



# Dos autômatos e robôs às redes difusas de agência no racismo algorítmico

**Tarcízio Silva**

Curador e pesquisador na Desvelar, Mestre em Comunicação (UFBA) e Doutorando em Ciências Humanas e Sociais (UFABC).

Antes de Isaac Asimov, talvez o mundo ocidental só tenha se deparado com a profundidade dos impactos sociais e especulativos da ficção científica com Mary Shelley, que inventou Frankenstein e sua criatura. Quase 100 anos antes do nascimento de Asimov, Shelley é considerada a mãe da ficção científica ao construir uma narrativa em torno de uma criatura feita de pedaços de corpos reanimados. A abordagem ficcional especulava sobre os avanços da ciência e os desafios éticos de então, mas também era uma estória em torno da inevitabilidade do destino.

O monstro de Frankenstein e o imaginário sobre os robôs de hoje devem muito aos mitos do golem na Antiguidade e aos autômatos dos séculos XVII e XVIII. Na convergência com a Revolução Industrial, porém, aqueles últimos foram responsáveis por multiplicar as metáforas que promoveram o deslocamento da agência das máquinas, escondendo trabalho, expropriação de materiais e outras variáveis político-sociais do entendimento popular sobre artefatos com determinados graus aparentes de autonomia.

Talvez um dos mais interessantes autômatos do período foi o chamado “turco mecânico”. Criado pelo inventor húngaro Wolfgang Von Kempelen, simulava o exotismo de um autômato turco que jogava xadrez de forma impecável contra oponentes humanos. O autômato era

composto de uma peça humanoide com vestuário da Turquia e uma mesa sobre a qual o tabuleiro de xadrez era operado aparentemente de forma independente. Levado a turnês por vários países europeus e EUA entre 1770 e 1854, alegadamente jogou com figuras históricas como Napoleão Bonaparte e Benjamin Franklin. Entre o fascínio e curiosidade de filósofos e cientistas da época e o medo de parte da população de que havia demônios dentro do artefato, a solução era muito mais prosaica: tratava-se apenas de uma farsa. Com o uso de fímãs e intrincadas peças, um operador humano dentro da máquina movimentava as peças de xadrez a cada jogada.

Debruçado sobre os autômatos do século XVIII, o historiador de tecnologia Edward Jones-Imhotep conta a história da ideação de “fábricas fantasmas”, imaginadas por inventores empreendedores que vislumbravam um futuro em breve onde fábricas têxteis inteiras seriam totalmente automatizadas. Para seu funcionamento, seria necessária apenas uma pessoa – até mesmo apenas uma jovem garota, para choque redobrado na época – responsável por ligar, desligar e checar que nada atrapalhe as máquinas.

O vislumbre futurístico acima foi publicado em 1745 para anunciar as obras de Jacques de Vaucanson, inventor responsável por muitos mecanismos “autômatos” que fizeram história, mas, defende Jones-Imhotep, sobretudo popularizou um *modo de falar* sobre máquinas. Em convergência com a defesa iluminista de valores particulares de governança e individualidade, os discursos da elite europeia da época em torno dos autômatos estabeleceram as bases de como falamos de máquinas e inteligência artificial hoje. Ver “máquinas como autônomas, então, historicamente significou *não* ver certos tipos de trabalho e as pessoas que o performam”<sup>1</sup>.

O próprio nome da *Amazon Mechanical Turk*, o principal fornecedor de trabalho distribuído de micro-trabalho formatado pela égide do *crowdsourcing* demonstra o aceno explícito a esta história de ocultação de trabalho repetitivo e sub-valorizado. Inúmeros esforços tidos como inovadores na inteligência artificial tratam-se de camadas de qualificação, embalagem e marketing pelo Vale do Silício sobre literalmente milhões de pequenas decisões feitas por trabalhadores de países empobrecidos com baixa regulação trabalhista, como Filipinas, Marrocos e Índia.

Ideações sobre a inteligência artificial geral, aquela que algum dia poderia replicar de forma multidimensional fatores como criatividade, autonomia e autorreflexão dos humanos, são impulsionadas em grande medida por dois fatores: o fascínio – e lucratividade decorrente – dos produtos narrativos sobre robôs superinteligentes que se revoltam contra os humanos ou, ainda, partem seus corações; e a abundância de agentes do capital financeiro que acreditam que a inteligência artificial geral está logo ali – e que seus inventores e financiadores receberão vantagens astronômicas.

Entretanto, a inteligência artificial estreita – aquela que está no nosso dia a dia em recursos de visão computacional, reconhecimento facial, processamento de linguagem natural, escores de crédito, recomendação de conteúdo, predição de riscos e afins é tão ou mais interessante

que a geral. E, sobretudo, mais premente. Argumentaria até que em grande medida a obra de Isaac Asimov é tão ou mais relevante para um mundo pré-inteligência artificial geral (que provavelmente nunca irá existir) do que para a ideação de possibilidades efusivas de robôs verdadeiramente inteligentes. As incidências multidirecionais entre tecnologia, sociedade e política são necessárias para discutir o agora.

Debates sobre atribuição, responsabilidade e agência, porém, frequentemente referenciam e bebem das ideações e reflexões basilares de pensadores como Isaac Asimov. Não por acaso, um interessante diálogo entre John M. Balkin e Frank Pasquale surgiu a partir da proposição, pelo primeiro, de *Três Leis da Robótica na Era do Big Data*, proposição na qual tais leis basilares tratam da sociedade algorítmica como um todo, incluindo IA e robôs. Balkin argumenta que as três leis, ou melhor, os três princípios legais, são: detentores de sistemas algorítmicos são fiduciários de informação; que as pessoas que usam e implementam os sistemas tem deveres públicos também para as pessoas que não são clientes, consumidoras ou usuárias finais; e que o dever público central dos detentores dos sistemas envolve evitar externalizar os custos/danos de suas operações<sup>2</sup>.

Frank Pasquale, por sua vez, debate as proposições de Balkin com uma preocupação especial em torno de transparência e atribuição. A rastreabilidade de delegações dos sistemas é essencial e se torna mais difícil quando se trata de artefatos com mobilidade e letalidade individualizada, tais como drones. Deste modo, Pasquale propõe a importante inclusão de uma quarta lei às de Balkin, que seria enunciada como “um robô [ou sistema algorítmico] deve sempre indicar a identidade de seu criador, controlador ou proprietário”<sup>3</sup>.

Deste ponto podemos evocar o que Latour, e outros teóricos da tríade ciência-tecnologia-sociedade, defende ao afirmar que “tecnologia é sociedade feita durável”<sup>4</sup>. Apesar dos esforços de pesquisadores, ativistas, desenvolvedores e juristas, a assimetria de poder entre a cidadania global frente aos grandes oligopólios de tecnologia provenientes dos grupos neoimperialistas nos países centrais impõe barreiras quase intransponíveis. O *lobby* político, econômico e cultural da lógica technochauvinista em prol da implementação de mais e mais tecnologias de ordenação, vigilância e predição através do aprendizado de máquina para geração de mais lucro e desigualdade parece implacável.

Com a continuidade de séculos de supremacismo branco a partir dos esforços coloniais e novas colonialidades centradas na Europa e EUA, não foi surpresa aos atentos a emergência do racismo algorítmico. De plataformas de mídias sociais a tecnologias carcerárias de vigilância, a distribuição racializada de poder e opressão reproduz privilégios e discriminações em prol do fornecimento de capital – financeiro e social – a poucos, em detrimento das maiorias vulnerabilizadas do Sul Global e das periferias do Norte.

Não só o racismo algorítmico intensifica e brutaliza ainda mais a distribuição de poder em prol da branquitude, como também promove a opacidade dos seus mecanismos com o apoio da lógica do aprendizado de máquina. Sistemas de ordenação de valores, oportunidades,

vigilância e respeito à humanidade são alimentados com histórico de milhares de pontos de dados desiguais, reproduzindo as dinâmicas sociais-políticas do passado em um *loop* de feedback cada vez mais opaco. Os casos auditados ou desvelados são inúmeros<sup>5</sup>, mas parecem representar apenas uma pequena parte dos sistemas algorítmicos efetivamente implementados e ainda não-auditados ou analisados. Gostaria de focar em um caso especialmente pertinente ao tema em questão.

Em 2019, pesquisadores<sup>6</sup> publicaram na revista *Science* importante estudo sobre algoritmos comerciais de predição de necessidades de cuidados médicos, provando que milhões de pacientes negros receberam atribuição a escores de risco que os prejudicava. Em determinados escores de riscos atribuídos aos avaliados em triagens médicas, pacientes negros estavam na verdade muito mais doentes do que os pacientes brancos – e em índices alarmantes. Ao investigar a origem da disparidade na base de dados, os pesquisadores descobriram que os dados que alimentavam o sistema estavam absurdamente enviesados: médicos tomavam frequentemente decisões de atribuir menos recursos a pacientes negros. Como essa métrica foi aplicada acriticamente como um proxy para supor a condição real dos pacientes, o que o algoritmo fez foi reproduzir, intensificar e esconder as decisões racistas granulares dos médicos que trabalhavam nas clínicas e hospitais fontes de dados para o sistema.

Este caso nos aponta muitos fatos e variáveis sobre a desinteligência artificial. Os mais óbvios tratam da incompetência (no mínimo) dos desenvolvedores que consideraram a métrica de “recursos gastos” como equivalente a “condições de saúde” e sobre a desumana negligência dos provedores e hospitais particulares de saúde que usaram o sistema algorítmico para otimizar custos, sem exigir auditorias prévias, uma vez que deveriam ser conscientes da fatalidade discriminatória na saúde pública.

O caso também nos conta algo além. Se a comercialização de sistemas algorítmicos tem como característica fundamental a tentativa de impor opacidade nos fluxos de trabalho que os mantém, o que podemos dizer de sistemas algorítmicos baseados em aprendizado de máquina sobre milhares ou milhões de pontos de dados de decisões racistas que já estavam em andamento? Cada vez que um médico ignorou a dor de uma pessoa negra, escolheu um procedimento menos eficaz por ser mais barato ou deu atenção apenas a seus iguais, o impacto foi direto naquele paciente e se somou, como ponto de dado, à bases que permitiriam a automatização em escala das decisões racistas<sup>7</sup>. Ironicamente, a auditoria realizada no estudo jogou luz não só sobre os horrores da desinteligência artificial, mas também sobre os horrores das decisões realizadas por cossignatários de um contrato racial<sup>8</sup> em prol da branquitude violenta antes de sequer virarem dados.

Então, apesar do que leituras apressadas da obra de Isaac Asimov e outros escritores clássicos de ficção científica podem incentivar, é preciso ir além de debates óbvios sobre temas como “direitos dos robôs” ou impactos possíveis de uma improvável inteligência artificial geral. Antes precisamos concordar com Abeba Birhane e Jelle van Dijk ao defender que deveríamos focar no “bem-estar de grupos marginalizados [...] que são desproporcionalmente

impactados pela integração da tecnologia em nossas vidas cotidianas”<sup>9</sup>. Em um mundo onde decisões racistas são novamente normalizadas ao serem transformadas em dados, perdendo as conexões de delegação e ganhando ares de neutralidade, não podemos perder de vista o compromisso de discutir as redes de poder e opressão incorporadas na tecnologia *hoje*.

## Notas

1. JONES-IMHOTEP, Edward. “The ghost factories: histories of automata and artificial life”, *History and Technology*, vol. 36, n.1, 2020, p.10.
2. BALKIN, Jack M. “The Three Laws of Robotics in the Age of Big Data”, *Ohio State Law Journal*, vol. 78, n.5, 2017.
3. PASQUALE, Frank. “Toward a fourth law of robotics: preserving attribution, responsibility, and explainability in an algorithmic society”, *Ohio State Law Journal*, vol. 78, n.5, 2017.
4. LATOUR, Bruno. “Technology is society made durable”, *The Sociological Review*, vol. 38, n.1, 1990.
5. Mantenho um mapeamento contínuo em “Linha do Tempo do Racismo Algorítmico” em <https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo/>
6. OBERMEYER, Ziad et al., “Dissecting racial bias in an algorithm used to manage the health of populations”, *Science*, v. 366, n. 6464, 2019.
7. BENJAMIN, Ruha. “Assessing risk, automating racism”, *Science*, v. 366, n. 6464, 2019.
8. MILLS, Charles W. *The racial contract*, Cornell University Press, 2014.
9. BIRHANE, Abeba; VAN DIJK, Jelle. “Robot rights? Let’s talk about human welfare instead”, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, p.6.