



Entre o fogo e a esperança: o impacto transformador e os riscos da Inteligência Artificial

POR **THIAGO LOPES LIMA**

O surgimento da Inteligência Artificial (IA) como uma força transformadora nos últimos anos tem redefinido inúmeras esferas da sociedade contemporânea, ao mesmo tempo em que se apresenta como um dos grandes desafios da humanidade no futuro. A IA exhibe uma dualidade intrigante, similar à história de Prometeu e Pandora na mitologia grega, ao oferecer oportunidades inovadoras e, simultaneamente, desencadear ameaças que exigem uma compreensão mais profunda.

Prometeu, um titã da mitologia grega, ficou conhecido por roubar o fogo dos deuses e entregá-lo aos humanos, concedendo-lhes conhecimento e tecnologia. Esse ato desafiador enfureceu Zeus, o rei dos deuses, que condenou Prometeu a ser acorrentado a uma rocha, onde uma águia devorava seu fígado diariamente. Além disso, para punir a humanidade,

Zeus ordenou a criação de Pandora, a primeira mulher, dotada de curiosidade. Ela recebeu uma caixa e, ao abri-la impulsivamente, liberou todos os males do mundo, restando apenas a esperança em seu interior. Esses mitos ressaltam a busca humana por conhecimento e os desafios que surgem ao longo dessa jornada, expondo tanto os benefícios quanto os perigos de desafiar as ordens divinas.

O fogo, dado aos humanos por Prometeu, foi um divisor de águas no avanço da humanidade. Ele permitiu que nossos ancestrais cozinhassem alimentos, garantindo melhor nutrição e reduzindo a propagação de doenças. Também possibilitou o desenvolvimento de ferramentas e armas mais sofisticadas e foi crucial para a construção de abrigos mais seguros e aquecidos. Além disso, o fogo impulsionou a expansão cultural, pois ao seu redor histórias e conhecimentos eram compartilhados, moldando sociedades. No entanto, o mesmo fogo que trouxe progresso também carrega o potencial para destruição. Incêndios devastadores, guerras alimentadas pelo fogo e a poluição ambiental resultante da queima de combustíveis são exemplos de como esse elemento pode se tornar perigoso se mal utilizado. Da mesma forma, a IA, uma ferramenta moderna de imenso poder, tem o potencial de revolucionar a sociedade, mas, se mal utilizada ou negligenciada, pode trazer consequências desastrosas, levantando questões sobre responsabilidade e controle à medida que sua adoção cresce desenfreadamente.

Lidar com a IA exige uma compreensão profunda das “brasas” que alimentam sua chama, ou seja, dos processos internos que sustentam essa tecnologia, algo que ainda estamos longe de dominar plenamente. A complexidade dos algoritmos de IA, que muitas vezes operam como “caixas-pretas”, pode gerar “alucinações”, inventando fatos ou tomando decisões imprevisíveis, especialmente quando os dados de treinamento são de baixa qualidade ou contêm vieses. Embora o fogo seja um elemento mais simples, a humanidade levou tempo para aprender a usá-lo de forma segura. Com a IA, essa necessidade de adaptação é ainda mais urgente, já que os riscos de consequências negativas aumentam conforme a tecnologia avança sem o devido controle e compreensão. Precisamos desenvolver métodos eficazes para remover vieses e garantir que as decisões da IA sejam confiáveis, evitando que seu imenso poder cause mais danos do que benefícios, assim como o fogo em mãos despreparadas.

A adoção acelerada da IA tem evidenciado exemplos alarmantes, como os acidentes fatais causados por falhas no sistema de piloto automático da Tesla, que revelam os perigos de uma automação excessiva sem supervisão humana. No início de 2024, a IA do Google (AI

Overview) sugeriu que usuários colocassem cola na pizza e comessem pedras, exemplificando os riscos das “alucinações” dessas tecnologias, que podem resultar em recomendações absurdas e perigosas. Casos como o do iTutorGroup (empresa de educação online), onde a IA discriminou candidatos com base na idade, ou o episódio de reconhecimento facial em uma partida de futebol na Arena Batistão, em Aracaju, que levou à prisão injusta de um homem, mostram como o uso descuidado da IA pode perpetuar preconceitos e injustiças. Em outro incidente, um advogado nos EUA apresentou à corte casos jurídicos inventados pelo ChatGPT, sem perceber que eram fictícios, quase comprometendo o processo judicial. Esses episódios nos lembram que, assim como o fogo, a IA precisa ser manejada com extrema cautela e responsabilidade. Sem controles adequados, corremos o risco de ser consumidos por uma tecnologia que, embora poderosa, pode causar danos severos se não for devidamente compreendida e utilizada.

Se não tomarmos os devidos cuidados, o futuro pode trazer cenários ainda mais graves, em que falhas em IA ameacem vidas humanas, como veículos autônomos que, ao falharem, podem causar acidentes fatais em massa, ou sistemas de defesa militar que tomem decisões letais sem intervenção humana. Além disso, o uso de IA na área médica pode resultar em diagnósticos e tratamentos incorretos, agravando condições de saúde ou até mesmo levando à morte. Em outras áreas, como a jurídica, decisões automatizadas podem gerar injustiças severas, incluindo a condenação de inocentes. Essas possibilidades reforçam a necessidade de regulamentação adequada e de um desenvolvimento cuidadoso da IA para garantir que ela melhore nossas vidas, em vez de colocar em risco o que temos de mais precioso.

Certamente, ainda há muito a se desenvolver e muito aprendizado pela frente. Afinal, não foi da chama da primeira fogueira que lançamos um foguete à lua. Uma iniciativa promissora para reduzir os riscos associados à Inteligência Artificial é a Explainable Artificial Intelligence (XAI), ou Inteligência Artificial Explicável. A XAI busca desenvolver modelos de IA que sejam não apenas precisos, mas também compreensíveis, permitindo que os humanos entendam as decisões tomadas pelos sistemas. Isso é fundamental para mitigar os riscos da IA, tornando suas operações mais transparentes e confiáveis. Outra iniciativa interessante é o AI Incident Database, um banco de dados de código aberto alimentado de forma colaborativa. Essa plataforma é relevante para catalogar falhas e evitar sua repetição, inspirada em setores como a aviação, onde a análise sistemática de falhas tem melhorado significativamente a segurança ao longo do tempo. Ao aplicar a IA em diversos contextos, é

crucial adotar uma abordagem semelhante para garantir que suas futuras implementações sejam mais seguras e menos propensas a erros recorrentes. Assim, o progresso da IA depende não apenas de avanços tecnológicos, mas também de um aprendizado constante a partir dos erros do passado, removendo vieses e garantindo que os sistemas sejam treinados com dados de alta qualidade.

O imperativo é garantir que as chamas da inovação tecnológica e da transformação digital iluminem o caminho da humanidade, em vez de incendiá-lo. A esperança, que permaneceu na caixa de Pandora, é nossa aliada: com responsabilidade e supervisão, podemos assegurar que a Inteligência Artificial seja uma força positiva e uma forte aliada no enfrentamento dos Grandes Desafios da Humanidade.

PARA SE APROFUNDAR NO TEMA

“Medo, frustrado e constrangido”, diz homem detido por engano em estádio após erro do sistema de reconhecimento facial. G1 Fantástico, 2024. Disponível em: <https://g1.globo.com/fantastico/noticia/2024/04/21/medo-frustrado-e-constrangido-diz-homem-detido-por-engano-em-estadio-apos-erro-do-sistema-de-reconhecimento-facial.ghtml>. Acesso em: 29/08/2024.

Eat glue and rocks? Google’s AI hallucinations lead to bizarre recommendations. BBC News, 2024. Disponível em: <https://www.bbc.com/news/articles/cd11gzejgz4o>. Acesso em: 29/08/2024.

EEOC. Settles first-of-its-kind AI bias lawsuit for \$365,000. Bloomberg Law, 2023. Disponível em: <https://news.bloomberglaw.com/daily-labor-report/eec-settles-first-of-its-kind-ai-bias-lawsuit-for-365-000>. Acesso em: 29/08/2024.

LONGO, Luca; BRCIC, Mario; CABITZA, Federico; et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, v. 106, p. 102301, 2024. Disponível em: <https://doi.org/10.1016/j.inffus.2024.102301>. Acesso em: 29/08/2024.

MCGREGOR, Sean. Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database. 2020. Disponível em: <https://arxiv.org/pdf/2011.08512v1.pdf>. Acesso em: 29/08/2024.

MENEZES, E. D. B. de. Prometeu e Pandora entre o espelho e a máscara ou Fantasia, ordem e mistério no moinho do sentido (Notas sobre Mito e Ideologia). *Revista de História*, [S. l.], n. 118, p. 97-156, 1985. Disponível em: <https://www.revistas.usp.br/revhistoria/article/view/61330/64270>. Acesso em: 29/08/2024.

Tesla autopilot fatal crash. *The Guardian*, 2024. Disponível em: <https://www.theguardian.com/technology/2024/apr/26/tesla-autopilot-fatal-crash>. Acesso em: 29/08/2024.

US lawyer faces sanctions after ChatGPT cites fake cases in court. BBC News, 2023. Disponível em: <https://www.bbc.com/news/world-us-canada-65735769>. Acesso em: 29/08/2024.

WORLD ECONOMIC FORUM. Building trust in AI means moving beyond black-box algorithms. Here's why. Disponível em: <https://www.weforum.org/agenda/2024/04/building-trust-in-ai-means-moving-beyond-black-box-algorithms-heres-why/>. Acesso em: 29/08/2024.

THIAGO LOPES LIMA é doutorando em Administração pela Fundação Dom Cabral. Mestre em Administração, pela Fundação Pedro Leopoldo. Empreendedor do Mercado Financeiro, gestor de fundos de investimentos e carteiras administradas.