

INTELIGÊNCIA ARTIFICIAL E MÁ-FÉ: SATRE E A CONSCIÊNCIA DA MÁQUINA

RAMIRO DA SILVA DUARTE¹; NUNO MIGUEL PEREIRA CASTANHEIRA².

¹Mestrando PPGFIL-UFPEL – ramirosduarte@gmail.com

²Professor PPGFIL-UFPEL – npcastanheira@gmail.com

1. INTRODUÇÃO

Tendo em vista o crescimento, tanto da produção como do uso, de ferramentas de inteligência artificial (IA), o presente estudo tem por objetivo discutir se tais ferramentas representam (como vulgarmente se afirma) uma forma de progresso para a humanidade, ou se aquilo que aparentemente representa uma conquista para o desenvolvimento humano é, em verdade, uma forma de retrocesso e dominação. Para buscar uma resposta a esta inquietante questão recorreremos ao trabalho de Jean-Paul Sartre intitulado *O Ser e O Nada*, publicado originalmente em 1943, mas que ainda se mostra bastante atual, em especial quando nosso tema relaciona-se ao problema da consciência humana.

Para que possamos prosseguir nosso percurso, faz-se necessário compreender, primeiramente, que Sartre aponta que a má-fé “pode até ser o aspect normal da vida para grande número de pessoas” (SARTRE, 2008), sendo esta “certa arte de formar conceitos contraditórios [...], que unam em si determinada ideia e a negação dessa ideia” (SARTRE, 2008). Ao que o autor acrescenta, em uma conferência proferida acerca do assunto: “A má-fé é, evidentemente, uma mentira, pois dissimula a total liberdade do engajamento” (SARTRE, 2010), porém esta não é uma simples mentira consciente, afinal nela o enganador e o enganado são o mesmo, o que caracteriza a má-fé é dar-se na “unidade de uma consciência” (SARTRE, 2008).

Todavia, ainda é possível questionar qual a relevância de tais afirmações para o assunto a ser abordado, ao que Podemos, brevemente, responder com palavras de Sartre: “Pode-se julgar um homem afirmando que ele age de má-fé [...] todo homem que inventa um determinismo, é um homem de má-fé. [...] não o julgo moralmente, mas define sua má-fé como um erro.” (SARTRE, 2010). E, este julgamento, mesmo que não-moral, encaminha perspectivas morais – no que tange a nossa temática, Podemos buscar a compreensão acerca de ações relacionadas ao uso e à criação de IAs. Nessa esteira, é premente que entendamos, mesmo que em primeira aproximação, como funcionam as IAs.

Em seu livro, intitulado *Algoritmos de destruição em massa*, Cathy O’Neil aponta, acerca do uso de IAs que “quando visitamos [um] site, passamos por postagens de nossos amigos. A máquina parece ser apenas um intermediário neutro. Muitas pessoas ainda acreditam que é.” (O’NEIL, 2020). As pessoas da qual O’Neil fala parecem ser usuários de sites e aplicativos que fazem uso de IAs, porém do outro lado destes mesmos sites e aplicativos não parece ser diferente, pois, falando acerca do Facebook, Max Fisher diz que “muitos na empresa pareciam quase ignorar que os algoritmos e o design da Plataforma moldavam propositalmente as experiências e os estímulos dos usuários” (FISHER, 2023).

Tais denúncias, em ambos os lados das ferramentas controladas por IAs, nos fazem questionar até que ponto a ignorância ou a crença justificam as afirmações de

ambos os autores. Pois, como ressalta O'Neil, “comparado ao cérebro humano, o machine learning [método utilizado para aprendizado em IAs] não é particularmente eficiente” (O'NEIL, 2020), também que “ter tanta fé na capacidade dos algoritmos [...] é um caminho certo para o desastre cultural e político” (MOROZOV, 2018), como aponta Morozov. Se tais perspectivas encaminham para a verdade é preciso afirmar que a ignorância ou a fé dos indivíduos precisa ser tratada como má-fé, aos moldes sartreanos, afinal tanto ao fazer uso como ao criar ferramentas comandadas por IAs é possível que estejamos produzindo distorções na própria compreensão do que é humano.

2. METODOLOGIA

Para a realização deste estudo, foi proposta uma exegese do texto sartreano, e a posterior aplicação da teoria na interpretação dos fenômenos apontados por autores dirigidos à temática da IA. Buscamos relacionar alguns trabalhos recentes relacionados à problemática das IAs com a teoria de Jean-Paul Sartre, em especial sua compreensão da má-fé. Sendo assim, realizamos pesquisas bibliográficas de textos pertinentes, bem como da obra do autor acima mencionado. Os demais autores constam referenciados no texto, da mesma forma que nas referências bibliográficas.

3. RESULTADOS E DISCUSSÃO

Cabe ressaltar aqui que este estudo é parte componente de um projeto mais abrangente, a ser apresentado em sua totalidade como dissertação de mestrado em Filosofia. Dessa forma, gostaríamos de apresentar alguns resultados preliminares acerca da problemática apresentada acima.

Vale apontar que, mesmo que não tenhamos nos atido mais profundamente em aspectos como o surgimento e a disseminação de aplicativos como ChatGPT, ou ferramentas de criação de fotografias e áudios a partir de IA, estes também podem ser compreendidos no mesmo escopo, afinal suas implicações para a convivência entre seres humanos têm sido grande. A má-fé que conduz os seres humanos à entrega de suas informações em troca de diversão, ou de facilidades e poupança de tempo, em troca de “mapear nossos pensamentos e as nossas amizades para promover nossa produtividade” (O'NEIL, 2020). Isto porque, revelando um outro traço de má-fé, “com seu dataísmo, o regime de informação revela traços totalitários. Aspira ao saber total. Mas o saber total dataísta não é alcançado pela *narração* ideológica, mas pela *operação* algorítmica” (HAN, 2022), sem esquecermos que tal regime é sempre operado por indivíduos humanos, da mesma forma que ferramentas de inteligência artificial.

Além disso, a própria concepção de que a inteligência humana possa ser terceirizada para as máquinas, ou de que estas podem transformar a realidade sem a ação humana, denuncia mais um destes momentos que não devem ser esquecidos. Como ressalta Gabriel, “a representação atual de que sistemas de IA e as descobertas

tecnológicas ligadas a eles e que chegariam até formas de superinteligência humana” (GABRIEL, 2021) capazes de solucionar os problemas humanos não se confirma na realidade, além de ser ingênua em sua concepção. Porém, ela tem passado à realidade, através – por exemplo – das compreensões dos indivíduos acerca de sua autoimagem, ou do grau de perfeccionismo cobrado em e por todos.

4. CONCLUSÕES

Há um aspecto complexo na relação dos seres humanos com a IA, pois não é possível afirmar, em um nível mais elementar que não exista qualquer acréscimo qualitativa para a realidade humana com o advento de tais instrumentos. Porém, é preciso ressaltar que, até onde os olhos alcançam, os desenvolvimentos e usos existentes até o presente momento parecem encaminhar para diversos problemas e crises – não necessariamente novos, mas acentuados – com relação à compreensão humana de si próprio e do mundo. De igual maneira, podemos ressaltar ainda que mesmo que se ignore (propositalmente) todo o potencial prejudicial de tais ferramentas, é preciso lembrar, como alguns dos autores mencionados citam textualmente, que a inteligência humana não pode ser substituída por uma IA (algo que hoje em dia é exteriorizado pelas próprias ferramentas).

O que nos leva a mais basilar de todas as conclusões possíveis após toda a apresentação anterior: o principal ato de má-fé a que um ser humano pode incorrer em tempos de digitalização é deixar-se determinar por ferramentas de IA, bem como fazer uso indiscriminado delas no intento de eximir-se da responsabilidade por suas ações.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- FISHER, M. **A máquina do caos: Como as redes sociais reprogramaram nossa mente e nosso mundo.** São Paulo: Todavia, 2023.
- GABRIEL, M. **O sentido do pensar: a filosofia desafia a inteligência artificial.** Petrópolis, RJ: Vozes, 2021.
- HAN, B.-C. **Infocracia: digitalização e a crise da democracia.** Petrópolis, RJ: Vozes, 2022.
- MOROZOV, E. **A ascensão dos dados e a morte da política.** São Paulo: Ubu Editora, 2018.
- O'NEIL, C. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia.** Santo André, SP: Editora Rua do Sabão, 2020.
- SARTRE, J.-P. **O existencialismo é um humanismo.** Petrópolis, RJ: Vozes, 2010.
- _____. **O ser e o nada – Ensaio de ontologia fenomenológica.** Petrópolis, RJ: Vozes, 2008.