

# Trustworthy AI for safe medicines



The pharmaceutical industry is rapidly adopting new and evolving applications of artificial intelligence (AI), and so we concur with the point made by Hines et al. (*Nat. Rev. Drug Discov.* **22**, 81–82; 2023)<sup>1</sup> that there is a need for regulatory agencies and industry to collaborate towards establishing a safety framework for this transition. However, the specifics of such a framework are yet to be defined. Here, we present our view of critical features of a potential regulatory framework for pharmaceutical applications of AI that improve understanding of the benefit–risk ratio of a medicinal product.

We use pharmacovigilance as an example to illustrate our view, as it is a field that is both well established and well regulated, but also one with matured usage of routinely collected healthcare data<sup>2</sup> but only emerging experience with AI applications. Our view overlaps with the view of Hines et al.<sup>1</sup>, particularly in our shared emphasis on a risk-based regulatory framework that implements proportionate precautionary measures to enable responsible innovation. We consider, however, that an industry-wide risk-based framework must grade requirements according to risk level and harmonize with existing pharmaceutical regulation, and should not require regulatory access to the underlying algorithms and datasets, particularly considering more effective alternatives.

## Developing a risk-based framework for AI

Grading requirements according to risk level. We concur with Hines et al.<sup>1</sup> that regulatory requirements should reflect the risk level of pharmaceutical AI. Although the authors focus on AI applications affecting the risk–benefit ratio of medicines, we note here the broader importance of establishing consistent enterprise-wide categorization based on the use context. Whereas high-risk applications should be subject to rigorous quality management standardized by industry regulators, low-risk applications in research and discovery that do not affect patient rights or interests only need to conform to [good machine learning practice](#)<sup>2</sup>. In either case, AI applications should not increase overall risk relative to relevant human benchmarks. This approach

will avoid unnecessary compliance costs, frequently cited as inhibiting innovation<sup>3</sup>, and encourage investment in the growing AI-based drug discovery ecosystem.

**Harmonization with existing pharmaceutical regulation.** In establishing a regulatory framework as Hines et al. propose<sup>1</sup>, we note that sectoral harmonization is critically important: we argue that it is imperative that the regulation of pharmaceutical AI is subsidiary to the existing pharmaceutical regulatory infrastructure, which already has sophisticated systems of control and validation in place. By contrast, a horizontal approach that imposes uniform regulations on AI applications across fields risks imposing inapt regulations on the specific needs of the pharmaceutical industry, as well as creating regulatory conflicts and uncertainty.

Instead, the industry requires clarity and regulatory harmony on the benefit–risk criteria of using AI in pharmacovigilance, particularly where AI is furthering understanding of the benefit–risk profile of a medicinal product. Regulatory mechanisms must also be suitably agile to keep pace with the evolution of pharmaceutical AI<sup>4,5</sup>. We look forward to this question being further pursued through international collaborative working groups such as the Council for International Organizations of Medical Sciences (CIOMS) Working Group on AI in Pharmacovigilance.

**Assuring safe and trusted use.** We propose that safe process oversight should focus on the outcomes of validation measures, which is an established practice across the drug development lifecycle. Whilst we recognize the regulatory authorities' obligation to ensure safety and transparency, requiring regulatory access to algorithms and datasets as proposed by Hines et al.<sup>1</sup> is neither necessary nor sufficient for this purpose. Furthermore, such requirements could inhibit innovation.

We believe the key motivation for requiring 'white box' access is to render a process explainable or interpretable<sup>6</sup>. However, there is a reason this is not the focus of current pharmaceutical regulation: although we cannot predict everything about a medicine's mechanism of action in a real-world scenario,

we can nevertheless ensure its safety and effectiveness through empirical testing and monitoring. We can trust medicines, therefore, because we trust the rigorous process that validates them. We assert that the same principle should apply to applications of AI in the pharmaceutical industry: regulatory scrutiny should focus on validating and monitoring the outcomes of a process for safety, reliability and effectiveness. Validation (as described above) is not particularly helped by access to algorithms or datasets<sup>7</sup>, which would not necessarily indicate how safe a pharmaceutical application of AI is.

Safety validation can adequately rely upon 'black box' assessment, which does not require access to algorithms or data<sup>6</sup>, complemented with methodological transparency, to guarantee medicine safety without adversely affecting innovation. Focusing on methodological transparency and outcome validation has important advantages: it improves the reproducibility of findings (per an assessment of 150 real-world data cases<sup>8</sup>), and it enables developers to use training datasets that evolve over time without having to store all data centrally or to lock their algorithms whilst data accrues<sup>4,9</sup>. Regulators might still need some insight into datasets to ensure that they are methodologically sound, but this is possible through transparency tools such as datasheets and summaries<sup>10</sup>.

Furthermore, the direct focus on access to algorithms and data by Hines et al.<sup>1</sup> seems anachronistic in view of a desired move towards a quality-management approach that will facilitate risk-based innovation, maximizing appropriate use of data, wherever it resides. It also conflicts with the current regulatory use of decentralized data networks, such as the [FDA's Sentinel Initiative](#) and the [DARWIN EU platform](#), which represent unique and potentially rich resources for advancing pharmacovigilance through AI-based technologies. An insistence on white box inspection would require centralizing data and therefore rule out these valuable resources for pharmacovigilance. Insofar as Hines et al.<sup>1</sup> represent the views of the European Medicines Agency (EMA), it would be in the interest of the pharmacovigilance regulatory ecosystem for the EMA to engage actively with industry

(as well as representative bodies, like the CIOMS) to achieve the kind of harmonized, pro-innovation framework described above.

## Conclusion

In a recent survey, global pharmaceutical companies cited regulatory and compliance concerns as one of the top reasons for not implementing AI<sup>3</sup>. Although we are in broad agreement with Hines et al.<sup>1</sup> about the need for a coherent and risk-based system of regulation, we suggest refinements to the specificities of such a system, with a view to unlocking the potential of AI and enabling proportionate and context-appropriate protections for patient safety.

## Note added in proof

The authors acknowledge and welcome the EMA's decision to join the ongoing CIOMS working group on AI, which occurred after acceptance of this publication, and look forward to active discussions in this forum in the future. The need for international alignment

for a coherent, risk-based regulatory framework remains a priority, and we encourage the EMA to engage with the pharmaceutical industry to ensure a harmonized approach moving forward.

**Jens-Ulrich Stegmann<sup>1</sup>, Rory Littlebury<sup>2</sup>, Markus Trengove<sup>3</sup>, Lea Goetz<sup>3</sup>, Andrew Bate<sup>4</sup> & Kim M. Branson<sup>5</sup>**

<sup>1</sup>Global Safety, GSK, Wavre, Belgium. <sup>2</sup>Global Safety, GSK, Stevenage, UK. <sup>3</sup>Artificial Intelligence and Machine Learning, GSK, London, UK. <sup>4</sup>Global Safety, GSK, Brentford, UK. <sup>5</sup>Artificial Intelligence and Machine Learning, GSK, San Francisco, USA.

✉ e-mail: [andrew.x.bate@gsk.com](mailto:andrew.x.bate@gsk.com)

Published online: 7 August 2023

## References

1. Hines, P. A., Herold, R., Pinheiro, L., Frias, Z. & Arlett, P. Artificial intelligence in European medicines regulation. *Nat. Rev. Drug Discov.* **22**, 81–82 (2023).
2. Bate, A. & Luo, Y. Artificial intelligence and machine learning for safe medicines. *Drug Saf.* **45**, 403–405 (2022).
3. Kassekert, R. et al. Industry perspective on artificial intelligence/machine learning in pharmacovigilance. *Drug Saf.* **45**, 439–448 (2022).
4. Bates, D. W. How to regulate evolving AI health algorithms. *Nat. Med.* **29**, 26 (2023).
5. Gerke, S., Babic, B., Evgeniou, T. & Cohen, I. G. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *npj Digit. Med.* **3**, 53 (2020).
6. Liu, X. et al. The medical algorithmic audit. *Lancet Digit. Health* **4**, e384–e397 (2022).
7. Beam, A. L., Manrai, A. K. & Ghassemi, M. Challenges to the reproducibility of machine learning models in health care. *JAMA* **323**, 305–306 (2020).
8. Wang, S. V., Sreedhara, S. K. & Schneeweiss, S., REPEAT Initiative. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. *Nat. Commun.* **13**, 5126 (2022).
9. Platt, R. et al. The FDA Sentinel Initiative – an evolving national resource. *N. Engl. J. Med.* **379**, 2091–2093 (2018).
10. Gebru, T. et al. Datasheets for datasets. *Commun. ACM* **64**, 86–92 (2021).

## Competing interests

All authors are employees of GSK and hold GSK stock and stock options.

## Additional information

**Peer review information** *Nature Reviews Drug Discovery* thanks David Bates and Rajesh Ghosh for their contribution to the peer review of this work.