

Reflexões sobre o Emprego de Inteligência Artificial em Ciência Participativa e Cidadã

Lesandro Ponciano¹

15 de agosto de 2023

Os potenciais da Inteligência Artificial (IA) têm despertado interesse de cientistas de diversas áreas do conhecimento. Recentes avanços na área de IA levaram ao advento de tecnologias capazes de produzir informação em linguagem natural, tal como os sistemas inteligentes [Bard](#) (da Alphabet-Google), [ChatGPT](#) e [DALL-E 2](#) (ambos da OpenAI). No meio científico, uma das expectativas é que o emprego de novas tecnologias, como aprendizagem profunda (*deep learning*) e grandes modelos de linguagem (*Large Language Models*, LLM), permitam automatizar tarefas e viabilizem novos tipos de análises. Tais sistemas e tecnologias podem resultar no aumento da produtividade dos cientistas e em novas descobertas científicas.

As abordagens de IA também têm sido empregadas em ciência participativa e cidadã. Esse uso é mais proeminente em projetos conduzidos *online* e baseados em *crowdsourcing*, como [The GalaxyZoo Project](#), [The MilkyWay Project](#), [CellSpotting](#) e [Sun4All](#). Nesses projetos, os voluntários executam tarefas como identificar uma célula cancerígena, o formato de uma mancha no sol ou em galáxias. Trata-se de milhões (em alguns casos bilhões) de tarefas. Na literatura, esse tipo de atividade é chamado de *human computation task*, pois o ser humano consegue executá-las facilmente, mas um computador não é capaz de fazê-lo com o mesmo nível de acurácia, nem com os melhores algoritmos.

Com a área de IA evoluindo, aos poucos sua acurácia melhora e pode-se então passar algumas tarefas para a IA em vez dos voluntários. Isso é positivo, pois direciona-se ao voluntário apenas tarefas que são essenciais e que precisam de um ser humano, ou seja, que o computador não é capaz de executar. Essa é uma premissa por trás do uso que muitos projetos de ciência cidadã fazem de IA. Filosoficamente, essa abordagem também deixa em aberto até onde a IA vai conseguir melhorar e avançar nas tarefas que são parte da pesquisa científica e para as quais até então se pressupõe a necessidade de uma inteligência humana.

Distinguir o conceito de aprendizagem de máquina (*machine learning*) facilita na compreensão de IA e seu uso. Enquanto área de estudo, IA é uma grande área, que busca dotar computadores com inteligência nos moldes da inteligência humana. Aprendizagem de máquina é uma subárea de IA em que a inteligência deriva de treinamento com dados. Nem todo sistema inteligente implementado com IA depende de dados de treinamento. O uso de dados traz preocupações

¹ Lesandro Ponciano é Doutor e Mestre em Ciência da Computação e Bacharel em Sistemas de Informação. Tem contribuições científicas nas áreas de e-ciência (e-science) e ciência cidadã (citizen science). Estuda iniciativas de participação popular em ciência e em design por meio de abordagens de Interação Humano-Computador e Sistemas Colaborativos. Atualmente, é professor na PUC Minas, lotado no Departamento de Sistemas de Informação e Engenharia de Software.
<https://orcid.org/0000-0002-5724-0094>

adicionais com a origem e a qualidade desses dados e os riscos aos indivíduos que os forneceram (por exemplo, privacidade).

O resultado da aprendizagem de máquina é um modelo preditivo. Por sua vez, esse modelo é uma parte do sistema inteligente capaz de aplicar o que aprendeu em cenários diferentes daqueles contidos nos dados usados para treinar a máquina. Técnicas atuais permitem que tais modelos sejam muito grandes, [com bilhões de parâmetros](#), o que possibilita a capacidade de aprendizagem notável que se observa nos sistemas inteligentes atuais. Esses sistemas são capazes de observar o participante ao executar a tarefa, aprender como essa execução é feita, e gerar modelos capazes de automatizar a execução. Assim, a partir de um certo grau de treinamento, o modelo segue a execução sem a necessidade do ser humano.

Quando falamos de IA baseada em aprendizagem de máquina, não falamos mais de um *software* que foi escrito por um programador. Nesse caso, trata-se de um modelo que foi treinado por um *software* usando dados e raciocínio simbólico. O *software* escrito por um programador pode conter um *bug*, mas nesse caso o programador conhece o código, verifica, detecta o *bug* e o corrige. O modelo não é gerado por um programador, mas produzido automaticamente a partir de dados e inferências próprias. Por isso, a compreensão e controle que profissionais da computação têm dos modelos é muito inferior ao controle e compreensão que eles têm de *software*.

Aí reside um problema, é difícil saber até mesmo se há um *bug* no modelo, mesmo quando se dispõe do modelo e é possível inspecioná-lo. O segundo problema é que na maior parte das vezes não se tem o modelo em mãos. Cientistas muitas vezes usam modelos criados por terceiros e esses modelos são servidos aos pesquisadores por uma interface de *software* (como o LLM servido pelo ChatGPT) sem que se tenha acesso ao código do modelo. É muito difícil saber se um modelo tem defeito em um caso desconhecido e não é possível testar exaustivamente todos os casos.

Como se certificar de que o modelo produz respostas confiáveis? Isso é um desafio. O modelo contido em um sistema inteligente pode estar sujeito a diversos tipos de imprecisões e vieses, que são oriundos dos dados ou do processo de treinamento. Imprecisões e vieses levam o modelo a produzir respostas erradas ou tendenciosas. Por exemplo, isso acontece quando o ChatGPT erra o nome do autor de um artigo científico ou prioriza informações de países desenvolvidos em detrimento dos países em desenvolvimento. É clássico o caso do [sistema inteligente chamado Tay](#), que foi lançado pela Microsoft em 2016. Ele era uma conta no Twitter que respondia perguntas textuais e aprendia com o que as pessoas diziam. Em um dia de operação, esse sistema aprendeu a postar conteúdos racistas e nazistas e precisou ser removido pela Microsoft. Esse caso ganhou notoriedade, mas não é único.

Há um grande debate sobre até que ponto se pode confiar e escrutinar um modelo gerado em um processo de aprendizagem de máquina. Nos últimos anos, muitas pesquisas se dedicaram a propor técnicas de transparência, explicabilidade e interpretabilidade de modelos. Trata-se da IA explicável (XAI, *eXplainable AI*). O objetivo é tornar os modelos menos opacos e mais transparentes quanto ao seu comportamento ou raciocínio, de modo que sua estrutura interna seja compreendida por quem o utiliza. Isso facilitaria a identificação de imprecisões, vieses e suas causas.

No entanto, esse objetivo ainda não foi atingido. Além disso, estudos recentes têm mostrado que os modelos podem conter informações que vão além da intenção de quem os treinou. Estratégias como inferência de associação (*membership inference*) e inversão do modelo (*model inversion*) têm mostrado que modelos podem aprender mais do que é o objetivo no treinamento. Em síntese, no estágio atual, nem mesmo os profissionais que coordenam o treinamento do modelo são capazes de controlar plenamente o seu aprendizado, suas imprecisões e seus vieses.

Também é importante considerar a natureza da participação das pessoas, a propriedade dos dados e a soberania de dados. Ciência participativa e cidadã enfatiza que a participação das pessoas não é meramente utilitária. A cada participação, o ser humano aprende e satisfaz motivações intrínsecas e extrínsecas. Ao usar sistemas inteligentes em projetos de ciência cidadã, é preciso ponderar os riscos e benefícios para esses participantes. A propriedade intelectual do modelo também é relevante. Modelos derivados do treinamento com dados de pessoas ou produzidos por pessoas podem conter informações de propriedade intelectual dessas pessoas.

Há também uma questão de soberania. Por exemplo, um modelo treinado com dados de milhões de imagens de aves brasileiras é capaz de capturar as peculiaridades desses dados. A soberania que se requer sobre esse tipo de dados também deve ser requerida sobre os modelos treinados com eles. Considerando os [10 princípios de ciência cidadã propostos pela European Citizen Science Association](#), aplica-se aos modelos o princípio 7: “Dados e metadados resultantes de projetos de ciência cidadã são tornados públicos e sempre que possível publicados num formato de acesso livre”. Nesse processo de publicação, deve-se mitigar os riscos de que os modelos contenham dados sensíveis dos participantes que atuaram no treinamento.

Não há dúvidas de que a IA é uma ferramenta que pode possibilitar avanços científicos em diversas áreas do conhecimento. Ela pode e deve ser uma grande aliada dos cientistas nas iniciativas de ciência participativa e cidadã. No entanto, o uso apressado dessa ferramenta sem uma profunda compreensão, acompanhamento e avaliação do seu estado da arte pode ser danoso. Por isso, é compreensível a [carta aberta de ativistas, desenvolvedores e empresários](#) que invocam uma pausa no desenvolvimento de abordagens de IA mais poderosas do que o GPT-4. Apesar de nobre, esse esforço é pouco efetivo.

A tecnologia continuará evoluindo. Mostrando-se útil, ela será cada vez mais usada. Esforços devem ser direcionados à conscientização e à regulamentação. Ao se usar um sistema inteligente, é necessário compreender e ponderar os riscos envolvidos. Em ciência participativa e cidadã, é preciso garantir a segurança dos cidadãos, a propriedade e soberania dos dados e modelos, e a validade dos resultados das pesquisas científicas.

Conforme publicado em:

PONCIANO, Lesandro. Reflexões sobre o Emprego de Inteligência Artificial em Ciência Participativa e Cidadã. Civis Blog, Brasília, 15 agosto. 2023. Disponível em: <https://www.civis.ibict.br/blog/2023/08/14/ia-ciencia-cidada/> Acesso em: 15 agosto 2023.

Referências

- Beaumont, C. N., Goodman, A. A., Kendrew, S., Williams, J. P., & Simpson, R. (2014). The milky way project: Leveraging citizen science and machine learning to detect interstellar bubbles. *Astrophysical Journal, Supplement Series*, 214(1). <https://doi.org/10.1088/0067-0049/214/1/3>
- OECD (2023), "Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research", OECD Publishing, Paris. Disponível em: <https://doi.org/10.1787/a8d820bd-en>. Acesso em: 7 Ago. 2023
- Ponciano, L. (2022). A participação popular nas Ciências Exatas e Informática e seus efeitos no conhecimento científico e tecnológico. In M. S. BRUCK, M. CARDOSO, & M. V. DOS-SANTOS (Eds.), *Dossiê contra o negacionismo da ciência: A importância do conhecimento científico* (Vol. 1). Editora PUC Minas. <https://books.google.de/books?id=uNluEAAAQBAJ>
- Ponciano, L., & Brasileiro, F. (2018). Agreement-based credibility assessment and task replication in human computation systems. *Future Generation Computer Systems*, 87, 159–170. <https://doi.org/10.1016/j.future.2018.05.028>
- Ponciano, L., Brasileiro, F., Andrade, N., & Sampaio, L. (2014). Considering human aspects on strategies for designing and managing distributed human computation. *Journal of Internet Services and Applications*, 5(1). <https://doi.org/10.1186/s13174-014-0010-4>
- Ponciano, L., & Brasileiro, F. (2014). Finding Volunteers' Engagement Profiles in Human Computation for Citizen Science Projects. *Human Computation*, 1(2). <https://doi.org/10.15346/hc.v1i2.12>
- Rafner, J., Gajdacz, M., Kragh, G., Hjorth, A., Gander, A., Palfi, B., ... Sherson, J. (2022). Mapping Citizen Science through the Lens of Human-Centered AI. *Human Computation*, 9(1), 66–95. <https://doi.org/10.15346/hc.v9i1.133>
- Ramos, H., Fonseca, M., & Ponciano, L. (2021). Modeling and Evaluating Personas with Software Explainability Requirements. In *Communications in Computer and Information Science* (Vol. 1478 CCIS, pp. 136–149). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-92325-9_11
- Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, 8, 42200–42216. <https://doi.org/10.1109/ACCESS.2020.2976199>
- SBPC. "Inteligência Artificial". *Jornal da Ciência*, São Paulo, Nº 804. 12 Jul. 2019. Disponível em: http://jcnoticias.jornaldaciencia.org.br/wp-content/uploads/2023/07/JC_804.pdf. Acesso em: 30 Jul. 2023.
- Torney, C. J., Lloyd-Jones, D. J., Chevallier, M., Moyer, D. C., Maliti, H. T., Mwita, M., ... Hopcraft, G. C. (2019). A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, 10(6), 779–787. <https://doi.org/10.1111/2041-210X.13165>