

Discriminação algorítmica: superando as limitações das estatísticas e cálculos probabilísticos

O necessário enfrentamento do conflito entre acurácia e justiça

Parte XI

Ana Frazão

Advogada. Professora de Direito Civil e Comercial da UnB. Ex-Conselheira do CADE.

Como se procurou demonstrar pelos artigos anteriores, julgamentos humanos quantitativos só aparentemente podem ser considerados neutros. Traduzir fenômenos humanos em números exige necessariamente escolhas valorativas. O problema é quando estas não são devidamente assumidas nem estão sujeitas ao devido escrutínio.

Também já se viu que julgar as pessoas apenas por critérios estatísticos é opção que, por si só, é fonte de grandes injustiças. Como alertam Kay e King¹, se a injustiça contra indivíduos é inerente a qualquer aplicação de discriminação estatística, a utilização dos sistemas algorítmicos aumenta ainda mais esse risco:

"The availability of big data, which enable us to learn far more about correlations - although not necessarily about causation - creates new opportunities for statistical discrimination and new dangers from its use. Related issues arise from the development of machine learning -

¹ KAY, John; KING, Mervin. *Radical Uncertainty. Decision-making beyond the numbers*. New York: W.W. Norton & Company, 2020, p. 208.

computers trained on historic data will develop algorithms reflecting past patterns of selection which may no longer be either appropriate or acceptable. Even if the explicit use of information such as gender or race is prohibited, the algorithms may have that consequence - without conscious intention for such discrimination on the part of anyone at all."

É importante ressaltar que o problema da discriminação estatística independe da qualidade dos dados e do acerto da metodologia, persistindo mesmo quando os resultados estatísticos refletem fielmente a realidade. Seja porque a individualidade da pessoa é desconsiderada, seja porque a estatística pode simplesmente retratar preconceitos que efetivamente ocorrem no mundo real, mas que não podem ser replicados de forma acrítica pelos modelos algorítmicos, é fundamental que existam análises complementares que possam contornar e suprir tais deficiências.

Outra grande dificuldade é que, diante do grande volume de dados e da potência cada vez maior dos sistemas algorítmicos, aumentam consideravelmente os riscos de correlações absurdas. É o que adverte Pedro Domingos, em seu famoso livro *The Master Algorithm*², no qual mostra os efeitos preocupantes dos *allucinating patterns*. Essa é mais uma razão para que sejam redobrados os cuidados com padrões e estatísticas que podem levar a correlações que não têm qualquer sentido e podem levar a conclusões totalmente equivocadas.

Se as estatísticas sobre o passado não resolvem, por si só, a solução de assuntos humanos complexos, o mesmo acontece com os cálculos probabilísticos que tentam antecipar o futuro. Esse é o argumento principal do livro de Kay e King³, ao mostrarem que tais metodologias são incompatíveis com as incertezas que caracterizam o mundo real:

"The appeal of probability theory is understandable. But we suspect the reason that such mathematics was, as we shall see, not developed until the seventeenth century is that few real-world problems can properly be represented

2 DOMINGOS, Pedro. *The Master Algorithm*. How the quest for the ultimate learning machine will remake our world. New York: Basic Books, 2015.

3 Op.cit., pp. 42-44.

in this way. The most compelling extension of probabilistic reasoning is to situations where the possible outcomes are well defined, the underlying processes which give rise to them change little over time, and there is a wealth history of information.

(...)

"Probabilistic reasoning may appear beautiful and appealing, but sadly its applicability to real-world problems is limited."

É por essas razões que os autores sustentam que a inferência estatística é – e precisa continuar sendo - uma opinião, até porque escolher as informações relevantes para uma decisão é uma importante forma de julgamento⁴. Daí o especial cuidado que se deve ter com as chamadas inferências causais⁵:

"Decisions about politics, finance and business should be made in the light of the best and most extensive data. But while data are essential, it is necessary to be careful in making inferences, and especially causal inferences, about the world based on data alone. The availability of what are now called 'big data' - the very large databases permitted by the power of modern computers - increases these dangers. The existence of a historic data set does not yield a basis for calculating a future probability distribution".

Kay e King⁶, na verdade, apresentam uma ampla crítica aos métodos matemáticos e quantitativos, como as análises de custo-benefício, mostrando várias de suas limitações, inclusive a de serem meras aparências de objetividade e neutralidade:

"We observe an obsession with cost-benefit analysis, or 'impact assessment' or even the 'business case' for major government decisions. Enthusiasm for 'evidence based policy' is seen as the hallmark of sophisticated decision-making. The problem is not so much that these models

4 Op.cit., p. 100.

5 Op.cit., pp. 389-390.

6 Op.cit., p. 404.

give rise to bad decisions but that they provide supposedly objective cover for bad decisions which have been made on quite different grounds."

Tal observação realça as críticas já feitas, ao longo da presente série, aos excessos do uso da matemática e dos métodos quantitativos. Daí a necessidade de que tais análises sejam complementadas por outros tipos de metodologias ou julgamentos, que privilegiem valores e aspectos qualitativos que foram desconsiderados pelas análises quantitativas, tal como igualmente já se demonstrou na série sobre Análises de Impacto Regulatório – AIRs⁷.

Ainda é preciso lembrar que o excesso de complexidade de vários sistemas de inteligência artificial, notadamente de *machine learning*, aumenta os riscos dos resultados não intencionais. Como apontam Michael Kearns e Aaron Roth⁸, a complexidade crescente dos modelos algorítmicos faz com que nem o executor nem o designer possam antecipar todos os seus resultados. Para Stuart Russel⁹, o problema dos resultados não intencionais é precisamente o mais grave das decisões algorítmicas:

"The debate has highlighted the conundrum we face: if we build machines to optimize objectives, the objectives we put into machines have to match what we want, but we don't know how to define human objectives completely and correctly."

Daí por que, além das limitações inerentes às metodologias estatísticas e probabilísticas, ainda é preciso levar em consideração os resultados não intencionais. Todos esses fatores mostram como a discussão ética e valorativa, longe de poder ser dispensada, é cada vez mais relevante para as decisões algorítmicas, a fim de assegurar que o propósito de acurácia não seja alcançado às custas de decisões injustas ou discriminatórias.

Como já se viu ao longo da série, para uma boa parte da literatura, critérios de justiça não podem ser simplesmente automatizados, o que já sugere

⁷ FRAZÃO, Ana. *Jota*. Perspectivas das Análises de Impacto Regulatório (AIRs) no Brasil.

. <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/perspectivas-das-analises-de-impacto-regulatorio-air-s-no-brasil-17022021>

⁸ KEARNS, Michael; ROTH, Aaron. *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*. New York: Oxford University Press, 2020, p. 69.

⁹ RUSSELL, Stuart. *Human Compatible. Artificial intelligence and the problem of control*. Viking, 2019, p. 70.

a necessidade de se complementar os julgamentos algorítmicos com julgamentos valorativos. Mesmo para autores que, como Kearns e Roth¹⁰, consideram ser possível expandir os princípios nos quais o *machine learning* é baseado, a fim de exigir que incorporem - de maneira quantitativa, mensurável e verificável - muitos dos valores éticos com os quais nos preocupamos como indivíduos e sociedade, entendem que o enfrentamento do problema exige necessariamente a intervenção humana. É o que se observa pela seguinte passagem, na qual os autores mostram que uma redução de acurácia pode ser justificável em prol dos ganhos na implementação da justiça¹¹:

"In other words, these kinds of more qualitative decisions and judgments - which type of fairness notion to use, when the reduction in accuracy is worth the gain of fairness, and many others - must remain firmly in the domain of human decision-making. The science part can begin only once society makes these difficult choices. Science can shed light on the pros and cons of different definitions, as we'll see, but it can't decide on right or wrong."

Se existe um tradeoff entre acurácia e justiça, Kearns e Roth¹² concluem que não há outra alternativa senão reconhecê-lo e enfrentá-lo:

"So there is simply no escaping that predictive accuracy and notions of fairness (and privacy, transparency, and many other social objectives) are simply different criteria, and that optimizing for one of them may force us to do worse than we could have on the other. This is a fact of life in machine learning. The only sensible response to this fact - from a scientific, regulatory, legal, or moral perspective - is to acknowledge it and to try to directly measure and manage the trade-offs between accuracy and fairness."

Vale ressaltar que os autores não apenas ressaltam a importância da decisão humana para a solução do conflito entre acurácia e justiça, como pontuam que tal solução envolve uma questão social e política – e não

10 Op.cit., p. 18.

11 Op.cit., p. 69.

12 Op.cit., p. 78.

propriamente científica – que deve ser informada por muitos fatores que não podem ser quantificados¹³:

"Making the trade-off between accuracy and fairness quantitative does not remove the importance of human judgment, policy, and ethics - it simply focuses them where they are most crucial and useful (...). Such decisions should be informed by many factors that cannot be made quantitative, including what the societal goal of protecting a particular group is and what is at stake."

"There are subjective, normative decisions that cannot (fruitfully) be made scientific (...)"

Dessa maneira, conclui-se que todos os esforços para quantificar e matematizar decisões complexas sobre seres humanos são insuficientes para abarcar a complexidade do problema e, mais ainda, para considerar os necessários valores e escolhas sociais e políticas que devem orientar esse tipo de julgamento.

O recurso às metodologias estatísticas e de cálculos probabilísticos, embora importante, não resolve parte importante dos assuntos humanos, seja em razão das limitações de tais abordagens, seja em razão dos riscos de resultados não intencionais, seja ainda em razão do fato de que, também nessa seara, deve haver o necessário enfrentamento das questões valorativas desde a programação do sistema algorítmico até a sua execução, sendo fundamental complementar os resultados com outros tipos de julgamento. Afinal, de que adianta assegurar acurácia ao preço de decisões algorítmicas injustas e discriminatórias?

Em qualquer caso, fica claro que a participação humana é fundamental nesse processo, como se continuará a examinar no próximo artigo, em que se explorará em que aspectos as decisões humanas, apesar de suas falhas e limitações, são superiores aos julgamentos algorítmicos e não podem ser simplesmente substituídas por eles.

¹³ Op.cit., p. 82 e p. 86, respectivamente.

Publicado em 25/08/2021

Link: <https://www.jota.info/opiniao-e-analise/artigos/discriminacao-algoritmica-superando-estatisticas-e-calculos-probabilisticos-25082021>