

A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais

Daniel de Araujo Dourado^{I,II} , Fernando Mussa Abujamra Aith^{I,II,III} 

^I Universidade de São Paulo. Centro de Pesquisa em Direito Sanitário. São Paulo, SP, Brasil

^{II} Universidade de São Paulo. Faculdade de Medicina. Programa de Pós-Graduação em Saúde Coletiva. São Paulo, SP, Brasil

^{III} Universidade de São Paulo. Faculdade de Saúde Pública. Departamento de Política, Gestão e Saúde. São Paulo, SP, Brasil

RESUMO

A inteligência artificial se desenvolve rapidamente e a saúde é uma das áreas em que as novas tecnologias desse campo são mais promissoras. O uso de inteligência artificial tem potencial para modificar a forma de prestação da assistência à saúde e do autocuidado, além de influenciar a organização dos sistemas de saúde. Por isso, a regulação da inteligência artificial na saúde é um tema emergente e essencial. Leis e normas específicas são elaboradas em todo o mundo. No Brasil, o marco inicial dessa regulação é a Lei Geral de Proteção de Dados Pessoais, a partir do reconhecimento do direito à explicação e à revisão de decisões automatizadas. É preciso debater a abrangência desse direito, considerando a necessária instrumentalização da transparência no uso da inteligência artificial na saúde e os limites atualmente existentes, como a dimensão caixa-preta inerente aos algoritmos e o *trade-off* existente entre explicabilidade e precisão dos sistemas automatizados.

DESCRITORES: Pesquisa sobre Serviços de Saúde. Inteligência Artificial, legislação & jurisprudência. Aprendizado de Máquina. Direito Sanitário.

Correspondência:

Daniel A. Dourado
Av. Dr. Arnaldo, 715
01246-904 São Paulo, SP, Brasil
E-mail: dadourado@usp.br

Recebido: 4 dez 2021

Aprovado: 4 fev 2022

Como citar: Dourado DA, Aith FMA. A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. Rev Saude Publica. 2022;56:80. <https://doi.org/10.11606/s1518-8787.2022056004461>

Copyright: Este é um artigo de acesso aberto distribuído sob os termos da Licença de Atribuição Creative Commons, que permite uso irrestrito, distribuição e reprodução em qualquer meio, desde que o autor e a fonte originais sejam creditados.



INTRODUÇÃO

A inteligência artificial (IA) começa a mudar o mundo como conhecemos e é uma das tecnologias mais promissoras para a área da saúde. O uso da IA na saúde, particularmente o subtipo *deep learning*¹, deve ter impacto significativo nos próximos anos na prática clínica, na gestão de sistemas de saúde e na relação entre os pacientes e a rede assistencial – ao permitir que eles processem seus próprios dados para promover a saúde². A Saúde Digital vai transformar a estrutura dos serviços de saúde e dos sistemas nacionais de saúde, com grande potencial para melhoria de qualidade e redução de custos na assistência³.

Por isso, a regulação da IA é hoje tema essencial no campo da saúde. Como toda intervenção que afeta a saúde, a incorporação dessas novas tecnologias precisa ser estimulada ao mesmo tempo em que se organiza uma estrutura regulatória capaz de assegurar que seu uso seja impreterivelmente em benefício dos seres humanos. Os sistemas de IA devem ter qualidade e segurança comprovadas e deve-se reconhecer que as ações e serviços, que sempre foram prestados principalmente por pessoas, começam a ser fortemente influenciados e até mesmo executados por sistemas automatizados. É um cenário que desafia pressupostos básicos da regulação em saúde⁴.

Até o início do ano de 2022, ainda não havia diretrizes ou leis específicas para regular o uso da inteligência artificial na assistência à saúde⁵. Esse debate está em andamento em muitos países e entidades internacionais e chega ao Brasil com a Lei Geral de Proteção de Dados Pessoais (LGPD – Lei Federal nº 13.709/2018), passando a vigorar e instituindo o direito à explicação e à revisão de decisões automatizadas no ordenamento jurídico brasileiro. Trata-se da expressão normativa do princípio da transparência de algoritmos, que é ponto central na regulação dos sistemas de IA.

Este artigo tem por objetivo discutir a abrangência do direito à explicação e à revisão de decisões automatizadas na regulação da IA na saúde no Brasil a partir do marco da LGPD, considerando o debate internacional sobre o tema os limites atualmente existentes para IA explicável na área da saúde.

Princípios para Regulação da Inteligência Artificial na Saúde

É importante diferenciar a IA tradicional – usada desde a década de 1950 – das técnicas mais recentes de *machine learning* e *deep learning*, porque são elas que representam o grande desafio regulatório. *Machine learning* é um tipo de IA que permite aos computadores aprender automaticamente, sem programação explícita. *Deep learning* é um subtipo de *machine learning* que consiste em uma classe de algoritmos que usam modelos inspirados no sistema nervoso central de organismos vivos, denominados redes neurais artificiais.

Os algoritmos de *deep learning* podem aprender relações extremamente complexas para reconhecimento de padrões e, por isso, podem ajudar a fazer previsões clinicamente relevantes a partir de dados complexos e heterogêneos, gerados em atendimento clínico, como registros médicos, imagens clínicas, dados de monitoramento contínuo de sensores e dados genômicos⁶. Diferentemente dos objetos da regulação tradicional, como medicamentos e dispositivos médicos, esses algoritmos “autodidatas” estão em constante mudança. A capacidade desses sistemas de aprender com a experiência do mundo real (treinamento) e melhorar continuamente o desempenho (adaptação) torna essas tecnologias únicas. Regulá-los é como atingir um alvo em contínuo movimento.

A regulação de algoritmos tem se tornado uma preocupação relevante nos sistemas jurídicos em todo o mundo e, atualmente, está em construção em diferentes países e órgãos internacionais. Os instrumentos normativos vinculantes (leis) são escassos e têm se concentrado principalmente no campo da privacidade dos dados. Nos demais aspectos, a abordagem regulatória começa a se estruturar por meio de códigos de conduta e diretrizes não vinculantes (*soft law*), produzidos por entidades governamentais, conselhos de especialistas para assessoramento de entes públicos, institutos de pesquisa e empresas privadas⁷.

A Organização Mundial da Saúde tem buscado convergência para orientar os governos e as demais entidades internacionais no uso da IA na saúde. Nesse sentido, a partir de princípios éticos gerais para desenvolvimento de IA e da incorporação de elementos da bioética e da atual regulação em saúde, identificam-se seis princípios-chave para a regulação dos sistemas de IA na saúde: 1) autonomia; 2) não-maleficência/beneficência; 3) transparência; 4) responsabilidade; 5) equidade; 6) responsividade/sustentabilidade⁵. Esses princípios estão interligados, sem hierarquia entre eles, e precisam ser instrumentalizados conjuntamente.

Transparência Algorítmica e Direito à Explicação

A transparência é o princípio ético mais frequentemente encontrado nos códigos de diretrizes gerais para o uso da IA⁷ e é um princípio-chave para a IA na saúde. Transparência significa que informações suficientes sobre as tecnologias de IA sejam documentadas antes da implantação de modo a facilitar consulta pública e entendimento sobre seu funcionamento no mundo real. Espera-se que os sistemas sejam inteligíveis e explicáveis para desenvolvedores, profissionais de saúde, pacientes, usuários e reguladores, de acordo com a capacidade de compreensão de cada grupo e mesmo de cada indivíduo.

Instrumentalizar a transparência dos algoritmos é necessário para que outros princípios-chave do uso da IA na saúde tenham eficácia⁸: a proteção da autonomia humana, para assegurar que as pessoas permaneçam no controle dos sistemas de saúde e das decisões médicas; os requisitos regulatórios de segurança e eficácia, que garantam que a IA não prejudique as pessoas e promova o bem-estar; a prestação de contas (*accountability*) no uso de tecnologias de IA⁹; a busca de equidade, que promova inclusão social e assegure que algoritmos não reproduzam qualquer tipo de preconceito e discriminação. A expressão de todos esses princípios pressupõe transparência dos sistemas de IA.

Atualmente, o principal mecanismo para expressão da transparência algorítmica tem sido o direito à explicação (*right to explanation*) sobre decisões automatizadas, considerado elemento fundamental na regulação de algoritmos. Esse conceito vem se consolidando desde a elaboração do Regulamento Geral sobre a Proteção de Dados da União Europeia (mais comumente identificado pela sigla GDPR, do inglês *General Data Protection Regulation*), em vigor desde maio de 2018. O titular dos dados deve ter “o direito de obter a intervenção humana, de manifestar o seu ponto de vista, de obter uma explicação sobre a decisão tomada na sequência dessa avaliação e de contestar a decisão”. Ou seja, além de receber uma explicação inteligível, cria-se o direito à oportunidade de ser ouvido, de questionar e pedir revisão da decisão automatizada. É o que vem sendo chamado de “devido processo algorítmico”¹⁰.

Em suma, o direito à explicação diz respeito ao reconhecimento de que deve ser assegurado a todas as pessoas o direito de saber como são tomadas as decisões baseadas em IA que afetam suas vidas. Desde a publicação do GDPR, mesmo antes de sua entrada em vigor, há intensas discussões sobre a existência e o alcance do direito à explicação nas decisões automatizadas. Como um algoritmo de IA pode usar inúmeras variáveis para chegar a um determinado resultado, a forma complexa de representação matemática é, na maior parte das vezes, ininteligível para os humanos – por isso os algoritmos são comumente chamados de sistemas de “caixa-preta”¹¹.

Em linhas gerais, o debate atualmente se divide em duas interpretações. Por um lado, os que defendem a viabilidade e o escopo do direito à explicação apenas no que diz respeito à funcionalidade geral do sistema, ao invés de sobre decisões específicas e circunstâncias individuais¹². Por outro lado, há o entendimento de que a explicação também deve incluir decisões específicas, com transparência limitada apenas pela dimensão intrinsecamente de caixa-preta dos algoritmos¹³.

A importância do direito à explicação é dar aos pacientes a possibilidade de compreender a lógica de decisões automatizadas que impactem condutas tomadas nos seus cuidados

de saúde. Tal preocupação deve estar cada vez mais presente em diversas situações clínicas. Por exemplo, atualmente já existem algoritmos de *deep learning* capazes de definir critérios para transplantes de órgãos, como alocação, correspondência entre doador e receptor e de previsão de sobrevida dos pacientes transplantados¹⁴. É bem possível que em breve esses algoritmos possam ser usados para essa finalidade e que haja diferenças de ordem nas filas de transplantes em comparação com as definidas por critérios clínicos feitas apenas por humanos. O direito à explicação relaciona-se à dignidade humana e não será aceitável que decisões dessa natureza sejam tomadas apenas com base em sistemas do tipo caixa-preta.

Direito à Explicação a Partir da Lei Geral de Proteção de Dados Pessoais

No Brasil, a LGPD foi promulgada em agosto de 2018 e entrou em vigor em setembro de 2020 (os dispositivos referentes às sanções administrativas só entraram em vigor em agosto de 2021). Fazendo jus a seu nome, a LGPD é essencialmente dedicada à privacidade de dados pessoais e não trata especificamente da regulação de IA – os termos “inteligência artificial” e “algoritmo” sequer aparecem no texto. Mas como foi abertamente inspirada no GDPR europeu e incorporou muito da sua racionalidade, a lei introduz no direito brasileiro o direito à explicação e à revisão de decisões automatizadas.

O direito à revisão de decisões automatizadas está definido explicitamente no texto do art. 20 da LGPD, que concede ao titular o direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, a exemplo do GDPR. Diferentemente da legislação europeia, a lei brasileira não prevê o direito de não se sujeitar a decisão exclusivamente automatizada e nem de obter intervenção humana em caso de revisão. A redação original aprovada pelo Congresso Nacional previa o direito do titular de solicitar revisão de decisões automatizadas “por pessoa natural”, mas o dispositivo foi alterado por medida provisória, posteriormente convertida em lei. Vale ressaltar que, embora a obrigação de supervisão humana tenha sido excluída da LGPD, não há vedação para que essa exigência seja feita na regulamentação infralegal.

O direito à explicação não aparece textualmente (assim como no GDPR), mas decorre da interpretação sistemática da própria LGPD em conjunto com dispositivos constitucionais e da legislação de proteção ao consumidor¹⁵. A lei brasileira garante a todo aquele afetado por decisões automatizadas o direito a obter informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados. Isso é a expressão do princípio da transparência, que só pode ser garantido por meio da explicação.

A LGPD ressalva os segredos comercial e industrial nesse e em diversos outros dispositivos, de modo que essa ponderação deverá ser feita na regulamentação infralegal e mesmo na análise de casos concretos. A proteção dos segredos de negócio na LGPD pode ser vista como uma forma de fomentar o modelo empresarial baseado em algoritmos, mas deve necessariamente ser sopesada com o direito à explicação sobre decisões automatizadas a fim de observar os princípios éticos de uso da IA em harmonia com os direitos humanos. A própria lei sinaliza nesse sentido ao prever a realização de auditoria em caso de suspeita de discriminação.

Os direitos à explicação e à revisão de decisões de sistemas de IA estão necessariamente vinculados e precisam ser compreendidos conjuntamente. Assim como deve acontecer em outros países, com base no modelo europeu, a configuração desses direitos no Brasil ainda necessita de regulamentação e de futura elaboração doutrinária e jurisprudencial.

O Desafio da Inteligência Artificial Explicável na Saúde

O direito à explicação está vinculado aos limites da transparência algorítmica. A transparência de um sistema de IA se concentra principalmente no processo, ou seja, trata-se de permitir que as pessoas entendam como os algoritmos são desenvolvidos e

implantados em termos gerais. Eventualmente pode incluir elementos sobre fatores de uma previsão ou decisão específica, mas em regra não inclui o compartilhamento de códigos ou conjuntos de dados.

Logo, a existência de alguma opacidade é inevitável. Essa opacidade tem relação com a dimensão de caixa-preta, pela complexidade dos sistemas, mas não só. Há também opacidade imposta por sigilo corporativo ou de Estado (intencional), pois o compartilhamento de códigos ou conjuntos de dados específicos pode revelar segredos comerciais ou divulgar dados confidenciais do usuário. E opacidade decorrente do próprio “analfabetismo técnico” dos usuários¹⁶.

Nesse sentido, não é viável nem necessário que uma explicação forneça todo o processo de tomada de decisão do sistema. A explicação é essencial para situações em que alguma falha precisa ser determinada em uma instância específica do sistema, sobretudo quando algoritmos começam a ser cada vez mais usados para fazer recomendações ou decisões atualmente sujeitas à discricção humana. Mas é suficiente que a explicação consiga responder a um dos seguintes pontos¹⁷: 1) principais fatores de decisão: indicar os fatores importantes para uma previsão de IA, preferencialmente ordenados por significância; 2) fatores determinantes de decisão: esclarecer fatores que afetam decisivamente o resultado; 3) resultados divergentes: explicar por que dois casos de aparência semelhante podem resultar em resultados diferentes¹⁸.

O campo de IA explicável está em acelerada expansão. Atualmente há muitas pesquisas técnicas em desenvolvimento por empresas, órgãos de normalização, organizações sem fins lucrativos e instituições públicas com intuito de criar sistemas de IA que possam explicar suas previsões. Projetar um sistema para fornecer explicações é complexo e caro, tanto se forem desenvolvidos prevendo a possibilidade de fornecer um certo tipo de explicação (“explicabilidade inerente”) quanto principalmente se explicações forem buscadas após a decisão algorítmica (“explicabilidade *post hoc*”)¹⁹. Por isso, as pesquisas têm sido impulsionadas sobretudo na busca de modelos explicáveis para áreas de alto risco, como é o caso do uso na assistência à saúde²⁰.

No entanto, os limites para a IA explicável na saúde são bastante relevantes.

Primeiramente, é preciso considerar que existe um *trade-off* entre explicabilidade e precisão²¹. Para que um sistema de IA seja explicável, na maioria das vezes é necessária uma redução das variáveis da solução a um conjunto pequeno o suficiente para que fique acessível ao entendimento humano. Isso pode inviabilizar o uso de alguns sistemas em problemas complexos. Alguns modelos de *deep learning* podem prever probabilidades de diagnósticos clínicos com precisão, mas serem humanamente incompreensíveis. Nesse sentido, um direito à explicação mais amplo, baseado na máxima transparência, pode ser incompatível com o uso de sistemas automatizados que busquem alta acurácia preditiva.

Além disso, as técnicas atualmente disponíveis para explicabilidade são capazes de oferecer descrições amplas de como o sistema de IA funciona em sentido geral, mas são muito superficiais ou não confiáveis para decisões individuais²². Na prática, as explicações podem ser muito úteis em processos globais de IA, como desenvolvimento de modelo e auditoria, mas raramente são informativas com relação a resultados específicos dados pelos algoritmos.

Portanto, a atual falta de transparência deve persistir, ao menos por algum tempo. Em certa medida, opacidade é uma característica usual na atividade clínica. A medicina tradicionalmente adota práticas que envolvem mecanismos não totalmente compreendidos, mas que continuam sendo amplamente usados, devido aos efeitos comprovados, como é o caso de muitos medicamentos. Os obstáculos para desenvolver uma IA explicável na saúde devem ser reconhecidos e bem ponderados na construção de mecanismos regulatórios que considerem os limites da explicabilidade e conseqüentemente a abrangência do direito à explicação e à revisão de decisões automatizadas na assistência à saúde.

CONCLUSÃO

A aplicação do direito à explicação na saúde deverá contemplar complexidades específicas da regulação da IA para uso clínico. Partindo do entendimento de que esse direito está agora presente na legislação brasileira, caberá aos órgãos reguladores delimitar a sua amplitude e mecanismos para que possa ser instrumentalizado. Além da atuação da Autoridade Nacional de Proteção de Dados, deve haver intervenção de outros órgãos reguladores, como a Agência Nacional de Vigilância Sanitária e das autoridades reguladoras de profissões regulamentadas, como os conselhos de medicina.

O exercício do direito à explicação na saúde depende da criação de mecanismos para criação de sistemas de inteligência artificial explicáveis e do reconhecimento dos limites da explicabilidade de algoritmos. A abrangência desse direito deve ser definida a partir de critérios a serem elaborados por autoridades reguladoras e precisam ser amplamente discutidos com a sociedade. Esse debate está apenas começando.

REFERÊNCIAS

1. Obermeyer Z, Emanuel EJ. Predicting the future: big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-9. <https://doi.org/10.1056/NEJMp1606181>
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56. <https://doi.org/10.1038/s41591-018-0300-7>
3. World Health Organization. mHealth: use of appropriate digital technologies for public health: report by the director-general. In: 71. WHO Assembly; 2018 Mar 26; Geneva, Switzerland. Geneva (CH): WHO; 2018 [citado 1 fev 2022]. Disponível em: https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_20-en.pdf
4. Richman B. Health regulation for the digital age: correcting the mismatch. *N Engl J Med*. 2018;379(18):1694-5. <https://doi.org/10.1056/NEJMp1806848>
5. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva (CH): WHO; 2021 [citado 1 fev 2022]. Disponível em: <https://www.who.int/publications/i/item/9789240029200>
6. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-58. <https://doi.org/10.1056/NEJMra1814259>
7. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1:389-99. <https://doi.org/10.1038/s42256-019-0088-2>
8. Watson DS, Krutzina J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ*. 2019;364:l886. <https://doi.org/10.1136/bmj.l886>
9. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med*. 2018;15(11):e1002689. <https://doi.org/10.1371/journal.pmed.1002689>
10. Kaminski ME. The right to explanation, explained. *Berkeley Technol Law J*. 2019;34(1):189-218. <https://doi.org/10.15779/Z38TD9N83H>
11. Price WN II. Medical malpractice and black-box medicine. In: Cohen IG, Lynch HF, Vayena E, Gasser U, editors. *Big data, health law, and Bioethics*. Cambridge (UK): Cambridge University Press; 2018. Chapter 20; p. 295-306. <https://doi.org/10.1017/9781108147972.027>
12. Wachter S, Mittelstadt B, Floridi L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *Int Data Priv Law*. 2017;7(2):76-99. <https://doi.org/10.1093/idpl/ix005>
13. Selbst AD, Powles J. Meaningful information and the right to explanation. *Int Data Priv Law*. 2017;7(4):233-42. <https://doi.org/10.1093/idpl/ix022>
14. Khorsandi SE, Hardgrave HJ, Osborn T, Klutts G, Nigh J, Spencer-Cole RT, et al. Artificial intelligence in liver transplantation. *Transplant Proc*. 2021;53(10):2939-44. <https://doi.org/10.1016/j.transproceed.2021.09.045>

15. Monteiro R, Machado C, Silva L. The right to explanation in Brazilian data protection law. RIDDN [Internet]. 2021 [citado 01 fev 2022];7(1):119-136. Disponível em: <https://ojs.imodev.org/?journal=RIDDN&page=article&op=view&path%5B%5D=406>
16. Burrell J. How the machine ‘thinks’: understanding opacity in machine learning algorithms. Big Data Soc. 2016;3(1):1-12. <https://doi.org/10.1177/2053951715622512>
17. Organisation for Economic Co-Operation and Development. Artificial Intelligence in society. Paris (FR): Éditions OCDE; 2019.
18. Doshi-Velez F, Kortz M, Budish R, Bavitz C, Gershman S, O’Brien D, et al. Accountability of AI under the law: the role of explanation. arXiv:1711.01134v3 [Preprint]. 2017 [citado 3 nov 2017; revisado 20 dez 2019]. Disponível em: <https://arxiv.org/abs/1711.01134v3>
19. Lipton ZC. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue. 2018;16(3):31-57. <https://doi.org/10.1145/3236386.3241340>
20. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv:1712.09923v1 [Preprint]. 2017 [citado 28 dez 2017]. Disponível em: <https://arxiv.org/abs/1712.09923v1>
21. London AJ. Artificial Intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep. 2019;49(1):15-21. <https://doi.org/10.1002/hast.973>
22. Ghassemi M, Oakden-Rayner Luke, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-50. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9)

Contribuição dos Autores: Concepção e planejamento do estudo: DAD, FMAA. Coleta, análise e interpretação dos dados: DAD. Elaboração ou revisão do manuscrito: DAD, FMAA. Aprovação da versão final: DAD, FMAA. Responsabilidade pública pelo conteúdo do artigo: DAD, FMAA.

Conflito de Interesses: Os autores declaram não haver conflito de interesses.