

Responsabilidade Epistêmica: os processos de conhecer mediados por sistemas de inteligência artificial

Elizabeth Maria Freire de Jesus¹, Maira Monteiro Fróes², Angélica Fonseca Dias³

^{1,2,3}Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais (NCE)
Universidade Federal do Rio de Janeiro (UFRJ)

beth@nce.ufrj.br, froes@nce.ufrj.br, angelica@nce.ufrj.br

Abstract. *Artificial intelligence technologies increasingly mediate and shape our contemporary epistemic practices. The aim of this work is to contribute to reflections on responsibilities in the processes of knowing, considering the epistemic responsibilities distributed by the individual perspectives and the governance of environments and systems.*

Resumo. Tecnologias de inteligência artificial cada vez mais medeiam e moldam as nossas práticas epistêmicas contemporâneas. O objetivo deste trabalho é contribuir com reflexões sobre as responsabilidades nos processos de conhecer, considerando as responsabilidades epistêmicas distribuídas pelas perspectivas individual e da governança de ambientes e sistemas.

1. Introdução

Os avanços em estudos de subáreas da inteligência artificial (IA) como aprendizado de máquina (*machine learning*) e seus “modos de aprender” e “produzir conhecimento”, trazem em seu bojo diversas discussões, em especial, quando se considera que a depender da tecnologia de *machine learning*, como o uso de modelos de redes neurais (*deep learning*), alguns sistemas de IA são considerados “caixas-pretas”, difíceis de serem abertas e escrutinadas mesmo diante de situações em que, com o uso da IA, alguma coisa sai errado ou produz efeitos negativos ao ser humano, ao meio ambiente e aos direitos humanos fundamentais.

As preocupações em torno da IA ética e responsável passam a ocupar espaço na agenda de discussão de diversos atores da sociedade. Em geral, diante dos benefícios das tecnologias de IA, que são inúmeros nos mais variados domínios, a discussão sobre a questão da atribuição de responsabilidade não tem recebido tanto atenção, a não ser para fins de promoção, crédito e divulgação dessas tecnologias e de grandes e poderosas corporações que não estão no Sul Global.

A ubiquidade e pervasividade das tecnologias baseadas em IA podem ser constatadas pela incorporação cada vez maior destas tecnologias em dispositivos tecnológicos e ferramentas digitais que fazem parte do nosso cotidiano. Com a captura de uma miríade de dados do ambiente, estas “aprendem” com e sobre o próprio ambiente, podendo, em alguns casos, tomar decisões e agir dentro de um determinado escopo de ação. Ou seja, sendo capazes de agir realizando algumas tarefas ou funções específicas definidas *a priori* e que são similares às realizadas pelos humanos, como interagir (chatbots), gerar frases ou textos (geração de linguagem natural), reconhecer pessoas (sistemas de reconhecimento facial) entre outros. Esse tipo de IA, denominada de IA

fraca, é aquela com que atualmente lidamos e que tem se tornado cada vez mais complexa e integrada nas nossas vidas [Ganascia 2018].

Em grande medida, a presença e agência da IA se dá sem a ciência ou consciência das pessoas, que sequer sabem que suas ações estão sendo mediadas por este tipo de tecnologia. Muitas pessoas que usam tecnologias de IA não sabem o que estão fazendo, no sentido em que não sabem o que a IA está fazendo e muito menos quais são os seus possíveis efeitos. Tem-se, portanto, aqui, um problema na dimensão do agir; do agir consciente que pressupõe conhecer as razões do agir e a autonomia; um problema eminentemente ético, na medida em que tais situações limitam, pelo menos, duas das condições basilares da ética centrada no agente, consideradas essenciais para a atribuição de responsabilidade pela ação – condição de “controle” e condição “epistêmica” [Rudy-Hiller 2018]. Uma diz respeito ao controle e a necessidade de identificar o agente responsável, e a outra diz respeito ao conhecimento do agente.

Transparência e explicabilidade dos sistemas de IA são princípios considerados essenciais para garantir certo conhecimento e controle do agente humano no e sobre o uso de tecnologias de IA. Em uma perspectiva ética, transparência e explicabilidade, no caso específico dos sistemas de IA, “não dizem respeito a como “isso funciona”, mas de como eu, um ser humano de quem se espera a capacidade de justificar-se e agir responsabilmente, posso explicar as minhas decisões” [Coeckelbergh, 2023, p.114]. Em ações mediadas por tecnologias de IA, a efetividade e instanciação destes princípios juntamente com a autonomia e o conhecimento do agente, colocam-no em condições de ser considerado como moralmente responsável por suas ações.

Em contextos onde as ações humanas são cada vez mais mediadas e atravessadas por tecnologias de IA sendo quase inescapavelmente expostas às práticas de modulação algorítmica que potencialmente influenciam suas condutas e comportamentos, incluindo desejos e pensamentos, e, portanto, os seus modos de conhecer e agir. Esses imbricamentos e seus efeitos nos modos de conhecer, e por conseguinte, nos modos de agir responsável e eticamente, colocam a necessidade de ampliar a discussão acerca da responsabilidade moral.

O objetivo deste trabalho é contribuir com algumas reflexões sobre as responsabilidades nos processos de conhecer, lançando luz sobre as responsabilidades epistêmicas em contextos onde as tecnologias de IA cada vez mais fazem parte destes processos. Acentuam-se aqui duas perspectivas distintas e inter-relacionadas: a individual e a governança, com foco na questão de como os sistemas e ambientes epistêmicos devem ser projetados para habilitar a responsabilidade de conhecer.

2. Responsabilidade epistêmica e responsabilidade epistêmica distribuída

Conhecer é um processo inerentemente social. Em grande medida, a construção de conhecimento se dá com e a partir do conhecimento de outrem, um conhecimento de segunda mão nos termos de Patrick Wilson (1983). O autor apresenta o conceito de autoridade cognitiva para referir-se a quem ou o quê atribuímos credibilidade, competência e confiabilidade ao ponto de influenciar o nosso pensamento e o nosso processo de conhecer. Autoridades cognitivas possuem um valor epistêmico, podendo ser uma pessoa, instituição, mídia, grupo, líder etc. Por outro lado, as tecnologias digitais de informação e comunicação e as tecnologias de IA cada vez mais medeiam e moldam as nossas práticas epistêmicas contemporâneas, o que nos permite conceber o processo de

conhecer, produzir, compartilhar e comunicar conhecimentos como práticas epistêmicas sociotécnicas.

A noção de responsabilidade baseia-se na escolha, e a noção de escolha é essencial ao conceito de liberdade. Na linguagem comum, chamamos alguém de “responsável” quando se pretende dizer que a pessoa em questão inclui nos motivos de seu comportamento, a previsão de possíveis efeitos dele decorrentes considerado um elemento essencial da justificação (*ex ante*) e para a responsabilização (*ex post*).

O sujeito cognoscente é epistemicamente responsável quando assume intencionalmente a responsabilidade pelo que afirma saber [Simon 2015]. Nos processos de conhecer, em contextos contemporâneos intensamente orientados e dependentes de um intrincado sistema sociotécnico, dependemos de vários agentes humanos e não-humanos tanto quanto esses agentes dependem de nós [Simon 2015]. Sistemas de IA dependem essencialmente dos nossos dados para serem treinados, “aprender” e “conhecer”, como também, muitos deles são úteis em nossos processos de conhecer e produzir conhecimento, seja na pesquisa ou na vida cotidiana.

No entanto, estamos vivenciando ou observando uma era onde sofisticados sistemas de IA, com ou sem a devida atenção e responsabilidade sobre seus usos, fazem parte da cadeia metodológica da pesquisa científica, como é o caso do modelo de linguagem ChatGPT. Tendo uma única função, qual seja, atuar como um gerador de palavras prováveis a partir de uma entrada (prompt) sem nenhum compromisso com a qualidade e a veracidade de seu “conhecimento”, o ChatGPT vem sendo utilizado em várias etapas do ciclo da pesquisa, desde a elaboração de projeto de pesquisa, como na coleta e análise de dados, na redação e comunicação [Unesco 2023]. Os usos e a utilidade dessa tecnologia de IA requer e depende uma postura crítica e responsável do pesquisador, o acionamento de um repertório de conhecimentos adequados e evidências suficientes para contradizer e questionar, colocando em prática o exercício crítico da dúvida sistemática cartesiana que recomenda que não se aceite algo como verdade antes de um exame minucioso.

O debate sobre as implicações dos emaranhados sistemas sociotécnicos cada vez mais complexos e dinâmicos que participam ou interferem nos processos de conhecer, requer colocar em perspectiva a inter-relação da responsabilidade epistêmica do sujeito cognoscente e a questão das condições e possibilidades de os sistemas e os ambientes serem projetados para promover, garantir e incentivar condutas e comportamentos responsáveis nos processos de conhecer. Nessa direção, Simon [2015] traz importantes indagações considerando duas perspectivas, individual e da governança. Na perspectiva do processo de conhecer dos indivíduos mediados pelas tecnologias de IA, a questão colocada pela autora é “o que significa ser responsável em saber?”. Na perspectiva da governança, com foco nos sistemas e nos ambientes, Simon coloca como questão principal, “o que é necessário para habilitar a responsabilidade de saber?”

A ideia de infraética de Floridi [2013] entendida como sendo uma infraestrutura ética pautada em valores, expectativas, atitudes e práticas, que se bem desenhados e alinhados com um conjunto de capacitadores morais, podem orientar, promover e incentivar ações moralmente boas e inibir ou restringir ações moralmente más, se apresenta como uma possibilidade organizativa e pragmática para a promoção de ambientes e ambiências que propiciem diálogos, compreensões e ações epistemicamente responsáveis, no nível individual e de governança. Ambientes e ambiências que facilitem

e garantam o diálogo interdisciplinar e intercultural que deem conta de reconhecer e considerar diferentes conhecimentos, bem como diferentes condições de possibilidade de conhecer, que busque formas aceitáveis de lidar com diferentes visões de mundo, ideologias e crenças e, sobretudo, que possibilite modos e meios de comunicação, de compartilhamento de informações, conhecimentos e práticas.

A necessidade premente de se construir e promover ambientes e ambiências que favoreçam as ações moralmente boas e restrinjam ou neutralizem possíveis efeitos das ações moralmente más, que são os objetivos da infraética, remete-nos à importância de se promover uma cultura de responsabilidade epistêmica como efeitos individuais e de governança, como forma a contribuir para a garantia tanto da construção de IA ética e responsável como de seus usos individuais conscientes e responsáveis.

3. Considerações finais

Tecnologias de inteligência artificial cada vez mais medeiam e moldam as nossas práticas epistêmicas contemporâneas. A discussão sobre IA ética e responsável é sobretudo uma discussão sobre responsabilidade (Jesus, 2024). As discussões em torno do comportamento humano ético nos modos de conhecer, produzir e compartilhar conhecimento em ambientes sociotécnicos é sobretudo uma discussão sobre responsabilidades epistêmicas distribuídas, onde as perspectivas individuais e de governança, em suas inter-relações são essenciais para a promoção de estruturas, políticas e ações concretas para a construção de ambientes que estimulem e possibilitem o agir ético. O reconhecimento do imbricamento dessas duas perspectivas pode ser constatado pela constituição de espaços locais, nacionais e internacionais de reflexão e discussão sobre os desafios e as implicações éticas da IA com importantes contribuições.

.

4. Referências

- Coeckelbergh, Mark. (2023). *Ética na Inteligência Artificial*. Rio de Janeiro: Editora PUC-Rio.
- Floridi, Luciano (2013). Distributed Morality in an Information Society. *Sci Eng Ethics* (19): p.727-743.
- Ganascia, Jean-Gabriel. (2018). *Inteligência Artificial: entre o mito e a realidade*. In: UNESCO. *Correio da Unesco*, n.3, p.7-9, 2018.
- Jesus, Elizabeth. M. F. de. (2024). *Responsabilidade Moral Distribuída: Contribuições para o Debate sobre Inteligência Artificial Ética e Responsável*. In: *Workshop sobre as Implicações da Computação na Sociedade (WICS)*, 5.
- Rudy-Hiller, Fernando. (2018). The epistemic condition for moral responsibility. *Stanford Encyclopedia of Philosophy*.
- Simon, Judith. (2015). Distributed Epistemic Responsibility in a Hyperconnected Era. In: Floridi, L. (eds) *The Onlife Manifesto*. Springer, Cham.
- Wilson, Patrick (1983). *Second-hand knowledge: an inquiry into cognitive authority*. Connecticut: Greenwood Press.