 19 out. 2023. Disponível em: <https://www.conjur.com.br/2023-out-19/opinio-ia-concorrencia-navegando-mar-aberto/>. Acesso em: 15 ago. 2024.

MARANHÃO, JULIANO; NAVAS, JOÃO. **Certificação como instrumento de regulação da Inteligência Artificial no AI Act**. In *Comentários ao EU AI Act*. Rony Vainzof, Andriei Gutierrez, Gustavo Godinho e Alexandra Krastins (Coords.) Publicação prevista para 2024.

MARGONI, T.; PERRY, M. **From Music Tracks to Google Maps: Who Owns Computer Generated Works?** *Computer Law and Security Review*, v. 26, p. 621-629, 2010. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1647584.

MARTINS, José Carlos Cordeiro. **Diálogos com a inteligência artificial: usando o chatbots de inteligência artificial ChatGPT, Bard e Claude em diversos campos do conhecimento**. 2ª Edição. São Paulo. Setembro de 2023.

MAZZI, Francesca. **Authorship in artificial intelligence-generated works**: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. *The Journal of World Intellectual Property*. 2024. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jwip.12310>.

MIT - MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Generative AI & Your Course**. Disponível em: <<https://tll.mit.edu/teaching-resources/course-design/gen-ai-your-course/>>. Acesso em 6 de ago. 2024.

MIT - MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Generative AI for Teaching & Learning**: Explore our resources to unlock AI's potential for enhancing teaching and learning at MIT Sloan. Disponível em: <<https://mitsloanedtech.mit.edu/ai/>>. Acesso em 6 de out. 2024.

MOLINA, E.; COBO, C.; PINEDA, J.; ROVNER, H. (2024). AI revolution in education: What you need to know. In Digital Innovations in Education. World Bank.

MOLLEMA, Warmhold Jan Thomas. **Decolonial AI as Disenclosure**. 2024. Open Journal of Social Sciences, 12, 574-603

MULLIGAN, Deirdre K.; BAMBERGER, Kenneth A., **Procurement As Policy**: Administrative Process for Machine Learning. Berkeley Technology Law Journal, Vol. 34, 2019.

NABI, Javaid. All You Need to Know About LLM Text Generation. **Medium**, 7 ago. 2024. Disponível em: <<https://medium.com/@javaid.nabi/all-you-need-to-know-about-llm-text-generation-03b138e0ed19>>. Acesso em: 7 ago. 2024.

OECD. **Artificial intelligence, data and competition**. OECD Artificial Intelligence Papers, n. 18, Paris: OECD Publishing, 2024. Disponível em: <https://doi.org/10.1787/e7e88884-en>. Acesso em: 19 ago. 2024.

OLHAR DIGITAL. **ChatGPT bate recorde como plataforma com crescimento mais rápido da história**. Disponível em: <https://olhardigital.com.br/2023/02/02/internet-e-redes-sociais/chatgpt-bate-recorde-como-plataforma-com-crescimento-mais-rapido-da-historia/>. Acesso em 18 de ago/2024.

PAGALLO, Ugo; SCIOLLA, Jacopo Ciani; DURANTE, Massimo. The environmental challenges of AI in EU law: lessons learned from the Artificial Intelligence Act (AIA) with its drawbacks. Transforming Government: People, Process and Policy, v. 16, n. 3, p. 359-376, 2022.

PARGENDLER, Mariana. The corporate governance obsession. **J. Corp. L.**, v. 42, p. 359, 2016.

PLAGIARISMCHECK.ORG. **Google AI chatbot Gemini caught plagiarizing: PlagiarismCheck's research**. 18 de novembro de 2024. Disponível em: <<https://plagiarismcheck.org/blog/google-ai-chatbot-bard-generated-to-45-of-plagiarism-plagiarismchecks-research/>>. Acesso em 18 nov. 2024.

PEARLMAN, R. **Recognizing Artificial Intelligence (AI) as Authors and Inventors Under U.S. Intellectual Property Law**. Richmond Journal of Law & Technology, v. 24, n. 2, p. 1-38, 2018. Disponível em: <<https://jolt.richmond.edu/recognizing-artificial-intelligence-ai-as-authors-and-inventors-under-u-s-intellectual-property-law/>>.

RIGANO, Christopher, "Using Artificial Intelligence to Address Criminal Justice Needs," NIJ Journal 280, January 2019, <<https://www.nij.gov/journals/280/Pages/using-artificialintelligence-to-address-criminal-justice-needs.aspx>>.

RUSSEL GROUP. New principles on use of AI in education. 4 de julho de 2023. Disponível em: <<https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/>>. Acesso em 2 de ago. 2024.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence: a modern approach**. 3. ed. Upper Saddle River: Prentice Hall, 2010

RUSSELL, Stuart. **Artificial Intelligence and the Problem of Control**. Perspectives on Digital Humanism, v. 19, p. 1-322, 2022

SKITKA, Linda; MOSIER, Kathleen; BURDICK, Mark. **Accountability and automation bias**. International Journal of Human-Computer Studies, v. 52, p. 701–717, 2000.

SMITH, Carol. **Trustworthy by Design**. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering. 2024. p. 1-4.

STANFORD UNIVERSITY – Human Centered Artificial Intelligence (HAI). **AI Index Report**. 7ª edição, 2024. Disponível em: <<https://aiindex.stanford.edu/report/>>. Acesso em 03 jul. 2024.

SUSARLA, A.; THATCHER, R.; SARKER, S. (2023) **Editorial: the janus effect of generative AI**: charting the path for responsible conduct of scholarly activities in information systems. Inf Syst Res 34(2): 399–408. Disponível em: <<https://doi.org/10.1287/isre.2023.ed.v34.n2>> Acesso em 11 de jul. 2024. p. 400

TIMES HIGHER EDUCATION. **World University Rankings 2024**. Disponível em: <<https://www.timeshighereducation.com/world-university-rankings/2024/world-ranking>>. Acesso em 6 de ago. 2024.

UNESCO. **Artificial Intelligence and Education**: Guidance for Policy-makers. Paris: UNESCO, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000376709>. Acesso em: 15 ago. 2024.

UNESCO. **Consultation paper on AI Regulation**: emerging approaches across the world, Digital Transformation, Agosto de 2024.

UNESCO. **Guidance for generative AI in education and research**. Paris: UNESCO, 2023. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000386693>>. Acesso em 13 out. 2024. p. 28

UNIVERSIDADE DE BERKELEY, CALIFORNIA. **Appropriate Use of Generative AI Tools**. Disponível em: <<https://oercs.berkeley.edu/privacy/privacy-resources/appropriate-use-generative-ai-tools>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE BERKELEY, CALIFORNIA. **Understanding AI Writing Tools and Their Uses for Teaching and Learning at UC Berkeley**. Disponível em: <<https://teaching.berkeley.edu/understanding-ai-writing-tools-and-their-uses-teaching-and-learning-uc-berkeley>>. Acesso em 6 de out. 2024.

UNIVERSIDADE DE CAMBRIDGE. **How we use AI tools**. Disponível em: <<https://www.communications.cam.ac.uk/generative-ai-tool-guidelines>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE CAMBRIDGE. **Plagiarism and Academic Misconduct**. Disponível em: <<https://www.plagiarism.admin.cam.ac.uk/what-academic-misconduct/artificial-intelligence>>. Acesso em 1 de out. 2024

UNIVERSIDADE DE CAMBRIDGE. **Using Generative AI**: General Principals for use of Generative AI. Disponível em: <<https://blendedlearning.cam.ac.uk/guidance-support/ai-and-education/using-generative-ai>>. Acesso em 1 de out. 2024

UNIVERSIDADE DE CHICAGO. **Generative Artificial Intelligence (AI) Guidance**. Disponível em: <<https://its.uchicago.edu/generative-ai-guidance/>>. Acesso em 17 de set. 2024.

UNIVERSIDADE DE COLUMBIA – Columbia Business Scholl. **Generative AI Policy**. Disponível em: <<https://students.business.columbia.edu/office-of-student-affairs/academic-advising-and-student-success/academic-integrity/generative-ai>>

[policy#:~:text=The%20use%20of%20generative%20Artificial,the%20course%20instructor%20granting%20permission>](#). Acesso em 2 de out. 2024.

UNIVERSIDADE DE HARVARD. **Generative AI Tool Comparison**. Disponível em: <https://huit.harvard.edu/ai/tools>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE HARVARD. **Ideas for experimenting with Generative AI: Use cases and things to keep in mind**. Disponível em: <https://huit.harvard.edu/news/ai-use-cases>>. Acesso em: 6 de ago. 2024.

UNIVERSIDADE DE HARVARD. **Initial guidelines for the use of Generative AI tools at Harvard**. Disponível em: <https://huit.harvard.edu/ai/guidelines#block-boxes-1694711840>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE NOVA YORK (NYU). **Teaching with Generative AI**. Disponível em: <https://www.nyu.edu/faculty/teaching-and-learning-resources/teaching-with-generative-tools.html>>. Acesso em: 28 de set. 2024.

UNIVERSIDADE DE OXFORD. **AI in teaching and assessment**. Disponível em: <https://academic.admin.ox.ac.uk/ai-in-teaching-and-assessment>>. Acesso em 5 de jul. 2024.

UNIVERSIDADE DE OXFORD. **An introduction to the use of generative AI tools in teaching**: ten ideas on the use of AI to support students in their learning, together with suggested responses for teaching. Disponível em: <https://www.ctl.ox.ac.uk/ai-tools-in-teaching>> Acesso em 2 de out. 2024

UNIVERSIDADE DE OXFORD. **Beyond ChatGPT: State of AI in Academic Practice**. Outubro de 2023. Disponível em: https://www.ctl.ox.ac.uk/sites/default/files/ctl/documents/media/beyond_chatgpt_-_state_of_ai_for_autumn_2023_correct.pdf>. Acesso em 6 de jul. 2024.

UNIVERSIDADE DE OXFORD. **Four lessons from ChatGPT: Challenges and opportunities for educators**. Disponível em: <https://www.ctl.ox.ac.uk/article/four-lessons-from-chatgpt-challenges-and-opportunities-for-educators>>. Acesso em 27 de set. 2024.

UNIVERSIDADE DE OXFORD. **Publication and Authorship**. Disponível em: <https://researchsupport.admin.ox.ac.uk/governance/integrity/publication>>. Acesso em 10 de out. 2024.

UNIVERSIDADE DE OXFORD. **Use Generative AI services such as ChatGPT safely**. Disponível em: <https://infosec.web.ox.ac.uk/use-generative-ai-services-such-as-chatgpt-safely>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE OXFORD. **Use of generative AI tools to support learning**. Disponível em: <https://www.ox.ac.uk/students/academic/guidance/skills/ai-study>>. Acesso em 2 de out. 2024

UNIVERSIDADE DE PRINCETON. **Generative AI in Teaching and Learning Fall 2023, Faculty and Student Memo**. Disponível em: <https://odoc.princeton.edu/about/official-deans-communications/2023/generative-ai-teaching-and-learning-fall-2023-faculty-and>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE PRINCETON. **Generative AI in Teaching and Learning Fall 2023, Faculty and Student Memo**. Disponível em: <https://odoc.princeton.edu/about/official-deans-communications/2023/generative-ai-teaching-and-learning-fall-2023-faculty-and>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE PRINCETON. **Generative AI**. Disponível em: <https://mcgraw.princeton.edu/generative-ai>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE PRINCETON. **Rights, Rules, Responsibilities:** Academic Regulations. Disponível em: <<https://rrr.princeton.edu/students-and-university/24-academic-regulations>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE STANFORD. **Artificial Intelligence Teaching Guide.** Disponível em <<https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide>>. Acesso em 1 de out. 2024.

UNIVERSIDADE DE STANFORD. **Generative AI Policy Guidance.** Guidance adopted on February 16, 2023. Disponível em: <<https://communitystandards.stanford.edu/generative-ai-policy-guidance>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE STANFORD. **Responsible AI at Stanford:** Enabling innovation through AI best practices. Disponível em: <<https://uit.stanford.edu/security/responsibleai>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE YALE. **AI Guidance for Teachers.** Disponível em: <<https://poorvucenter.yale.edu/AIguidance>>. Acesso em 6 de ago. 2024.

UNIVERSIDADE DE YALE. **Guidelines for the Use of Generative AI Tools.** 20 de setembro de 2023. Disponível em: <<https://provost.yale.edu/news/guidelines-use-generative-ai-tools>>. Acesso em 6 de ago. 2024.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A.N.; KAISER, Ł.; POLOSUKHIN, I. **Attention Is All You Need.** Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.

WILLIAMS, C. (2023). **Hype, or the future of learning and teaching?** 3 Limits to AI's ability to write student essays. London School of Economics internet blog. Disponível em: <<https://kar.kent.ac.uk/99505/>>. Acesso em 11 de jul. de 2024. p. 3-4

WISCHMEYER, T and RADEMACHER, T. **Regulating Artificial Intelligence.** Springer, 2020 p. 76 a 97

WISCHMEYER, Thomas. **Artificial intelligence and transparency:** opening the black box. Regulating artificial intelligence, p. 75-101, 2020.

WORLD ECONOMIC FORUM. **The Future of Jobs Report 2023.** Geneva: World Economic Forum, 2023. Disponível em: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf. Acesso em: 15 ago. 2024.

ZAO-SANDERS, Marc. **100 Applications of Generative AI:** How people are really using the technology in the wild. Disponível em: <<https://learn.filtered.com/thoughts/ai-now-report>>. Acesso em 15 de jul. 2024.

ZARIFHONARVAR, Ali. **Economics of chatgpt:** A labor market view on the occupational impact of artificial intelligence. Journal of Electronic Business & Digital Economics, v. 3, n. 2, p. 100-116, 2024.

ZHOU, Michelle; XUE, Michelle. **Democratizing AI for Legal Professionals:** Creating Cognitive AI Legal Assistants with No Coding. In: ASAIL/LegalAIIA@ ICAIL. 2021. p. 50.

ANEXOS

ANEXO I - Governança da Inteligência Artificial como gestão de riscos

O conceito de governança assume diferentes significados a depender do contexto de uso¹²², como em seus usos em *governança corporativa*¹²³ ou nos debates envolvendo a *governança* pública, como instrumento político organizacional.¹²⁴ A governança pode se referir a práticas organizacionais em diferentes níveis, como em organizações privadas, órgãos públicos e na relação entre países.¹²⁵

No campo da Inteligência Artificial responsável, o conceito de governança diz respeito à implementação de valores, por meio de regras, práticas e procedimentos que alinhem o uso da tecnologia às melhores práticas para mitigação dos riscos inerentes ao seu desenvolvimento e emprego.¹²⁶ Portanto, a compreensão dos principais tipos de riscos inerentes ou associados à inteligência artificial é chave para desenvolver mecanismos de governança adequados.

Podemos apontar, primeiramente, os riscos inerentes à tecnologia que podem trazer consequências negativas à direitos humanos. São eles: (a) os riscos de legitimação de processos ligados à falta de transparência; (b) os riscos de legitimação de processos ligados à privacidade e proteção de dados (c) os riscos instrumentais de danos materiais ou morais, decorrentes do *design* equivocado

¹²² BEVIR, M. (2007). Governance. In BEVIR, M. **Encyclopedia of governance** (pp. 364-381). 2 volume set. Thousand Oaks, California: SAGE Publications, Inc.

DE LIMA LÚCIO, Magda et al. **Sentidos e significados de se planejar estrategicamente nas organizações públicas-planejamento estratégico sociotécnico (PLANES): análise de uma experiência.** Nau Social, v. 5, n. 9, 2014.

¹²³ PARGENDLER, Mariana. The corporate governance obsession. **J. Corp. L.**, v. 42, p. 359, 2016.

¹²⁴ BEVIR, M. **A Theory of Governance**. Los Angeles: University of California Press, 2013.

¹²⁵ BEVIR, Mark. **Governance: A Very Short Introduction**. Oxford: Oxford University Press, 2012.

¹²⁶ MÄNTYMÄKI, Matti et al. Defining organizational AI governance. **AI and Ethics**, v. 2, n. 4, p. 603-609, 2022.

DALY, Angela et al. AI, governance and ethics: global perspectives. **University of Hong Kong Faculty of Law Research Paper**, n. 2020/051, 2020.

BIRKSTEDT, Teemu et al. AI governance: themes, knowledge gaps and future agendas. **Internet Research**, v. 33, n. 7, p. 133-167, 2023.

ou de erros no treinamento ou na implementação, que resultem em falhas, (d) os riscos instrumentais ligados à discriminação.

A falta de transparência é um risco inerente à tecnologia, na medida em que a IA emprega modelos complexos de aprendizado de máquina, que trazem correlações estatísticas entre milhares de parâmetros, nem sempre inteligíveis para o raciocínio humano.¹²⁷

Além disso, também são inerentes preocupações quanto à privacidade e proteção de dados. Isso porque o desenvolvimento de ferramentas de IA usa grandes quantidades de dados que podem conter dados pessoais. E uma vez que ferramentas de IA têm grande poder inferencial, são capazes de traçar perfis ou inferir padrões de comportamento ou características individuais, dificultando o controle de finalidade do emprego dos dados.¹²⁸

Os riscos de dano advêm da própria implementação de modelos estatísticos para estipular indutivamente funções objetivas que expliquem as observações escolhidas. Uma função explicativa, que será a base das previsões e decisões, será sempre inferior ao ideal, uma vez que as diferentes funções alternativas são sempre aproximações para explicar adequadamente os dados observados e sempre há limitações quanto às informações disponíveis acerca do domínio de aplicação.¹²⁹

Por fim, os riscos ligados à discriminação decorrem da incorporação, nos dados de treinamento, de vieses sociais e estruturais, diretos ou indiretos, além

¹²⁷ WISCHMEYER, T. **Artificial Intelligence and Transparency**: opening the black box, em WISCHMEYER, T and RADEMACHER, T. *Regulating Artificial Intelligence*, Springer, 2020 p. 76 a 97

¹²⁸ MARANHÃO, Juliano e ALMADA, Marco. **Contribuições e Limites da Lei Geral de Proteção de Dados para a Regulação da Inteligência Artificial no Brasil**. *Revista Direito Público*, v. 20, p. 385-413, 2023.

¹²⁹ RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence**: a modern approach. 3. ed. Upper Saddle River: Prentice Hall, 2010

RUSSELL, Stuart. **Artificial Intelligence and the Problem of Control**. *Perspectives on Digital Humanism*, v. 19, p. 1-322, 2022

LEHR, David; OHM, Paul. **Playing with the Data**: what legal scholars should learn about machine learning. *UC Davis Law Review*, v. 51, n. 2, 2017

da possibilidade de vieses cognitivos dos próprios desenvolvedores na concepção do modelo e na escolha dos dados relevantes para treinamento.¹³⁰

Além desses riscos inerentes à inteligência artificial, há o risco de cybersegurança, decorrente do fato de que muitos sistemas de IA, sejam *softwares stand-alone* ou *softwares* embarcados em equipamentos, operarem de modo conectado, estando vulneráveis a ataques externos com objetivo malicioso de manipulação de seus resultados, o que pode trazer graves consequências em sistemas de IA voltados para a tomada de decisões automatizadas.¹³¹

É em torno desses riscos que se apontam os valores correspondentes à chamada IA confiável ou IA responsável¹³², dentre eles: transparência (ligada ao risco de opacidade), confiabilidade (ligada ao risco de erro), segurança (ligada ao risco de ataque cibernético), *fairness* ou não discriminação, privacidade e proteção de dados pessoais, e responsabilidade (no sentido de prestação de contas e reparação possíveis danos).

Os modelos de governança desenvolvidos por entidades de referência internacional como o Institute for Electric and Electronic Engineers-IEEE, the National Institute of Standards and Technology-NIST, and the International Organization for Standardization-ISO¹³³, assim como as propostas na literatura, estabelecem

¹³⁰ **ISO 24027** - Information technology – Artificial intelligence (AI) – Bias in AI systems and AI aided decision making.

MULLIGAN, Deirdre K.; BAMBERGER, Kenneth A., **Procurement As Policy**: Administrative Process for Machine Learning. Berkeley Technology Law Journal, Vol. 34, 2019.

LEHR, David; OHM, Paul. **Playing with the Data**: what legal scholars should learn about machine learning. UC Davis Law Review, v. 51, n. 2, 2017.

CITRON, Danielle Keats; PASQUALE, Frank. **The scored society**: Due process for automated predictions. Wash. L. Rev., v. 89, p. 1, 2014.

¹³¹ BONFANTI, Matteo E.; CAVELTY, Myriam Dunn; WENGER, Andreas, **Artificial intelligence and cyber-security**, in: The Routledge Social Science Handbook of AI, [s.l.]: Routledge, 2021.

¹³² EC HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. **Orientações éticas para uma IA de confiança**. 2018. Disponível em: <https://doi.org/10.2759/2686>. Acesso em: 15 jul. 2024.

¹³³ Ver por exemplo: **ISO 24027** - Information technology – Artificial intelligence (AI) – Bias in AI systems and AI aided decision making; **ISO 38507:2022** (Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations); **IEEE P2863™** - Recommended Practice for Organizational Governance of Artificial Intelligence; **IEEE P2894™** - Guide for an Architectural Framework for Explainable Artificial Intelligence; **IEEE Std 7000™-2021** - Model Process for Addressing Ethical Concerns During System Design; **IEEE 7001™-2021** - Standards for Transparency of Autonomous Systems; **IEEE P7003™** - Standard for Algorithmic Bias Considerations.

mecanismos para estruturar organizacionalmente os agentes e introduzir elementos técnicos no design de ferramentas de IA¹³⁴. Embora haja algum consenso quanto aos tipos de medida de governança para mitigação dos riscos da IA, há uma série de desafios em sua especificação e implementação para diferentes setores de atividade econômica.¹³⁵

Iniciativas de legislação em diferentes países tem adotado abordagem baseada em riscos,¹³⁶ com imposições de obrigações de governança conforme a categoria de riscos das aplicações, de modo a comprometer os agentes desenvolvedores e aplicadores de inteligência artificial com as melhores práticas de gestão e mitigação dos riscos associados à tecnologia. Baseamos a listagem abaixo nas medidas de governança previstas no *Artificial Intelligence Act Europeu*, por ser, até o agosto de 2024, a legislação mais robusta de IA no âmbito internacional e o PL 2.339/2023¹³⁷, em tramitação e debate no Senado Brasileiro. Destacam-se as seguintes medidas de governança, obrigatórias para IAs consideradas de alto risco:

(i) **sistema de gestão de risco**¹³⁸: procedimento contínuo e executado ao longo do ciclo de vida de sistema de inteligência artificial, com revisões e atualizações regulares, para identificação, estimação e análise de riscos em

¹³⁴ SMITH, Carol. **Trustworthy by Design**. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering. 2024. p. 1-4.

¹³⁵ JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. **The global landscape of AI ethics guidelines**. Nature machine intelligence, v. 1, n. 9, p. 389-399, 2019.

¹³⁶ UNESCO, Consultation paper on AI Regulation: emerging approaches across the world, Digital Transformation, Agosto de 2024.

¹³⁷ Nas medidas elencadas são destacadas as medidas de governança previstas para o AI Act. Em nota de rodapé, para além da indicação da referência da legislação europeia, também são indicadas as medidas análogas prevista no projeto de lei brasileiro. Apesar da semelhança, a legislação europeia é mais detalhada quanto às medidas e obrigações de governança a serem implementadas em comparação ao PL 2.338/2023, o qual apresenta uma redação mais genérica.

¹³⁸ AI Act, Artigo 9º.

O PL 2.338/2023 prevê especificamente na Seção IV do Capítulo IV (arts. 25 e ss.) a adoção de avaliação de impacto algorítmico para os sistemas de IA de alto risco a ser executado ao longo de todo o ciclo de vida do sistema de IA de forma que sejam implementadas, ao menos, avaliação dos riscos e benefícios aos direitos fundamentais, medidas de atenuação e efetividade das medidas de gerenciamento. Ademais, o art. 18, VI estabelece a adoção de medidas para mitigar vieses discriminatórios, incentivar a diversidade nas equipes de desenvolvimento e a política de gestão de governança para promoção da responsabilidade social e sustentável.

condições normais e de uso indevido, considerando dados de acompanhamento após a comercialização dos sistemas;

(ii) **gerenciamento de dados**¹³⁹: adoção de práticas de governança adequada aos dados utilizados para treinamento, testes e validação do sistema. Esses conjuntos de dados devem ser geridos por práticas adequadas de governança, envolvendo escolhas de concepção, origem e coleta de dados, preparação dos dados e avaliação de disponibilidade e adequação que garantam o controle, a prevenção e a correção de vieses;

(iii) **documentação**¹⁴⁰: elaboração de documento técnico contendo a descrição geral do sistema de inteligência artificial de alto risco, com respeito ao funcionamento do sistema e decisões relevantes no ciclo de vida do *software*, incluindo o processo de desenvolvimento do sistema, métodos utilizados, especificações de concepção, arquitetura do sistema, requisitos de dados, medidas de supervisão humana, alterações predeterminadas e procedimentos de validação e testagem;

(iv) **manutenção de logs automáticos (registro) de eventos**¹⁴¹: as capacidades de registro durante a operação do sistema devem permitir a identificação de situações de risco ou modificações substanciais, facilitar o acompanhamento pós-comercialização e controlar seu funcionamento;

(v) **transparência**¹⁴²: os sistemas devem ser desenvolvidos para garantir transparência suficiente, permitindo que os responsáveis pela implantação

¹³⁹ AI Act, Artigo 10º.

O PL 2.338/2023 em seu artigo 18, V determina que os sistemas de IA utilizem dados de treinamento, validação e teste que sejam adequados e representativos em relação às pessoas afetadas, seu contexto geográfico, comportamental ou funcional.

¹⁴⁰ AI Act, Artigo 11º.

O Art. 18, I do PL 2.338/2023 exige que os agentes de IA adotem documentação quanto ao funcionamento do sistema e das decisões envolvidas em sua construção durante todo seu ciclo de vida.

¹⁴¹ AI Act, Artigo 12º.

O PL 2.338/2023 em seu art. 18, II determina que as ferramentas de registro automático da operação do sistema garantam acompanhar sua acurácia e robustez e apurar resultados discriminatórios. Já o art. 18, IV estabelece o registro de fontes automatizadas e do grau de supervisão humana que tenham contribuído para os resultados.

¹⁴² AI Act, Artigo 13º.

interpretem e utilizem os resultados adequadamente, além de incluir instruções de uso; as instruções devem incluir a identidade e contato do prestador, características, capacidades e limitações do sistema, níveis de exatidão, robustez e cibersegurança, riscos conhecidos e previsíveis, capacidades técnicas para explicar resultados, desempenho em relação a grupos específicos, especificações dos dados de entrada, informações para interpretar os resultados, entre outros;

(vi) ferramentas apropriadas para interfaces homem-máquina¹⁴³: para permitir a supervisão humana em seu processo de utilização para minimizar os riscos para a saúde, a segurança ou os direitos fundamentais, por meio de medidas que permitam ao supervisor humano compreender suas capacidades e limitações, controlar seu funcionamento, detectar e corrigir anomalias, interpretar corretamente os resultados e intervir ou interromper o sistema de forma segura;

(vii) teste de segurança¹⁴⁴: incluindo mecanismos para garantir níveis adequados de precisão e cobertura; medidas de redundância que incluam planos de reserva para assegurar sua solidez; e medidas de cibersegurança destinadas a preservar a integridade do sistema contra intervenções de terceiros não autorizados.

O art. 18, VII do PL 2.338/2023 prevê a adoção de medidas técnicas para viabilizar a explicabilidade dos resultados e de medidas para disponibilizar informações adequadas que permitam a interpretação dos seus resultados e funcionamento.

¹⁴³ AI Act, Artigo 14º.

Considera-se que as medidas previstas no art. 18, VII também englobem a adoção de ferramentas apropriadas para interfaces homem-máquina.

¹⁴⁴ AI Act, Artigo 15º.

O art. 18, III do PL 2.338 estabelece a adoção de realização de testes para avaliação de níveis apropriados de confiabilidade.

ANEXO II - Introdução aos Sistemas de Inteligência Artificial Generativa

Sistemas de Inteligência Artificial Generativa (IAG) ganharam popularidade a partir de 2022 com o lançamento de ferramentas geradoras de imagens a partir de requisições em texto (*prompts*) e, principalmente no final de 2022 e início de 2023, com o lançamento e explosão do uso do ChatGPT, disponibilizado pela OpenAI.

Tais sistemas são capazes de realizar uma grande variedade de tarefas que simulam o exercício de criatividade humana,¹⁴⁵ como a manipulação de atributos visuais para geração de imagens realistas, permitindo uma série de explorações criativas por artistas e a elaboração de textos com notável fluência e coerência, além de capturar e simular diferentes estilos de redação, fazendo associações que permitem novas dimensões de análise e estimulam a criatividade humana. Os modelos de IAGs trazem as seguintes inovações com potencial transformador, em relação aos modelos tradicionais de aprendizado de máquina:¹⁴⁶

(i) são sistemas genéricos ou de propósito geral, em vez de sistemas treinados para desempenhar uma tarefa específica, sendo capazes de realizar diferentes funções, a partir de treinamento com metodologias de aprendizado de máquina em grandes bases genéricas de dados;

(ii) são capazes de gerar conteúdo novo que simula, com elevada acurácia, conteúdo de textos, imagens, sons e vídeos produzidos por humanos, sendo o conteúdo automatizado indiscernível daquele realizado por humanos sem análise especializada;

¹⁴⁵ ALEMANHA. Federal Office for Information Security. **Generative AI Models: Opportunities and Risks for Industry and Authorities.** Disponível em: <https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4> Acesso em 20 de jul. de 2024.

¹⁴⁶ BRIGGS, J. KODNANI, D. **The potentially large effects of artificial intelligence on economic growth**, Goldman Sachs Global economic Analyst, 2023.

(iii) possuem interfaces amigáveis, que respondem a *inputs* formulados em linguagem natural (texto ou voz) e, assim, não necessitam do conhecimento de códigos ou elementos específicos da interface de *software*.

Existem diferentes métodos e arquiteturas computacionais para que esses sistemas gerem novos pontos de dados a partir de dados disponíveis:¹⁴⁷ Variational autoencoders (VAEs), com uma arquitetura de codificação-decodificação e inferência variacional, em que o modelo aprende representações dos dados no *input* e gera novos dados ao selecionar amostras dos dados aprendidos; *Generative Adversarial Networks*-GANs,¹⁴⁸ compostos por um gerador que cria aleatoriamente novos conteúdos e um discriminador treinado para selecionar conteúdos similares a conteúdos criados por humanos; *Diffusion Models*, em que há de, modo iterativo, um passo de geração de “ruídos” (*noising*) nos dados seguido de um passo de filtragem de ruído (*denoising*) de forma a aperfeiçoar a qualidade do *output*; Modelos de fluxo normalizados aprendem distribuições probabilísticas complexas transformando iteradamente uma base de distribuição mais simples e são treinados por estimativa de maximização de proximidade; *Recurring Neural Networks*-RNNs, geram sequências de símbolos predizendo o próximo item (token) de forma recorrente baseados em arquitetura de *codificação-decodificação*, sendo treinados por meio de aprendizagem supervisionada; *Large Language Models* (LLMs), historicamente ligados à Transformers, também são arquiteturas de *codificação-decodificação* que utilizam mecanismos de atenção para capturar dependências entre sequências do *input* e do *output*, de modo a refletir coerência contextual, possibilitando expressividade bastante superior e mais sofisticada em relação a RNNs .

¹⁴⁷ BANDI, A.; ADAPA, P.V.S.R.; KUCHI, Y.E.V.P.K. **The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges**. Future Internet 2023, 15, 260. <https://doi.org/10.3390/fi15080260>

¹⁴⁸ Tem-se usado o termo “modelos discriminativos” para se referir a IAGs, em oposição aos modelos descritivos tradicionais. Preferimos, porém, usar o termo modelos gerativos, tendo em vista que a principal diferença está justamente na capacidade desses modelos em gerar conteúdo novo, ao passo que o termo discriminativo se circunscreve aos modelos generativos adversariais, que compõem apenas uma das metodologias ou arquiteturas empregadas.

O quadro abaixo sintetiza as diferenças principais no desenvolvimento e emprego de IAGs em relação a modelos tradicionais:

	Treinamento	Modelo	Interface	Aplicações
Modelos Descritivos	Ocorre em bases de dados ligados a determinado domínio e voltado para propósitos específicos	Metodologias diversas baseadas em correlações estatísticas observadas nas bases de dados do domínio	Usuário deve adquirir conhecimento específico do sistema para realizar os comandos ou requisições	Classificadores de dados que detectam padrões e fazem predições ou adotam decisões e fazem recomendações baseadas nas predições
Modelos Generativos	Ocorre em bases de dados enormes e generalizadas (e.g. conteúdo disponível na internet); pode ser complementado com treinamento para aplicações específicas (<i>fine tuning</i>)	Metodologias e arquiteturas diversas, não determinísticas e recorrentes, de geração e seleção do conteúdo gerado otimizando aquele que mais se aproxima do conteúdo gerado por humanos	Usam grandes modelos de linguagem e processamento de linguagem natural, de modo que o usuário pode interagir e fazer requisições em linguagem ordinária e os outputs são de fácil compreensão	Geram conteúdo novo de texto, imagem, áudio, vídeo e códigos de programação indistinguíveis do conteúdo produzido por humanos

Tabela 1 Principais diferenças entre modelos de inteligência artificial descritivos e generativos

De particular interesse para os usos na prática jurídica e nas universidades são os modelos geradores de textos como o ChatGPT, Gemini, Llama, Copilot, BingAI, Claude, LeChat e Grok (xAI), treinados para gerar textos em estilo chat ou conversa online.¹⁴⁹ Os Chatbots de IAG utilizam-se de *Large Language Models* (LLMs) para gerar respostas aos *prompts* submetidos pelos usuários em seus sistemas. Usuários inserem textos ou fazem requisições em linguagem natural – os chamados *prompts* – para que os Chatbots de IAG respondam perguntas ou realizem tarefas, como sintetizar documentos ou sistematizar em tabelas, extrair excertos relevantes, produzir redações, poemas, letras de músicas, receitas

¹⁴⁹ O termo para softwares que usam linguagem natural para interagir com seus usuários é “Conversational agent”.

culinárias, realizar recomendações etc. Caso necessário, os usuários podem refinar os *prompts* inseridos para receber respostas mais adequadas aos seus interesses. Tais ferramentas têm sido integradas a outras ferramentas como editores de textos e pacotes para atividades de escritório (Copilot) ou buscadores (Gemini e BingAI).

Dizendo de modo simplificado, os grandes modelos de linguagem, em geral, aprendem a prever estatisticamente a próxima melhor palavra a partir de uma sequência de palavras.¹⁵⁰ Tais modelos beneficiaram-se do desenvolvimento de redes neurais recorrentes, em sua capacidade de conectar eventos passados e futuros em uma cadeia de processamento computacional.

O grande passo para capturar e simular habilidades humanas deu-se com a estrutura de rede neural chamada de *Transformer*, em particular por sua capacidade de estabelecer *mecanismos de atenção* (*self-attention mechanism*) no processamento de textos.¹⁵¹ A arquitetura dos Transformers possui um codificador (*encoder*) que, a partir de um *input* de texto, entrega como *output* não já o resultado, mas, para cada palavra ou unidade de texto, um estado oculto (*hidde state*) que será descartado no processo, e um decodificador (*decoder*) que, a partir do conjunto de estados ocultos que representam o *contexto*, produz os *outputs*. Graças a esse mecanismo é possível correlacionar estatisticamente palavras ou passagens de diferentes porções do texto de modo simultâneo ao se fazer predições sobre textos correlatos a serem gerados no *output* (de uma tradução ou geração de texto) e não mais apenas relacionar palavras de entrada

¹⁵⁰ GIMPEL, H.; HALL, K.; DECKER, S.; EYMAN, T.; LÄMMERMANN, L.; MÄDCHE, A.; RÖGLINGER, R.; RUINER, C.; SCHOCH, M.; SCHOOP, M.; URBACH, N.; VANDIRK, S. (2023). **Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers.** University of Hohenheim. Disponível em <https://digital.uni-hohenheim.de/fileadmin/einrichtungen/digital/Generative_AI_and_ChatGPT_in_Higher_Education.pdf>. Acesso em 10 de jul. 2024.

¹⁵¹ Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. **Attention Is All You Need.** Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.

com palavras de saída de modo sequencial, como nos modelos de redes neurais recorrentes.

Em 2018, para aperfeiçoar seus mecanismos de busca e seu tradutor, o Google desenvolveu o modelo BERT (*Bidirectional Encoder Representation from Transformers*), com uma arquitetura de larga escala de múltiplas camadas de *transformers* bidirecionais, que revolucionou o setor de Processamento de Linguagem Natural, não só pelo alto desempenho demonstrado, como também pela abertura do seu código à comunidade de programadores(as), criando o caminho para o desenvolvimento de grandes modelos de linguagem e suas aplicações. O modelo emprega aprendizado não supervisionado e seu pré-treinamento para “aprender” a língua ocorre por meio de dois processos: (i) aprender a descobrir palavras que são ocultadas aleatoriamente em textos (*Mask Language Modelling*) e (ii) aprender se duas sentenças distintas podem ser consecutivas em um texto (*Next Sentence Prediction*). Após o pré-treinamento, o modelo fundacional desenvolvido pode ser refinado para tarefas mais específicas, a partir de bases de dados de diferentes domínios.

As aplicações dos grandes modelos de linguagem permitem a produção de palavras e construção de sentenças por predição, a partir do prompt e das palavras subsequentes geradas no processo, operação que, realizada repetidas vezes, formam mensagens e conteúdos coerentes que dão fluidez ao diálogo, até que a predição do algoritmo indique que o texto deve ser finalizado, com o conteúdo mais adequado ao conteúdo (string de símbolos) presente no prompt.¹⁵²

Para seu desenvolvimento e produção de resultados de alto desempenho, os modelos fundacionais são treinados sobre enormes conjuntos de textos obtidos da própria internet, como páginas de sites, livros disponíveis *online*

¹⁵² NABI, Javaid. All You Need to Know About LLM Text Generation. **Medium**, 7 ago. 2024. Disponível em: <<https://medium.com/@javaid.nabi/all-you-need-to-know-about-llm-text-generation-03b138e0ed19>>. Acesso em: 7 ago. 2024.

(Wikipedia e BookCorpus) e postagens em redes sociais, formando-se bases de dados com bilhões de parâmetros.¹⁵³

O quadro abaixo sintetiza o volume de dados, parâmetros e capacidade computacional exigida para o desenvolvimento desses modelos:

Model	Parameters	Training data (in tokens)	Training time (in days)	Hardware (GPUs/TPUs)
LLaMA ¹⁷ (Meta)	65B	1400B	21	2048 A100 GPU
LaMDA ¹⁸ (Google)	137B	2810B	57.5	1024 TPU v3
GPT-3 ¹⁹ (OpenAI)	175B	300B	34 [estimated]	1024 A100 GPU [estimated]
MT-NLG (Microsoft/ NVIDIA) ²⁰	530B	270B	90	4480 A100 GPU

Figura 1: o volume de dados, parâmetros e capacidade computacional exigida para o desenvolvimento dos principais modelos de IA (Fonte: Competition Market Authority)¹⁵⁴

Não é difícil notar o enorme potencial de aumento de produtividade quanto a geração de conteúdo novo em tarefas que exigem criatividade. Essas ferramentas podem ser utilizadas em uma ampla gama de atividades, como desenvolvimento de atividades escolares, auxílio na redação dos mais diversos tipos de textos como livros, artigos científicos, poesias, letras de músicas, a partir da geração e edição de textos, assistente administrativo para elaboração, classificação e sistematização de documentos, programação e desenvolvimento de softwares aplicação em ciências de dados e em áreas de negócios.¹⁵⁵ Esses sistemas possuem, ainda, usos e capacidades emergentes não previstos

¹⁵³ SUSARLA, A.; THATCHER, R.; SARKER, S. (2023) **Editorial: the janus effect of generative AI: charting the path for responsible conduct of scholarly activities in information systems.** Inf Syst Res 34(2): 399–408. Disponível em: <<https://doi.org/10.1287/isre.2023.ed.v34.n2>> Acesso em 11 de jul. 2024. p. 400

¹⁵⁴ CMA. **AI Foundation Models: Initial Report.** Competition & Markets Authority, set. 2023. Disponível em: https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf Acesso em 18 de ago. 2024.

¹⁵⁵ MARTINS, José Carlos Cordeiro. **Diálogos com a inteligência artificial: usando o chatbots de inteligência artificial ChatGPT, Bard e Claude em diversos campos do conhecimento.** 2ª Edição. São Paulo. Setembro de 2023.

inicialmente, a exemplo do desenvolvimento de novas áreas do conhecimento na psicologia¹⁵⁶, aplicação em diagnósticos clínicos^{157,158,159}

Dadas essas potencialidades e facilidade de uso da ferramenta, observou-se uma explosão do uso do ChatGPT, logo após seu lançamento, tanto no setor privado, quanto no setor público.¹⁶⁰ Assim, como visto, esse recurso também pode ser aplicado às atividades vinculadas ao setor jurídico, na produção de textos, incluindo minutas de petições ou decisões judiciais, elaboração, edição e classificação de contratos, resumos, sistematização e categorização de textos, revisão e triagem de documentação, triagem de processos e, inclusive, auxílio na programação e elaboração de códigos para desenvolvimento de *softwares* voltados à automação de atividades legais.¹⁶¹

¹⁵⁶ HAGENDORFF, Thilo. **Machine psychology**: Investigating emergent capabilities and behavior in large language models using psychological methods. **arXiv preprint arXiv:2303.13988**, 2023.

¹⁵⁷ GHIM, Jong-Lyul; AHN, Sangzin. **Transforming clinical trials**: the emerging roles of large language models. *Translational and Clinical Pharmacology*, v. 31, n. 3, p. 131, 2023.

¹⁵⁸

¹⁵⁹ ALQAHTANI, Tariq et al. **The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research**. *Research in Social and Administrative Pharmacy*, v. 19, n. 8, p. 1236-1242, 2023.

¹⁶⁰ Olhar Digital. **ChatGPT bate recorde como plataforma com crescimento mais rápido da história**. Disponível em: <https://olhardigital.com.br/2023/02/02/internet-e-redes-sociais/chatgpt-bate-recorde-como-plataforma-com-crescimento-mais-rapido-da-historia/>. Acesso em 18 de ago/2024.

¹⁶¹ ZHOU, Michelle; XUE, Michelle. **Democratizing AI for Legal Professionals**: Creating Cognitive AI Legal Assistants with No Coding. In: ASAIL/LegalAIIA@ ICAIL. 2021. p. 50.

ANEXO III - Riscos e Governança de sistemas de IA Generativa de textos

IAGs de textos trazem riscos importantes, não só agravando os riscos já apurados quanto ao emprego de sistemas de aprendizado de máquina como também riscos novos, que já vem sendo apontados por relatórios internacionais.¹⁶² Tais riscos, destacados a seguir, decorrem principalmente (i) de sua natureza estocástica (quando o resultado envolve aspectos de incerteza e de variabilidade) e (ii) dos dados de treinamento.¹⁶³

(a) Viés de automação: excesso de confiança nos resultados da ferramenta

Este talvez seja o risco mais preocupante no que tange à prestação de serviço jurisdicional, tendo em vista a relevância e impacto dos documentos produzidos nos processos de adjudicação.

Os Chatbots de IAG são treinados para simular respostas e redações humanas, de modo que o elemento discriminador seleciona conteúdo que seja similar ao humano sem compromisso com a verificação dos fatos descritos no texto. Além disso, a ferramenta é incapaz de compreender os textos ou realizar inferências lógicas sobre eles.

Como indicado acima, o resultado gerado é apenas aquilo que a ferramenta prevê ser a combinação mais provável de palavras subsequentes a partir do texto formulado no *prompt* com base em treinamento realizado a partir de enorme volume de textos. Por consequência, o resultado gerado, apesar de ser indistinguível de conteúdo humano, pode conter referências falsas e despregadas da realidade, chamadas de “alucinações”.¹⁶⁴ Como a base de dados para treinar

¹⁶² ALEMANHA. Federal Office for Information Security. **Generative AI Models: Opportunities and Risks for Industry and Authorities**. Disponível em: <https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4> Acesso em 20 de jul. de 2024.

¹⁶³ BOWMAN, Samuel R. **Eight Things to Know about Large Language Models**. *Preprint, submetido em 2 abr. 2023*. 10.48550/arXiv.2304.00612.

¹⁶⁴ HILLIER, Mathew. **Why does ChatGPT generate fake references?** - TECHE. Macquarie University's Learning and Teaching blog, 10 mar. 2023. Disponível em: <<https://teche.mq.edu.au/2023/02/why-does-chatgpt-generate-fake-references/>>. Acesso em: 29 mai. 2023.

esses algoritmos é massiva, o risco permanece ainda que o treinamento do algoritmo inclua textos contendo informações sólidas.

A possibilidade de “alucinações” alia-se a imprecisões e equívocos conceituais, chamados de “vazamentos semânticos”, quando a ambiguidade ou múltiplos sentidos de uma palavra pode gerar sequências de texto impróprias que são observados principalmente em requisições aplicadas a domínio técnicos do conhecimento (e.g. sequencias que tomam o termo “expirar”, ligado no campo jurídico a prazos de validade, gerando sequencias de texto ligadas a expirar como parte do processo respiratório).¹⁶⁵

Também em função de serem probabilísticos os modelos que embasam sistemas de IAG, tais ferramentas podem apresentar falhas em aplicações mais refinadas e sofisticadas de domínios técnicos, além de serem incapazes de realizar inferências lógicas ou de senso comum. Ainda que o texto produzido possa passar tal impressão, sistemas baseados em LLMs não realizam raciocínio correspondente ao humano, apenas correlacionam estatisticamente sequências de símbolos (palavras, trechos de textos).¹⁶⁶ Por consequência, IAGs não fazem propriamente distinção entre fatos, contraposição de opiniões, deliberações ou exercício de juízo, compreensão ou análise, muito embora o resultado gerado pareça mimetizar esses processos do pensamento humano.¹⁶⁷

A seguir, pode-se ver dois exemplos de *prompts* para o ChatGPT em que há falha na resposta pela ferramenta não deter, em sua base, uma representação de conhecimento de senso comum, ou ainda, por não realizar inferências lógicas.

¹⁶⁵ SUSARLA, A.; THATCHER, R.; SARKER, S. (2023) **Editorial: the janus effect of generative AI: charting the path for responsible conduct of scholarly activities in information systems.** Inf Syst Res 34(2): 399–408. Disponível em: <<https://doi.org/10.1287/isre.2023.ed.v34.n2>> Acesso em 11 de jul. 2024. p. 404

¹⁶⁶ FLORIDI, Luciano. **AI as agency without intelligence: on ChatGPT, Large Language Models, and other generative models.** Philosophy & Technology, v. 36, n. 1, 2023

¹⁶⁷ WILLIAMS, C. (2023). **Hype, or the future of learning and teaching?** 3 Limits to AI's ability to write student essays. London School of Economics internet blog. Disponível em: <<https://kar.kent.ac.uk/99505/>>. Acesso em 11 de jul. de 2024. p. 3-4

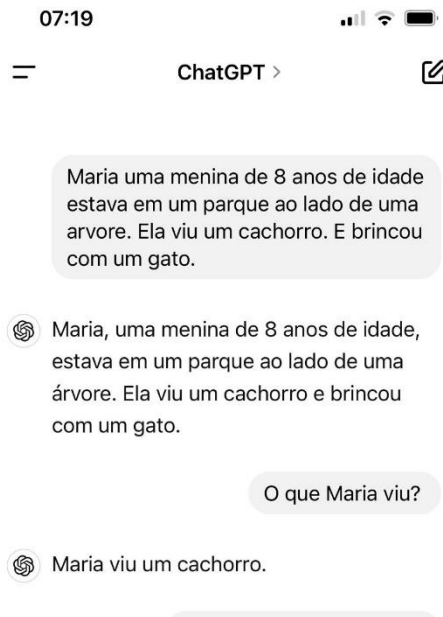


Figura 2: Exemplo de falha em senso comum

No exemplo acima, falta à ferramenta o senso comum de que, em geral, uma menina que brincou com um gato, viu o gato. Embora a ferramenta seja capaz de fazer síntese por abstração, resposta dada pela ferramenta é extrativa, relacionando apenas o pronome substantivo, com o nome correspondente no início do texto. A resposta mais adequada ao senso comum diria que Maria viu um cachorro e um gato.

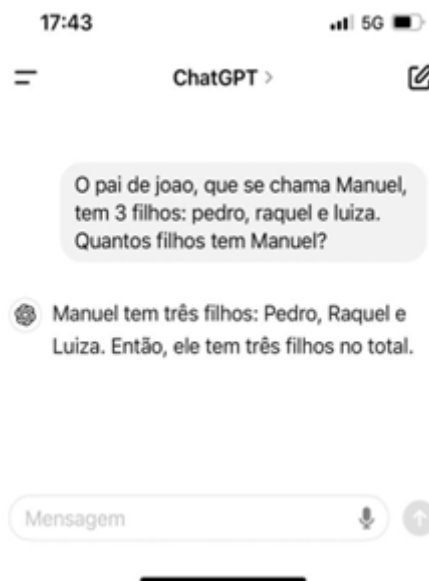


Figura 3: Exemplo de falha em inferência lógica

O exemplo traz uma limitação quanto a inferência lógica. Em qualquer ontologia simples para os conceitos de parentesco, baseada em lógica de predicados, a relação “*x é pai de y*” permitiria inferência de que “*y é filho de x*”, o que instanciaría João também como filho de Manuel, uma informação implícita que a ferramenta não foi capaz de detectar ou inferir.

Isso se aplica também ao raciocínio jurídico, de modo que IAGs aplicadas a textos ou com requisições de conteúdo jurídico são incapazes de realizar a subsunção de normas gerais à descrição de elementos presentes em casos concretos de conflito.¹⁶⁸

Apesar dessas limitações, dada a rápida difusão e popularização das ferramentas de IAGs de textos, sem o necessário treinamento e informação, aliada à capacidade de gerar textos linguisticamente corretos e convincentes sobre ampla variedade de tópicos, pode haver confiança excessiva nos resultados e desempenho do modelo ou aceitação e emprego dos *outputs*, sem questionamento, uma instância do “viés de automação” (*automation bias*).¹⁶⁹

O risco de viés de automação traz a necessidade de adoção de medidas organizacionais de treinamento dos usuários e de instâncias de revisão, quando a ferramenta é utilizada. Medidas técnicas podem ser adotadas para reduzir “alucinações” ou imprecisões e erros técnicos, a partir de treinamento específico em conjuntos de documentos de determinado domínio de aplicação (*fine-tuning*).

(b) Falta de Transparência

A transparência certamente está dentre as preocupações mais relevantes no uso de IAGs para fins acadêmicos.

Boa parte da literatura sobre transparência relacionada à modelos de aprendizado de máquina concentra-se na dificuldade de compreensão humana

¹⁶⁸ MARANHÃO, JULIANO; ABRUSIO, J. ; ALMADA, M. **Inteligência artificial aplicada ao direito e o direito da inteligência artificial**. Suprema - Revista de Estudos Constitucionais, v. 1, p. 154-180, 2021.

¹⁶⁹ SKITKA, Linda; MOSIER, Kathleen; BURDICK, Mark. **Accountability and automation bias**. International Journal of Human-Computer Studies, v. 52, p. 701–717, 2000.

dos critérios determinantes dos *outputs* em modelos mais complexos, o que vale também para aplicações à prática jurídica.¹⁷⁰ Porém, em relação aos ChatBots de IAG, como tratamos de ferramenta de conversação, que pode ser utilizado como mecanismo auxiliar a tarefas humanas na elaboração de ampla variedade de textos, inclusive textos com conteúdo jurídico, não temos aqui, ou pelo menos não deveríamos ter, decisões automatizadas que impactam direitos, de modo que a opacidade quanto ao modelo e critérios de decisão não é propriamente o fator de preocupação.

O ponto fundamental de transparência no emprego de IAGs de textos está na informação quanto a seu uso. Tal preocupação liga-se ao ponto anterior, na medida em que a composição de conteúdo técnico pela ferramenta pode conter falhas, imprecisões e “alucinações”, que demandam governança para revisão dos conteúdos gerados. A falta de transparência quanto ao uso, impede uma estruturação organizacional de revisão, o que pode resultar em falhas quanto ao resultado. Para as universidades, a falta de transparência no uso das ferramentas de IAG possui ainda implicações quanto à violação de políticas de integridade acadêmica, sujeita a sanções.

Transparência quanto ao uso diz respeito a transmitir informação sobre como e para quais finalidades a ferramenta foi usada para diferentes interlocutores. Assim, medidas relativas à transparência devem ser especificadas para cada interlocutor(a) relevante. Um dos problemas relativos à transparência sobre o uso da IA em organizações diz respeito a dúvidas de seus membros sobre a licitude desse uso, o que motiva o usuário a omitir o seu emprego. Daí a importância de divulgação na organização sobre os usos aceitos e diretrizes sobre como utilizar as ferramentas de IAGs.

Ou seja, a institucionalização da ferramenta pelas universidades, com indicações claras sobre sua aplicabilidade e limites a seus alunos pode ser

¹⁷⁰ WISCHMEYER, Thomas. **Artificial intelligence and transparency: opening the black box.** Regulating artificial intelligence, p. 75-101, 2020.
KAMINSKI, Margot. **The right to explanation, explained.** Berkeley Technology Law Journal, v. 34, n. 1, 2019

medida, ao mesmo tempo, técnica e organizacional que incentiva seu uso transparente.

(c) Riscos relacionados à proteção de privacidade e confidencialidade

Outra preocupação relacionada à IAG diz respeito a violações de privacidade e finalidade do uso dos dados.

As requisições feitas por meio de uso de *prompts* às ferramentas podem incluir dados pessoais ou mesmo conteúdo confidencial da organização¹⁷¹. Tais informações podem ser coletadas pela empresa que disponibiliza o sistema e, em alguns, casos, pode ser usada para retreinamento, fazendo com que o conteúdo confidencial possa ser incorporado ao modelo e “reproduzido” em outros usos por outros usuários. Alguns desenvolvedores e fornecedores de ferramentas de IAG de textos abrem a opção para que não sejam utilizados os dados presentes nos *prompts*, o que pode ser uma alternativa dentro da política da organização que adotar a ferramenta ou para usuário, muito embora possa haver alguma perda de desempenho tendo em vista que, para algumas aplicações, a ferramenta pode também ser aperfeiçoada com o conteúdo regularmente solicitado pelo usuário.

Dados os riscos à privacidade e confidencialidade é usualmente recomendado, como medida de governança, que as informações inseridas não façam referência a elementos identificadores pessoais ou a aspectos confidenciais da organização.¹⁷²

A mitigação desses riscos dá-se por meio de treinamento dos usuários e centralização do uso pela disponibilização “oficial” da ferramenta pela universidade, que pode analisar os termos de uso e adotar as opções de contratação menos invasivas ou arriscadas do ponto de vista de proteção da

¹⁷¹ IBM. Confidential Data in Prompt. Disponível em: <https://www.ibm.com/docs/en/watsonx/saas?topic=atlas-confidential-data-in-prompt>. Acesso em: 15 ago. 2024.

¹⁷² LORENZ, P.; PERSET, K.; BERRYHILL, J. **Initial policy considerations for generative artificial intelligence**. OECD Artificial Intelligence Papers, No. 1, OECD Publishing, Paris, 2023. Disponível em: <https://doi.org/10.1787/fae2d1e6-en>. Acesso em: 15 ago. 2024.

confidencialidade e proteção de dados pessoais, em consonância com a estrutura e política de governança de privacidade e proteção de dados da universidade.

(d) Infrações a direitos autorais

A geração de conteúdo por ferramentas de IAG tem suscitados debates quanto a autoria dos *outputs*, tema que levou a demandas de indenização autoral em tribunais estrangeiros.¹⁷³ Discute-se, entre diversas possibilidades, se a autoria caberia ao desenvolvedor do software¹⁷⁴ ou se caberia ao autor dos *prompts*,¹⁷⁵ se haveria proteção autoral em relação aos dados protegidos usados para treinamento,¹⁷⁶ em particular se a “reprodução” de estilo de autor que possa ser mencionado no *prompt* seria objeto de proteção,¹⁷⁷ se a autoria seria da

¹⁷³ Class Action. U.S. Court District of California (**Case 3:23-cv-00201-WHO**)

The New York Times Company v. OpenAI. (**Case 1:23-cv-11195**)

Umg Recordings, Inc., Capitol Records, LLC, Sony Music Entertainment, Atlantic Recording Corporation, Atlantic Records Group LLC, Rhino Entertainment LLC, The All Blacks U.S.A., Inc., Warner Music International Services Limited, And Warner Records Inc. v. Suno Inc. (**Case 1:24-cv-11611**) U.S. Copyright Office. **Zarya of the Dawn**. Washington, D.C.: U.S. Copyright Office, 2023.

¹⁷⁴ Craig, Carys J. and Kerr, Ian R., **The Death of the AI Author**. Osgoode Legal Studies Research Paper (March 25, 2019), (2021) 52(1) Ottawa Law Review 31 (<https://rdo-olr.org/2021/the-death-of-the-ai-author/>), Available at SSRN: <https://ssrn.com/abstract=3374951> or <http://dx.doi.org/10.2139/ssrn.3374951>

GINSBURG, J. C.; BUDIARDJO, L. A. **Authors and Machines**. Berkeley Technology Law Journal, v. 34, n. 2, p. 343-456, 2018. Disponível em: <https://scholarship.law.columbia.edu/faculty_scholarship/2323/>.

¹⁷⁵ ESTADOS UNIDOS. Copyright Office. **Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence**. 16 mar. 2023. Disponível em: <https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence>. Acesso em: 16 ago. 2024. U.S. Copyright Office. **Zarya of the Dawn**. Washington, D.C.: U.S. Copyright Office, 2023.

MAZZI, Francesca. **Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection**. The Journal of World Intellectual Property. 2024. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/jwip.12310>>.

BBC News. **Sony World Photography Award 2023: Winner refuses award after revealing AI creation**. Disponível em: <https://www.bbc.com/news/entertainment-arts-65296763> Acesso em 18 de ago/2024.

¹⁷⁶ The New York Times Company v. OpenAI. **Case 1:23-cv-11195**.

MARANHÃO, JULIANO. **The New York Times versus OpenAI**. Jota, 13 jan. 2024. Disponível em: <<https://legalgroundsinstitute.com/blog/the-new-york-times-versus-openai/>>.

¹⁷⁷ KUPER, Simon. **This artist is dominating AI-generated art, and he's not happy about it**. MIT Technology Review, 16 set. 2022. Disponível em: <https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>. Acesso em: 16 ago. 2024.

The New York Times Company v. OpenAI. **Case 1:23-cv-11195**

Class Action. U.S. Court District of California (**Case 3:23-cv-00201-WHO**)

própria IA,¹⁷⁸ ou ainda, se deveria ser obra de domínio público.¹⁷⁹

Ainda que o tema não tenha definição clara, o treinamento de grandes modelos de linguagem envolve grande volume de documentos, possivelmente protegidos por direitos autorais, o que pode acarretar questionamentos quanto ao conteúdo gerado.¹⁸⁰ Os questionamentos de autores têm sido dirigidos contra as empresas que disponibilizam o software e, em menor grau, a obras resultantes de *prompts* que mencionem explicitamente o estilo de determinado autor com obras protegidas.

Como trata-se de tema em aberto, é recomendável postura conservadora, assumindo-se que o uso de materiais protegidos por direitos autorais, como artigos, livros, códigos, pinturas ou músicas, por fornecedores dos modelos de IA pode infringir os direitos de propriedade intelectual, sendo adequado verificar a política e os termos de uso dos fornecedores dos sistemas em relação a abordagem para direitos autorais e como é tratada a autoria do conteúdo gerado.

(e) Reprodução de vieses

A preocupação com a incorporação de vieses também está presente em sistemas de IAG, agravado pelo fato de que o modelo é treinado com ampla base de documentos disponíveis na internet e em redes sociais, que podem deter

¹⁷⁸ PEARLMAN, R. **Recognizing Artificial Intelligence (AI) as Authors and Inventors Under U.S. Intellectual Property Law.** Richmond Journal of Law & Technology, v. 24, n. 2, p. 1-38, 2018. Disponível em: <<https://jolt.richmond.edu/recognizing-artificial-intelligence-ai-as-authors-and-inventors-under-u-s-intellectual-property-law/>>.

Abbot, Ryan. **Rational Robots: a response to interesting, renegotiated, and relational robots.** Jerusalem Review of Legal Studies, n. 1 (2022), p. 40-53. <https://doi.org/10.1093/jrls/jlac004>

DAVIES, C. R. **An Evolutionary Step in Intellectual Property Rights – Artificial Intelligence and Intellectual Property.** Computer Law & Security Review, v. 27, p. 601-630, 2011

¹⁷⁹ CLIFFORD, R. D. **Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?** Tulane Law Review. Vol. 71. 1997. p. 1675-1703. https://scholarship.law.umassd.edu/fac_pubs/77/

MARGONI, T.; PERRY, M. **From Music Tracks to Google Maps: Who Owns Computer Generated Works?** Computer Law and Security Review, v. 26, p. 621-629, 2010. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1647584.

HUSON, G. I. **Copyright.** Santa Clara High Technology Law Journal, v 35, n. 2, p. 54-78, 2018. <https://digitalcommons.law.scu.edu/chtlj/vol35/iss2/2/>.

¹⁸⁰ LORENZ, P.; PERSET, K.; BERRYHILL, J. **Initial policy considerations for generative artificial intelligence.** OECD Artificial Intelligence Papers, n. 1. Paris: OECD Publishing, 2023. Disponível em: <https://doi.org/10.1787/fae2d1e6-en>. Acesso em: 15 jul. 2024.

vieses e estereótipos quanto a grupos ou populações minorizadas, sendo mais difícil, em relação a sistemas de aprendizado desenvolvidos especificamente para determinada aplicação, fazer a curadoria para eliminar tais vieses.¹⁸¹

Diversos fatores restringem a participação de minorias na Internet, sendo mais difícil para ferramentas treinadas em grande volume de dados capturar diversidade, a não ser que haja treinamento específico para detectar padrões ligados a determinadas minorias.¹⁸² Ademais, a internet também perpetua visões de mundo hegemônicas e como IAGs líderes de mercado são, em sua grande maioria, oferecidas por empresas estrangeiras com base em documentos em língua inglesa traduzidas por IA, desdobramento do fenômeno chamado de colonialismo digital,¹⁸³ além de possíveis imprecisões ou falhas decorrentes de tradução, estereótipos ligados à cultura e visão de mundo constante naqueles documentos podem trazer notas depreciativas para minorias (em função de raça, gênero, condição econômica de grupos ou regiões etc.).¹⁸⁴

Embora em usos para domínios técnicos, como nas aplicações para fins acadêmicos, tais estereótipos discriminatórios possam ser menos frequentes, e os desenvolvedores tenham adotado cada vez mais ferramentas de moderação e detecção de vieses ou discursos abusivos tanto no conteúdo dos *prompts*, quanto nos resultados das requisições, é recomendável atentar para potenciais aspectos discriminatórios no conteúdo gerado com auxílio dessas ferramentas.¹⁸⁵

¹⁸¹ BAIDOO-ANU, David; OWUSU ANSAH, Leticia. **Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning** (January 25, 2023). Disponível em: <<https://ssrn.com/abstract=4337484>>. Acesso em 10 de jul. 2024. p. 56

¹⁸² BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. 2021. **On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?** In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. Disponível em: <<https://doi.org/10.1145/3442188.3445922>>. Acesso em 9 de jul. 2024. p. 613

¹⁸³ MOLLEMA, Warmhold Jan Thomas. **Decolonial AI as Disenclosure**. 2024. Open Journal of Social Sciences, 12, 574-603

¹⁸⁴ SUSARLA, A.; THATCHER, R.; SARKER, S. (2023) **Editorial: the janus effect of generative AI: charting the path for responsible conduct of scholarly activities in information systems**. Inf Syst Res 34(2): 399–408. Disponível em: <<https://doi.org/10.1287/isre.2023.ed.v34.n2>> Acesso em 11 de jul. 2024. p. 401

¹⁸⁵ KASNECI, E.; SEBLER, K.; KÜCHEMANN, S.; BANNERT, M.; DEMENTIEVA, D.; FISCHER, F.; GASSER, U.; GROH, G.; GÜNNEMANN, S.; HÜLLERMEIER, E.; et al. (2023) **ChatGPT for good? On**

(f) Sustentabilidade, impactos sobre o trabalho e concorrência

Além do risco de impactos a direitos individuais, as IAGs trazem mudanças substantivas em relação às preocupações ligadas a interesses coletivos, notadamente impactos ambientais e sociais.

Quanto a impactos sociais, as IAGs alteraram o espectro de profissões e ocupações ameaçadas pelo avanço tecnologia, com risco de redução de renda e aumento de desigualdades sociais.¹⁸⁶ Se, antes, os modelos de aprendizado de máquina traziam ameaça de supressão de profissões e desemprego para atividades repetitivas, agora também as profissões envolvendo elaboração criativa podem ser afetadas. Relatório recente do Goldman Sachs traz cenário contundente, em que o setor de serviços passa a ser o mais ameaçado, em particular de práticas jurídicas:

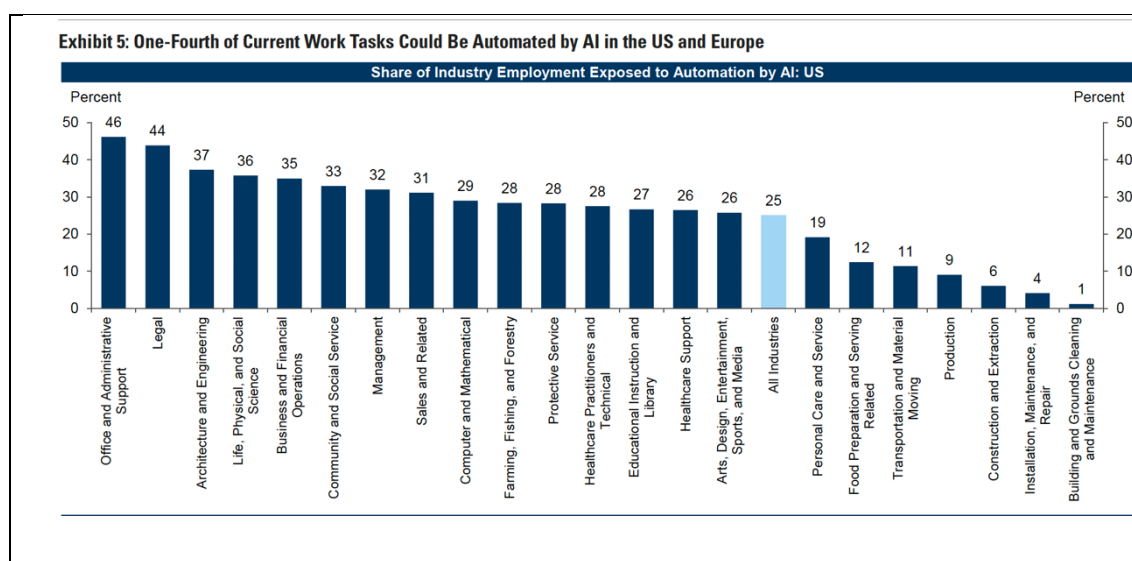


Figura 4: Projeções de supressão de profissões e desemprego por IA(Fonte: Goldman Sachs)¹⁸⁷

opportunities and challenges of large language models for education. Learn Individ Differ 103(102):274. p. 7

¹⁸⁶ ZARIFHONARVAR, Ali. **Economics of chatgpt**: A labor market view on the occupational impact of artificial intelligence. Journal of Electronic Business & Digital Economics, v. 3, n. 2, p. 100-116, 2024.

¹⁸⁷ GOLDMAN SACHS. 2023 **Global Economics Analyst**: The Potentially Large Effects of Artificial Intelligence on Economic Growth. GS Publishing. Disponível em: <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>. Acesso em: 15 ago. 2024.

Não apenas em relação à desocupação reside o risco a interesses coletivos, mas também no que toca à qualidade do trabalho. Se, de um lado, a implementação de ferramentas de inteligência artificial no ambiente de trabalho traz efeitos positivos sobre bem-estar e satisfação de trabalhadores – pela redução de tarefas perigosas e tediosas e o aumento da eficiência de grupos com menor especialização –, de outro lado, há efeitos negativos, como o subemprego e baixa remuneração das atividades de “tagueamento” no treinamento de sistemas de IA, que são aspectos explícitos da rede de trabalho, além de aspectos de trabalho implícitos, como a geração de conteúdo por usuários da internet que observados e utilizado no treinamento de sistemas de IAGs.¹⁸⁸

Embora haja bastante debate sobre o impacto social efetivo e se contraponha a possibilidade de criação de novas ocupações com o crescimento econômico propiciado pelo aumento de eficiência das ferramentas de IAG,¹⁸⁹ é importante que organizações atuem com responsabilidade social, no sentido de propiciar treinamento e educação a seus membros, tanto para que os mesmos ocupem novos espaços e oportunidades, como também para que possibilitem o melhor emprego e aproveitamento da tecnologia dentro da organização.

Além do impacto social, IAGs aumentam em muito o impacto sobre o meio ambiente em relação aos modelos tradicionais, surgindo a preocupação quanto a sua sustentabilidade (“sustentabilidade da IA”¹⁹⁰), muito embora a IA possa ser empregada para aumento de eficiência industrial, redução de emissão de carbono e combate a desmatamento e poluição ambiental (“IA para sustentabilidade”¹⁹¹).

¹⁸⁸ LORENZ, P.; PERSET, K.; BERRYHILL, J. Initial policy considerations for generative artificial intelligence. OECD Artificial Intelligence Papers, n. 1. Paris: OECD Publishing, 2023. Disponível em: <https://doi.org/10.1787/fae2d1e6-en>. Acesso em: 15 jul. 2024.

¹⁸⁹ WORLD ECONOMIC FORUM. **The Future of Jobs Report 2023**. Genebra: World Economic Forum, 2023. Disponível em: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf. Acesso em: 15 ago. 2024.

¹⁹⁰ Vide A. Van Wynsberghe, “Sustainable AI: AI for sustainability and the sustainability of AI”, cit.

¹⁹¹ Vide A. Van Wynsberghe, “Sustainable AI: AI for sustainability and the sustainability of AI”, 2021, *AI and Ethics*, Vol. 1 No. 3, pp. 213-218.

Modelos de IAG¹⁹² exigem volumosos conjuntos de dados e enorme quantidade de recursos naturais limitados, como eletricidade e água.¹⁹³ Não só o desenvolvimento de modelos de IAGs, como também sua aplicação na realização de inferências estatísticas requeridas por milhões de usuários, requerem recursos computacionais massivos, que demandam significativos recursos energéticos, associados à emissão de gases de efeito estufa.¹⁹⁴ Conforme declarado no Relatório de 2024 da International Energy Agency,¹⁹⁵ o consumo global de eletricidade dos data centers, criptomoedas e IA deve variar entre 620-1050 TWh em 2026 e, conseqüentemente, é questionável se as fontes de energia renovável poderão, nos próximos anos, satisfazer a demanda de energia dos data centers.¹⁹⁶

Com relação a recursos hídricos, estudo de 2023 revelou que o ChatGPT consome cerca de 500 ml de água para cada 20-50 perguntas e respostas simples.¹⁹⁷ A água é utilizada para resfriar os data centers e não pode ser reutilizada facilmente porque tem adição de produtos químicos. Além disso, há preocupações com o aumento de lixo eletrônico e o aumento da demanda de elementos minerais raros para produção de componentes.¹⁹⁸

Recomendações recentemente adotadas em nível internacional, como a Recomendação da UNESCO sobre a Ética da Inteligência Artificial,¹⁹⁹ a

¹⁹² IA é concebida como um termo abrangente que inclui aprendizado de máquina, redes neurais e todos os tipos de sistemas autônomos. Vide, nesse sentido, K. Crawford, **"Atlas of AI: power, politics, and the planetary costs of artificial intelligence"**, 2021, Yale University Press.

¹⁹³ Vide Seção 2 deste Capítulo.

¹⁹⁴ FREITAG, C.; BERNERS-LEE, M.; WIDDICKS, K.; KNOWLES, B.; BLAIR, G. S.; FRIDAY, A. The real climate and transformative impact of ICT: a critique of estimates, trends, and regulations. *Patterns*, v. 3, n. 8, 2022.

¹⁹⁵ International Energy Agency, **"Electricity 2024, Analysis and forecast to 2026"**, 2024, disponível em <https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf>.

¹⁹⁶ Acerca da relação entre IA e desenvolvimento de energia sustentável, vide C. Zhaoa, K. Donga, K. Wangb, R. Nepal, **"How does artificial intelligence promote renewable energy development? The role of climate finance"**, 2024, *Energy Economics*, Elsevier, vol. 133(C).

¹⁹⁷ Vide P. Li, J. Yang, M.A. Islam, S. Ren, **"Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models"**, 2023, disponível em <https://arxiv.org/pdf/2304.03271>.

¹⁹⁸ PAGALLO, Ugo; SCIOLLA, Jacopo Ciani; DURANTE, Massimo. The environmental challenges of AI in EU law: lessons learned from the Artificial Intelligence Act (AIA) with its drawbacks. *Transforming Government: People, Process and Policy*, v. 16, n. 3, p. 359-376, 2022.

¹⁹⁹ Vide a Recomendação da UNESCO sobre a Ética da Inteligência Artificial, adotada em 23 de novembro de 2021, ponto 84. 86, p. 30-31 (ponto 84).

Recomendação do Conselho da OCDE sobre Inteligência Artificial, alterada em 2024,²⁰⁰ ou as Diretrizes Éticas da UE para IA Confiável,²⁰¹ declaram expressamente que impacto ambiental dos sistemas de IA deve ser incluída entre os principais princípios de uma IA responsável e confiável.

Por fim, também em relação a interesses coletivos, as IAG provocaram enorme transformação na dinâmica concorrencial dos mercados de IA e mercados intensivos no uso de IA, ao criarem possíveis barreiras à entrada em um mercado que se acreditava ser pulverizado, em função da exigência de grandes quantidades de dados, dependência de profissionais de elevada *expertise* e capacidade computacional em larga escala. Tais características podem induzir concentração econômica,²⁰² com um fenômeno de plataformação dos mercados de IA semelhante à observada nos mercados *online*, muito embora ainda haja incerteza quanto a esse resultado,²⁰³ em função das características da oferta e emprego de IA que diferem dos serviços online e da adoção de modelos abertos por algumas empresas desenvolvedoras de IAGs.²⁰⁴

Nem todas as preocupações aqui apontadas em relação a IAGs são objeto

²⁰⁰ Vide a Recomendação do Conselho da OCDE sobre Inteligência Artificial, adotada em maio de 2019 e alterada em maio de 2024, ponto 1.1.

²⁰¹ As Diretrizes Éticas para IA Confiável foram elaboradas pelo Grupo de Especialistas de Alto Nível em Inteligência Artificial criado pela Comissão Europeia e foram publicadas em 2019, onde, entre a lista de requisitos concretos para obter uma IA confiável, bem-estar social e ambiental, incluindo sustentabilidade e respeito ao meio ambiente”, p. 14.

²⁰² FEDERAL TRADE COMMISSION; DEPARTMENT OF JUSTICE; CONSUMER FINANCIAL PROTECTION BUREAU. **Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems**. Washington, D.C.: FTC, 2023. Disponível em: https://www.ftc.gov/system/files/ftc_gov/pdf/ai-joint-statement.pdf. Acesso em: 19 ago. 2024. Competition & Markets Authority. **AI Foundation Models: Technical Update Report**. London: Department for Science, Innovation and Technology, 2023. Disponível em: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf. Acesso em: 19 ago. 2024.

OECD. **Artificial intelligence, data and competition**. OECD Artificial Intelligence Papers, n. 18, Paris: OECD Publishing, 2024. Disponível em: <https://doi.org/10.1787/e7e88884-en>. Acesso em: 19 ago. 2024.

²⁰³ ALMADA, Marco; MARANHÃO, Juliano; SARTOR, Giovanni. **Artificial intelligence**. In: PARCU, Pier Luigi; ROSSI, Maria Alessandra; BOTTA, Marco (Eds.). *Research handbook in competition & technology*. Edward Elgar Publishing, 2024. Draft, 22 September 2023.

²⁰⁴ MARANHÃO, Juliano; MENEZES, Josie; ALMADA, Marco. **Inteligência artificial e concorrência: navegando em mar aberto**. Consultor Jurídico, 19 out. 2023. Disponível em: <https://www.conjur.com.br/2023-out-19/opinio-ia-concorrencia-navegando-mar-aberto/>. Acesso em: 15 ago. 2024.

de imposição de obrigações de governança. No AI Act e no PL 2338 há medidas específicas para modelos de propósito geral, onde se enquadram as IAGs,²⁰⁵ a saber: manutenção de documentação técnica, incluindo detalhes do processo de treino, testagem e fornecimento de informações para permitir que os prestadores de sistemas de IA tenham uma boa compreensão das capacidades e limitações do sistema de IA, além do dever de cumprir a legislação de direitos autorais da UE e fornecer um resumo detalhado dos conteúdos protegidos usados no treinamento do modelo. No PL 2338, embora não haja exigência de medidas específicas de governança, é prevista diretriz quanto a sustentabilidade ambiental, inclusive com a previsão de certificação quanto ao consumo de energia limpa e políticas de emprego.²⁰⁶

Assim, embora os impactos ambientais, sociais e sobre a concorrência não estejam diretamente relacionados à aplicação de IAGs de textos pelas universidades brasileiras e não se desdobre especificamente em recomendações ou medidas pelos usuários, a atuação socialmente responsável das instituições de ensino, na contratação de sistemas de IAG, recomenda que sejam analisado, dentre os potenciais fornecedores, seu compromisso com sustentabilidade ambiental e medidas para propiciar o ambiente competitivo, como a adoção de modelos abertos, bem como a FDUSP adote iniciativas de treinamento e qualificação para seus alunos e funcionário para que estes se adaptem às novas exigências profissionais.

²⁰⁵ AI Act, Seções 2 e 3 do Capítulo V.

A Seção V do Capítulo IV do PL 2.338/2023 traz medidas de governança para sistemas de inteligência artificial de propósito geral e generativa. Os desenvolvedores desses sistemas devem demonstrar a identificação e redução de riscos, documentação de riscos, utilizar dados pessoais em conformidade com disposições da legislação de proteção de dados, reduzir o consumo de energia, utilização de recursos e geração de resíduos e fornecer documentação técnica e instruções para uso inteligível do sistema.

²⁰⁶ O Projeto de Lei nº 2338/2023 prevê, de um lado, a sustentabilidade, preservação do meio ambiente (art. 57, IV, e art. 58) e proteção do trabalho (art. 56) como diretrizes a para o desenvolvimento e emprego de IA, e, de outro, a certificação ampla abrangendo não apenas softwares como também agentes de IA (desenvolvedores, distribuidores e aplicadores). Além disso, contempla especificamente o fomento a programas de certificação para redução do impacto ambiental de sistemas de IA (art. 59). Conforme substitutivo consolidado disponibilizado pela Comissão Temporária Interna sobre Inteligência Artificial no Brasil do Senado Federal datado de 04 de junho de 2024.