

# Proposta de grupo de pesquisa no âmbito do IEA/USP

## *Inteligência artificial, ética e sociedade*

Os recentes progressos na área de inteligência artificial (IA) têm levantado questões sobre o que pode significar a “inteligência” das máquinas e que grau de liberdade ou autonomia elas poderiam vir a adquirir em um futuro próximo. Entre novembro e dezembro de 2014, o físico inglês Stephen Hawking disse à BBC que (tradução livre) "o desenvolvimento da inteligência artificial total pode significar o fim da espécie humana" [1]. Em janeiro de 2015, ele assinaria, na companhia de alguns cientistas de elite em inteligência artificial, uma carta aberta reivindicando uma agenda de pesquisa acerca dos impactos sociais da IA [2]. Esses foram eventos que podem ser vistos como compondo um marco histórico de guinada com relação à relevância e à seriedade com que o tema é tratado no domínio público. Em 28 de setembro de 2016, algumas das maiores empresas de tecnologia do mundo (Google, Facebook, Amazon, IBM, Microsoft) anunciaram a formação de um consórcio de (tradução livre) "parceria em inteligência artificial para beneficiar as pessoas e a sociedade" [3].

De fato, diversas realizações da inteligência artificial têm sido concretizadas, ainda que em data distinta daquela prevista por pioneiros como Herbert Simon [4]. Novas questões sobre as fronteiras da inteligência, da liberdade e dos direitos e deveres de robôs e IAs fazem voltar à tona a questão ontológica sobre o que o ser humano realmente é (digamos, em comparação com as máquinas), nesse novo contexto de uma civilização tecnológica [5]. Ao mesmo tempo em que é oportuno nos colocarmos tais questões fundamentais, cabe discutirmos, desde agora, sobre como a inteligência artificial *já* impacta em nosso cotidiano [6]. Pode-se dizer que o advento de tecnologias móveis e digitais de informação e comunicação das últimas décadas tiveram impacto nos modos como trabalhamos, como estudamos, como designamos nosso tempo de lazer, como escolhemos viagens, filmes, casas e até pessoas para nos relacionarmos (basta considerar a miríade de aplicativos disponíveis em *smartphones*). Há nesse cenário, é claro, uma questão mais ampla sobre os usos contemporâneos da tecnologia, que não envolve necessariamente aquilo que se pode chamar propriamente de inteligência artificial. Entretanto, encampando aquilo que há mais de ponta em termos de tecnologia e automação --- ao reproduzir modos de aprendizagem empregados por nós mesmos, seres humanos ---, cabe indagar, em cada um dos usos tecnológicos mencionados, sobre o papel e a importância que a inteligência artificial tem ou pode vir a ter.

Ora, o avanço da inteligência artificial deve-se em grande parte justamente a um aumento expressivo na disponibilidade de dados produzidos por nós mesmos através da expansão dos usos da Internet (por exemplo, em redes sociais, enciclopédias virtuais etc.; cf. [7]), bem como das tecnologias móveis e digitais já mencionadas. Tudo isso nos leva também a questionar como os dados são gerados, curados, administrados e comercializados, o que, como se sabe, está ligado à consciência (e ao consentimento) do usuário sobre usos de seus dados de maneira clara, inclusive por sistemas de inteligência artificial. Cabe, portanto, falar sobre uma ética dos dados, em particular, uma *ética da inteligência artificial* --- um tema que vem rapidamente ganhando destaque, sendo discutido em fóruns do próprio campo tecnocientífico [8], de fundações e institutos acadêmicos [9], bem como pelas grandes empresas de tecnologia [3].

Há também questões humanitárias e de soberania já pronunciadas em torno da IA e da automação. O problema das armas autônomas se apresenta noutra carta aberta notória [10], assinada por um sem-número de executivos-chefe de empresas de robótica, e direcionada à uma convenção especial das Nações Unidas sobre tipos de armas. Na carta, os signatários (sic.) "imploram" por medidas regulatórias para que o progresso tecnológico do setor não seja redirecionado para o desenvolvimento de armas autônomas. Enquanto isso, os investimentos em IA são vistos como estratégicos por nações industrializadas. Em 2017, o governo chinês iniciou uma ambiciosa política para se tornar, até 2030, (sic.) "o centro primário mundial de inovação em IA", quando ele prevê que sua indústria nacional de IA valha da ordem de 150 bilhões de dólares [11].

Por fim, vale lembrar que a questão da IA contemporânea --- com sua reflexão ontológica sobre o que o ser humano é, bem como seus impactos na estrutura ética e social da civilização global --- possui uma dimensão histórica com raízes pelo menos tão antigas quanto os primórdios da modernidade: dos autômatos discutidos por Descartes e D'Alembert nos séculos XVII e XVIII, ao engenho analítico de Babbage e as máquinas de Turing nos séculos XIX e XX, finalmente levando aos programas de aprendizagem de máquina do século XXI.

No contexto de todo esse cenário, o grupo de pesquisa aqui proposto no âmbito Instituto de Estudos Avançados da Universidade de São Paulo (IEA/USP) pretende desenvolver reflexões éticas e filosóficas, sociológicas e históricas, e de disseminação tecnocientífica sobre IA e tecnologias relacionadas. Uma lista não-exaustiva de tópicos e ações de interesse inclui (por ordenação arbitrária):

- I. Colocar em diálogo pessoas da academia, do governo, da indústria, da sociedade civil etc., que trabalhem em diferentes aspectos das questões levantadas pela IA;
- II. Servir como fórum de excelência para apresentação, análise e crítica de argumentos de cunho racional e/ou empírico acerca do papel da IA na sociedade;

- III. Reflexões acerca do valor das últimas realizações do campo para as pessoas e a sociedade, também levando em conta possíveis aspectos estratégicos, econômicos e (geo)políticos;
- IV. Estudo de implicações e riscos de tais avanços, prevendo cuidados que devem ser tomados (medidas técnicas, políticas, econômicas, regulatórias, sempre calcadas numa reflexão ética e filosófica), a fim de sinalizar uma condução equilibrada dessa tecnologia.

## Referências

- [1] Rory Cellan-Jones. "Stephen Hawking warns artificial intelligence could end mankind". *BBC*. Disponível em: <<http://www.bbc.com/news/technology-30290540>>. Último acesso em: 04.maio.2018.
- [2] Future of Life Institute. Carta aberta "Research priorities for robust and beneficial artificial intelligence". Disponível em: <<http://futureoflife.org/ai-open-letter/>>. Último acesso em: 04.maio.2018.
- [3] Alex Hern. "'Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft". *The Guardian*. Disponível em: <<http://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms>>. Último acesso em: 4.maio.2018. Ver também sítio oficial do consórcio: <<http://www.partnershiponai.org/>>.
- [4] Bernardo Gonçalves. *O Estado de S. Paulo* (Coluna de Cultura, Blog "O Estado da Arte"). Os primeiros 60 anos de feitos da inteligência artificial: Revisitando as previsões de Herbert Simon. Disponível em: <<http://cultura.estadao.com.br/blogs/estado-da-arte/os-primeiros-60-anos-de-feitos-da-inteligencia-artificial-revisitando-as-previsoes-de-herbert-simon/>>. Último acesso em: 04.maio.2018.
- [5] Hermínio Martins. *Experimentum Humanum: Civilização tecnológica e condição humana*. Coleção Sociedade e Cultura - Vol. 9. Editora Fino Traço, 2012.
- [6] Fabio Cozman. A inevitável vitória da inteligência artificial. *O Estado de S. Paulo* (Coluna de Cultura, Blog "O Estado da Arte"). Disponível em: <<http://cultura.estadao.com.br/blogs/estado-da-arte/a-inevitavel-vitoria-da-inteligencia-artificial/>>. Último acesso em: 04.maio.2018.
- [7] Eduard Hovy, Roberto Navigli, Simone Paolo Ponzetto. Collaboratively built semi-structured content and Artificial Intelligence: The story so far. *Artificial Intelligence* 194 (Jan.): 2-27, 2013. <<http://doi.org/10.1016/j.artint.2012.10.002>>.

[8] Primeira edição da conferência AAAI/ACM em Inteligência Artificial, Ética e Sociedade. Disponível em: <<http://www.aies-conference.com/>>. Último acesso em: 04.maio.2018.

[9] Ver, por exemplo, o já aqui citado "Future of Life Institute" (<http://futureoflife.org/>), que tem foco no assim-chamado "existential risk" (risco existencial para a espécie humana); o "Digital Ethics Lab" da Univ. de Oxford <<http://www.oii.ox.ac.uk/research/digital-ethics-lab/>>, e o "Kennedy Institute of Ethics" da Univ. de Georgetown <<http://kennedyinstitute.georgetown.edu/>>.

[10] Future of Life Institute. "An open letter to the United Nations Convention on certain conventional weapons". Disponível em: <<http://futureoflife.org/autonomous-weapons-open-letter-2017/>>. Último acesso em: 04.maio.2018.

[11] Christina Larson. "China's massive investment in artificial intelligence has an insidious downside". *Science*, 08.fev.2018. Disponível em: <<http://doi.org/10.1126/science.aat2458>>.