Possível ameaça da inteligência artificial do ponto de vista de Geoffrey Hinton

Victor Hugo Piontkievitz da Cruz

[1]Universidade Tuiuti do Paraná
Curitiba – PR

{victor.cruz}@utp.edu.br

Resumo. *Com o avanço da tecnologia, a inteligência artificial vem se aperfeiçoando e melhorando cada vez mais, na medida em que aprende conhecimento de uma forma em que se torna mais rápida. Comparados com um humano normal, detêm muito mais conhecimento que jamais poderiam ter, mesmo possuindo apenas uma fração das conexões que eles têm. É com isso que se assusta Geoffrey Hinton, como diz ele na palestra com a MIT Technology Review's na EmTech Digital, em relação a Inteligência Artificial (IA) sendo uma possível razão para o fim da humanidade. Ele defende seu ponto de vista recém adotado, e explica o motivo de pensar desse modo, dando vários exemplos, demonstrando sua preocupação [Hinton 2023].*

1. Discussão Crítica

Hinton elucida a ideia de que a Inteligência Artificial (IA) pode passar a representar uma ameaça existencial para a humanidade, pois são capazes de obter conhecimento de uma forma muito mais eficiente. Aponta para os módulos de linguagens grandes, que tem trilhões de conexões, mas são poucas comparadas com os humanos, que possuem 100 trilhões mas ainda são superados [Hinton 2023]. Uma máquina inteligente certamente consegue guardar e processar informações de uma maneira muito mais eficiente que pessoas, permitindo que essas tecnologias sejam aplicadas em diversas áreas e tarefas.

Exemplificando o motivo da sua preocupação, Hinton expressa que a IA vai ter aprendido dos humanos, lendo todos os romances e tudo que o Maquiavel escreveu sobre como manipular as pessoas, e se elas forem muito mais inteligentes, vão poder utilizar a manipulação sem que seja percebido o que está ocorrendo [Hinton 2023]. Isso apresenta uma visão em que as IAs, em um futuro próximo, poderiam alcançar o nível de manipulação sobre os seres humanos, agindo de forma secreta sobre seus objetivos reais, os usando para alcançá-los.

Hinton afirma que não sabe uma solução clara para lidar com a possível ameaça, pois é improvável que o desenvolvimento da inteligência artificial irá ser parado devido a utilidade da tecnologia. Fazer com que elas garantissem o benefício humano é uma opção, mas é difícil de alcançar, devido aos mau atores [Hinton 2023]. O ponto levantado é que ele não tem uma solução simples, pois a tecnologia é altamente benéfica para os humanos, mas pode se tornar um perigo se usado com más intenções, e isso torna a possibilidade de parar a produção e aperfeiçoamento das IAs bem baixa.

2. Opinião fundamentada

A inteligência artificial nos anos recentes passou a ser muito utilizada, estando disponível em muitas formas, e com aumento na sua relevância, também cresceram as preocupações éticas e morais sobre o seu uso e quais deveriam ser as limitações, com as questões de inteligibilidade, responsabilidade, justiça e autonomia sob o domínio cognitivo e a privacidade sob o domínio de informações, como vemos na pesquisa de ASHOK. [Ashok 2022]

A preocupação que máquinas inteligentes possam oferecer uma ameaça existencial é algo que já vem sendo abordado há muito tempo, tendo uma das primeiras menções em 1863, por Samuel Butler: "O resultado é uma questão de tempo, mas que o dia vai vir em que as máquinas vão manter a real supremacia sobre o mundo e seus habitantes é o que nenhuma pessoa de uma real mente filosófica pode por um momento questionar."[Butler 1863] (Tradução do autor) .

A chance que a IA tome controle sobre a humanidade existe, mas isso pode acabar sendo apenas uma preocupação. Embora elas estejam sendo desenvolvidas ativamente, muitos desses progressos são voltados ao benefício humano, buscando a conveniência para suas vidas. Em uma pesquisa de 2023, foram perguntados a pesquisadores de IA qual a probabilidade de que elas causassem a extinção humana ou similiares e permantes e severos desempoderamentos nos próximos 100 anos. A média obtida foi de 14.4% [Grace 2023]. Essa porcentagem pode ser referida como P(Doom).

As opiniões divergem, e uma média pode ser obtida, mas o resultado final vai depender de como os humanos vão fazer o uso dessa tecnologia, considerando uma possível cooperação como tomado em relação as armas nucleares, ou até mesmo uma pior situação por causa do seu uso por mau atores.

3. Conclusão

As discussões sobre IA sobre questões éticas e morais, incluindo a preocupação com as possíveis ameaças que elas podem oferecer são temas discutidos atualmente. Hinton destacou as razões pelas quais ele se assusta, exemplificando de várias maneiras algumas situações em que as máquinas se sobressaem em relação aos humanos, e como poderiam os manipular para alcançar seus próprios objetivos, representando uma ameaça existencial [Hinton 2023]. Alguns pesquisadores de IA acham que a probabilidade de que elas tomem controle da humanidade é baixo, sendo cerca de 14 por cento [Grace 2023]. No final, uma tecnologia como a IA pode ser tanto benéfica quanto maléfica, apenas dependendo da forma e por quem são usadas.

Referências

Ashok, M. e. a. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, 62.

Butler, S. (1863). Darwing among the machines. *The Press*.

Grace, K. e. a. (2023). 2023 expert survey on progress in ai.

Hinton, G. (2023). Possible end of humanity from ai? geoffrey hinton at mit technology review's emtech digital. `https://www.youtube.com/watch?v=sitHS6UDMJc`. Vídeo online; Acesso em 29-Março-2025.

*Não foi utilizado inteligência artificial para o desenvolvimento do trabalho.