



ARTÍCULO DE REFLEXIÓN / REFLECTIVE PAPER

<https://dx.doi.org/10.14482/inde.44.01.813.443>

# Transforming Drug Development with Artificial Intelligence: Reflections on Applications in Safety and Clinical Planning

*Transformando el desarrollo de fármacos con inteligencia artificial: reflexiones sobre aplicaciones en seguridad y planificación clínica*

M A R I A C A M I L A M A R E N C O \*

\* Master of Business Analytics, Massachusetts Institute of Technology. Independent author.  
Orcid-ID: <https://orcid.org/0009-0003-8094-7615>. [mariacamilamarenco@gmail.com](mailto:mariacamilamarenco@gmail.com)

**Correspondence:** 88 Ames Street, apt 1407, Cambridge, MA, 02142. Phone number: +17876385050



## Abstract

Pharmaceutical development faces mounting pressure from escalating costs, lengthy timelines, and stringent regulatory oversight. Artificial intelligence (AI) has emerged as a potential catalyst for redesigning this end-to-end process. This reflective article critically reviews the current state and future trajectory of AI in the pharmaceutical industry, illustrated by two projects led by the author: (i) an AI-empowered pharmacovigilance framework that forecasts adverse events before clinical manifestation and (ii) a dynamic optimization algorithm that redefines the strategy for selecting and activating clinical sites. Methodological challenges (data quality, interpretability, bias) and lessons learned for scalable adoption are discussed, as well as requirements for governance and regulatory collaboration. Evidence suggests that AI can reduce costs, compress timelines, and enhance patient safety; however, its ultimate value will depend on rigorous principles of transparency, validation, and ethical oversight.

**Keywords:** advanced analytics, artificial intelligence, clinical site allocation, optimization, pharmacovigilance, patient safety.

## Resumen

El desarrollo farmacéutico se enfrenta a presiones crecientes de costo, tiempo y complejidad regulatoria. La inteligencia artificial (IA) surge como catalizador potencial para reconfigurar este proceso a lo largo de todo su ciclo de vida. Este artículo de reflexión examina críticamente el estado actual y la trayectoria futura de la IA en la industria farmacéutica, ilustrado mediante dos casos que la autora lideró: (i) un marco de farmacovigilancia que predice eventos adversos antes de su manifestación clínica y (ii) un algoritmo de optimización dinámica que redefine la estrategia de selección y activación de sitios clínicos. Se discuten los retos metodológicos (calidad de datos, interpretabilidad, sesgo) y las lecciones aprendidas para la adopción escalable, así como los requisitos de gobernanza y colaboración regulatoria. La evidencia demuestra que la IA puede reducir costos, acelerar calendarios y mejorar la seguridad del paciente, pero su valor dependerá de principios rigurosos de transparencia, validación y supervisión ética.

**Palabras clave:** analítica avanzada, asignación de sitios clínicos, farmacovigilancia, Inteligencia artificial, optimización, seguridad del paciente.

## INTRODUCTION

The drug-development ecosystem is under unprecedented strain. A single investigational compound now costs an estimated USD 2 billion and 10–15 years to shepherd from discovery to market approval [1]. Yet approximately 90% of molecules that enter human trials ultimately fail to secure regulatory clearance [2]. Amid this attrition, sponsors are expected to uphold ever-higher standards of patient safety, data transparency, and operational efficiency. Over the past decade, artificial intelligence (AI)—an umbrella term encompassing machine learning, natural language processing (NLP), and knowledge graph techniques—has matured to the point where tangible impact is demonstrable across the pharmaceutical value chain.

## OVERVIEW OF AI APPLICATIONS IN DRUG DEVELOPMENT

AI applications now permeate every major phase of pharmaceutical R&D:

- **Compound discovery:** Deep-learning architectures, particularly graph neural networks, explore vast chemical spaces and propose synthetically accessible leads within weeks rather than years [3].
- **Protocol design:** Artificial intelligence and natural-language-processing methods are increasingly applied to historical study documents to refine eligibility criteria, visit schedules, and endpoint hierarchies, thereby enabling more patient-centric clinical-trial protocols [4].
- **Site selection and recruitment:** Predictive analytics leverage multi-sourced data (historic performance, demographic catchment, investigator fingerprints) to rank sites by expected enrollment yield [5].
- **Clinical trial enablement:** AI frameworks support patient recruitment through mining electronic health records, enable simulation of interventions, and facilitate remote participation via digital biomarkers and therapeutics [6].
- **Pharmacovigilance:** NLP and signal-detection algorithms continuously integrate electronic health records, spontaneous-report databases, and social media narratives to flag emerging safety concerns [3].

FDA recognizes the increased use of AI throughout the drug product life cycle and across several therapeutic areas, underscoring the importance of aligning innovation with evolving regulatory expectations [7].

## REFLECTIVE CASE STUDIES

### AI-Enabled Pharmacovigilance for Predicting Adverse Events

The first project, executed through a Massachusetts Institute of Technology (MIT) capstone in partnership with a global biopharmaceutical company, sought to anticipate adverse drug reactions (ADRs) before they manifest in clinical or post-marketing settings. A modular pipeline incorporated four complementary analytics models to act at three different stages of the patient-drug interaction process [8]:

#### *Model 1 – Early Development: Continue or Discontinue the Drug?*

An interpretable machine learning decision tree was employed to predict the likelihood of adverse events based on demographic data. This model supports early go/no-go development decisions by identifying patient populations at higher risk for specific ADRs.

#### *Model 2 – Causal Inference: Is There a Link Between the Drug and the Adverse Event?*

This model used knowledge graph techniques to analyze the influence of concomitant medications and assess the causal relationship between drug initiation and adverse events. It enables disentangling the effect of the investigational drug from other co-administered therapies.

#### *Model 3 – Pre-Commercialization: Should We Market This Drug?*

A second interpretable machine learning model incorporated both demographic and dosage information to improve the prediction of dosage-dependent ADRs. This model informs benefit–risk assessments at the pre-launch stage.

#### *Model 4 – Post-Marketing Surveillance: How Do We Improve the Drug Label?*

This model applied regression-based methods and causal comparison techniques to detect early safety signals with potential implications for drug labeling. It supports ongoing pharmacovigilance by identifying label update needs based on emerging real-world evidence.

Key challenges included extreme class imbalance, heterogeneity of unstructured narratives, and the need for clinician interpretability. Strategies such as adaptive

under-sampling, SHAP-value explanation, and ontology-anchored entity recognition mitigated these barriers.

To assess model performance, Receiver Operating Characteristic (ROC) curves were employed, providing a comprehensive view of classification performance across probability thresholds (0 to 1). The Area Under the Curve (AUC) was calculated as a quantitative measure, with the ROC curve plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis). The models' performances ranged from 0.80, with the best-performing model achieving an AUC of 0.95, indicating strong predictive capability. This top-performing model, an interpretable classification tree technique (OCT), provided visibility into the key contributing predictive factors. In addition, the models quantified a 12–30% incremental risk attributable to specific co-medications. One of these therapies impacts more than 31 million individuals annually across over 100 countries, underscoring the models' potential for substantial real-world impact.

### *AI-Driven Optimization of Clinical-Site Allocation*

The second initiative addressed the chronic inefficiency of activating underperforming investigative sites. Industry data indicate that a typical Phase II/III study costs USD 86 million, while sub-optimal site selection, among other factors, can prolong timelines by up to six months [9]. A four-tier solution was designed:

- 1) Binary “never-enroll” model (XGBoost) flagging sites unlikely to randomize any subject.
- 2) Multiclass enrollment-rate classifier forecasting low, medium, or high recruitment strata.
- 3) Accelerated failure-time (AFT) log-logistic model predicting the date at which each site will cease enrollment.
- 4) Dynamic optimization engine solving a mixed-integer programming formulation that maximizes expected enrollment under budgetary constraints and performance-risk penalties. The formulation incorporates a piecewise-linear approximation of the sigmoid function, enabling the optimization to capture non-linear relationships.

The classification models achieved AUC values ranging from 0.70 to approximately 0.93 in the best-performing model. Analysis of these models identified a subset of the most impactful site and study characteristics influencing enrollment. When incorporated into study-planning workflows, these insights have the potential to save up to USD 200 million over five years by avoiding the activation of underperforming sites.

## CONCLUSIONS

Artificial intelligence methods are already reshaping pharmaceutical R&D by shortening development timelines, strengthening pharmacovigilance, and enabling data-driven operational planning across the product life cycle. The FDA has proposed a comprehensive, risk-based credibility assessment framework to evaluate the trustworthiness of AI model outputs in specific drug development contexts, emphasizing the importance of early engagement with the Agency to align on model design and documentation [10]. Additionally, CDER has reported reviewing over 500 submissions involving AI/ML components between 2016 and 2023, highlighting the rapid growth of real-world AI integration across therapeutic and operational domains [11].

AI in drug development is no longer a distant vision—it is a present-day reality. However, achieving durable impact will require: (i) transparent and reproducible algorithm design, (ii) alignment with evolving regulatory guidance, and (iii) structured cross-sector collaboration. Strategic investment in these pillars will enable the reliable integration and scalable deployment of AI, advancing a smarter and safer paradigm for drug discovery and development.

## DISCLAIMER

The author is an employee and shareholder of Takeda Development Center Americas, Inc. The views expressed herein are solely those of the author and do not necessarily represent those of Takeda. All information referenced is derived from publicly available sources.

## REFERENCES

- [1] Deloitte Centre for Health Solutions, *Measuring the Return from Pharmaceutical Innovation*. London, U.K.: Deloitte Touche Tohmatsu Ltd., 2022.
- [2] Biotechnology Innovation Organization, *Clinical Development Success Rates 2006–2015*. Washington, DC, USA: BIO, 2016.
- [3] U.S. Food and Drug Administration, *Artificial Intelligence and Machine Learning in Drug Development*, Discussion Paper. Silver Spring, MD, USA, Mar. 2023. [Online]. Available: <https://www.fda.gov/media/167973/download> (accessed Jul. 26, 2025).
- [4] E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.
- [5] MITRE Corporation, *Artificial Intelligence in Clinical Trials*. McLean, VA, USA: MITRE, 2023. [Online]. Available: <https://mitre.org> (accessed Jul. 26, 2025).

- [6] M. I. Miller, L. C. Shih, and V. B. Kolachalama, “Machine Learning in Clinical Trials: A Primer with Applications to Neurology,” *PLoS Comput. Biol.*, vol. 18, no. 4, pp. e1010970, Apr. 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10228463/>. [Accessed: Jul. 30, 2025].
- [7] U.S. Food and Drug Administration, “Artificial Intelligence in Drug Development,” U.S. Food and Drug Administration, Silver Spring, MD, USA. [Online]. Available: <https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development>. [Accessed: Jul. 30, 2025]
- [8] M. C. Marengo, *Predicting drug adverse events using AI*, M.S. capstone project report, Massachusetts Inst. Technol., Cambridge, MA, USA, 2023.
- [9] M. C. Marengo, *An efficient strategy for clinical site allocation*, M.S. capstone project report, Massachusetts Inst. Technol., Cambridge, MA, USA, 2022.
- [10] U.S. Food and Drug Administration, “FDA Proposes Framework to Advance Credibility of AI Models Used in Drug and Biological Product Submissions,” *FDA News*, May 2025. [Online]. Available: <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions>. [Accessed: Jul. 30, 2025].
- [11] U.S. Food and Drug, “Responsive Regulation of Artificial Intelligence in Drug Development,” CDER Office of Medical Policy, May 29, 2024. [Online]. Available: <https://www.fda.gov/media/184256/download>. [Accessed: Jul. 30, 2025].