Artificial intelligence and machine learning in real-world evidence: Transforming data into actionable insights

Pattabhi Machiraju, Hemalatha Jayapalan, Giles Devasahayam, Manjari Deshmukh
Indegene, Bengaluru, India

doi: 10.56012/rimq2971

Correspondence to:
Giles Devasahayam
giles.devasahayam@indegene.com

Abstract

In recent years, the healthcare industry has experienced an exponential growth in the volume of real-world data (RWD) due to advancements in digital health, electronic health records (EHR), wearables, and other data-generating technologies. The integration of artificial intelligence (AI) and machine learning (ML) into real-world evidence (RWE) generation has the potential to revolutionise how clinical and healthcare decisions are made. AI and ML can effectively analyse large and complex datasets, identifying patterns and insights that were previously hidden or too difficult to detect using traditional analytical methods. For medical writers involved in regulatory submissions, clinical research documentation, and healthcare communications, understanding the application of AI and ML in RWE generation is essential. This publication explores the impact of AI/ML on RWE, its regulatory considerations, and best practices for integrating AI-driven insights into medical writing.

As the healthcare industry shifts towards data-driven decision-making, real-world data (RWD) has become a crucial resource for clinical research, regulatory evaluations, and health technology assessments. The global real-world evidence (RWE) solutions market is projected to reach $4.5 billion by 2029, growing at a compound annual growth rate of 16.9% from 2024 to 2029.[1] This growth reflects the increasing reliance on RWD to enhance patient care, refine treatment strategies, and improve healthcare outcomes. However, ensuring the quality of RWD sources and assessing the feasibility of cross-regional data integration remain key challenges.

RWE has emerged as a transformative force in healthcare, reshaping how medical treatments, interventions, and policies are evaluated.[2] Consequently, the demand for more comprehensive, patient-centred evidence has fuelled the rise of RWE, which leverages data from diverse sources such as electronic health records (EHRs), insurance claims, patient registries, wearable health devices, social media and patient-reported outcome, genomic and biomarker data.[3]

The advancement of RWE has been propelled by innovations in data collection, regulatory endorsement, and analytical methodologies. With the expansion of digital health technologies, vast amounts of patient data have become available, enabling large-scale and more rigorous RWE studies. Regulatory agencies, including the U.S. FDA and the EMA, now acknowledge the value of RWE in complementing traditional evidence for drug approvals, safety monitoring, and health policy decision-making.[4]

Unlocking real-world data: AI and ML's role in actionable insights

Unlocking the potential of RWD for actionable insights is a transformative opportunity. AI and ML are key to overcoming its challenges. Artificial intelligence (AI) and machine learning (ML) enhance data quality by automating error detection, managing missing values, and standardising formats. They also handle unstructured data like text, images, and audio through techniques such as natural language processing and computer vision, enabling analysis of medical records, social media, or sensor data. For data integration, AI resolves entity mismatches and maps diverse datasets to common frameworks, while federated learning allows decentralised training without sharing raw data, ensuring privacy.[5]

AI and ML enable predictive and prescriptive analytics, forecasting trends like disease outbreaks or customer behaviour and recommending optimal actions, such as personalised treatments or dynamic pricing.[6] They also provide real-time insights by processing streaming data from Internet of Things devices or social media, with edge computing enabling instant analysis on devices. To address bias and fairness, AI detects biases in datasets and designs fairness-aware algorithms, while synthetic data generation helps reduce biases and protect privacy.[7]

AI also supports personalisation, powering recommendation systems in retail or healthcare, and enhances scalability through automated machine learning (AutoML) and distributed computing. Privacy-preserving techniques like federated learning and differential privacy ensure secure data usage. Finally, explainable AI (XAI) tools make models transparent, building stakeholder trust and enabling actionable insights. By leveraging AI and ML, organisations can transform RWD into meaningful, innovative solutions across industries.

Applications of AI/ML in RWE

Predictive modelling for patient outcomes and disease progression

AI/ML can be applied to large-scale RWD datasets to forecast patient outcomes and understand disease progression. This predictive capability is crucial for personalising treatment strategies and identifying at-risk patients who may benefit from early intervention. Predictive models can estimate the likelihood of disease complications, hospital readmissions, or treatment responses, enabling healthcare providers to proactively tailor interventions.[9] This not only improves patient outcomes but also reduces unnecessary healthcare costs. Additionally, these models help in predicting disease progression, allowing clinicians to identify high-risk patients who may need more aggressive interventions.

For example, in patients with diabetes, AI/ML algorithms can detect patterns indicative of impending complications such as diabetic retinopathy, nephropathy, or cardiovascular disease by analysing trends in haemoglobin A1c levels, blood pressure readings, and renal function markers. Similarly, in patients with heart disease, predictive models can assess risk factors such as cholesterol levels, previous cardiac events, medication adherence, and electrocardiogram findings to estimate the likelihood of heart failure exacerbations or myocardial infarctions.[10]

Identifying treatment patterns and healthcare resource utilisation trends

AI/ML are powerful tools for uncovering trends in treatment patterns and healthcare resource utilisation. By analysing large-scale datasets, these technologies can track how treatment approaches evolve, identify the most effective real-world therapies, and reveal care patterns across diverse patient groups. For instance, ML can examine EHRs to determine the most frequently prescribed treatments for specific conditions, enabling clinicians and policymakers to evaluate the effectiveness of current practices. These insights can also support cost-effectiveness analyses, guiding healthcare systems in optimising resource allocation.[11]

Analysing RWD on hospital admissions and discharge rates can offer valuable insights into the effectiveness of post-surgery care plans. By monitoring readmission rates and length of stay, hospitals can evaluate whether patients receive adequate post-discharge care and pinpoint risk factors leading to complications. Identifying common causes of readmissions, such as infections or medication non-adherence, allows for targeted interventions to improve care co-ordination and prevent unnecessary hospitalisations. Predictive analytics can further enhance resource utilization by optimising staffing, bed management, and follow-up care strategies. Ultimately, leveraging RWD refines post-surgical care, reduces readmission rates, and improves both patient outcomes and healthcare system efficiency.[12]

Generative AI (GenAI) in generating RWE to inform and optimise clinical trial design using RWD

GenAI helps enhance the design of a clinical trial (e.g., inclusion/exclusion criteria, endpoints, comparator arms) by leveraging RWD such as EHR, claims data, and patient registries, transformed into actionable real-world evidence (RWE) leading to faster patient recruitment, cost reduction, dynamic protocol adjustments and higher external validity (Table 1). GenAI models process large-scale, unstructured RWD (e.g., clinician notes, lab results) aiding data ingesting and harmonization.[13] On the basis of target disease profiles, GenAI simulates patient cohorts from RWD to reflect "real-world" populations, identifies characteristics such as comorbidities, treatment patterns, and geographic distribution, which helps simulate patient cohorts.[14] In comparative effectiveness modelling, GenAI generates synthetic control arms using matched historical patient data, reducing or replacing the need for placebo groups and assesses comparative effectiveness of drugs, aiding in endpoint selection.[15] In trial feasibility and site

Table 1. AI/machine learning applications in real world evidence generation across healthcare

Use case	Application	Data source	AI/ML role	Benefits
Advancing cancer research and drug development	Processing genomic data to identify therapy-linked mutations	Genomic profiles from cancer patients	Analyse genetic mutations to develop targeted, personalised treatments	Enhanced treatment precision, improved outcomes
Forecasting hospital readmissions	Predicting which patients are at high risk of readmission	EHRs: comorbidities, past hospital stays, labs, social determinants	Identify high-risk patients by evaluating clinical and non-clinical risk factors	Lower readmission rates, better care quality, reduced healthcare costs
Monitoring drug safety in real-time	Continuous pharmacovigilance post-marketing	EHRs, insurance claims data	Detect adverse events and safety signals in near real-time	Improved patient safety, proactive regulatory decisions
Patient stratification and disease progression	Grouping patients by risk level and predicting disease progression	Clinical records, genetic markers, drug use, adherence, cost data	Stratify patients into risk categories and forecast disease trajectory	Better outcomes, personalised care, efficient resource use

Abbreviations: EHR, electronic health records; ML, machine learning

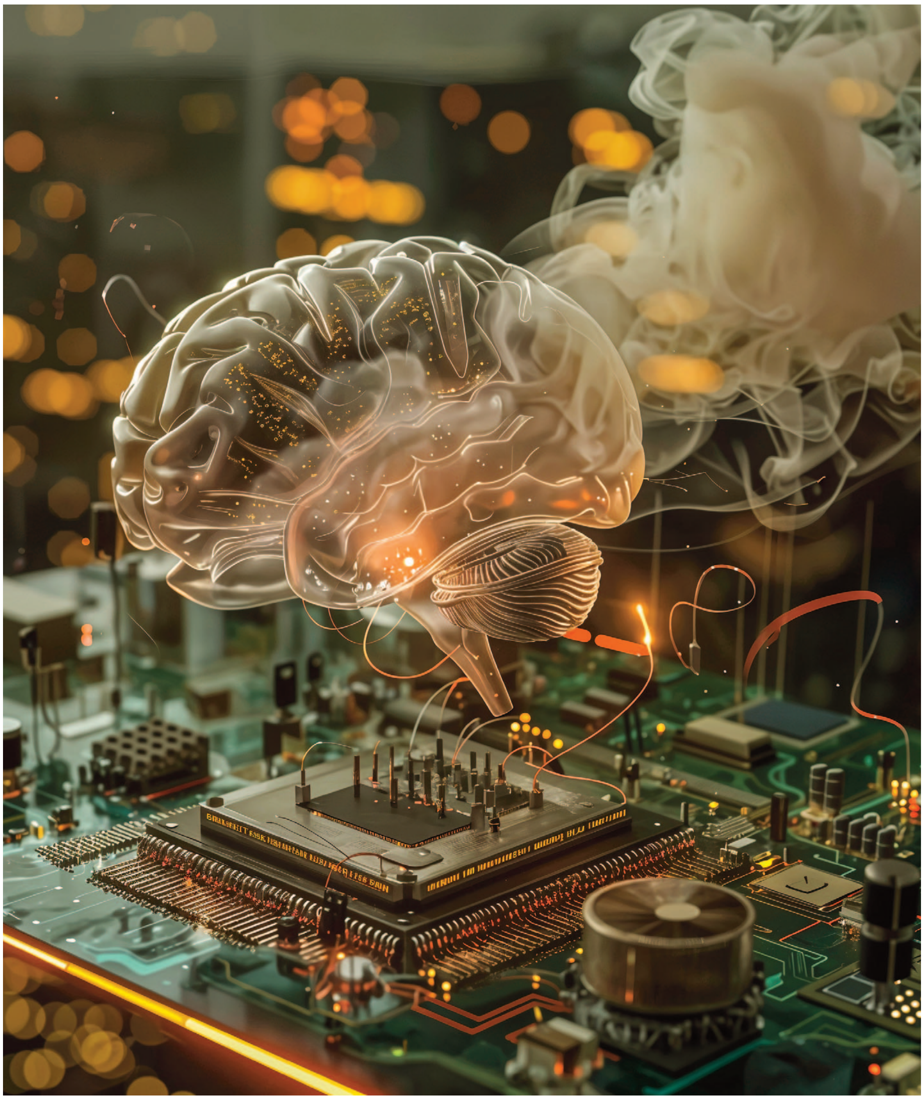

Image: Freepik

selection process GenAI models use RWE to predict patient availability by geography and site, optimising site selection and reducing recruitment delays process.[16,17] GenAI can simulate synthetic patient populations that reflect real-world demographics and disease patterns, in synthetic data generation process which enables simulated control arms for clinical trials and pre-trial feasibility assessments.[18] GenAI has been significantly used in protocol optimisation and feasibility studies to generate and refine trial protocols by predicting dropout rates, identifying high-performing sites, and assessing protocol feasibility based on RWD.[19]

Challenges and ethical considerations

Tackling bias in data and algorithms

A major challenge in applying AI and ML to RWE is ensuring the training data are unbiased. RWD often lack representation of certain populations, which can result in skewed outcomes and disparities in healthcare. When AI models are trained on such imbalanced data, they may reinforce these biases, producing less accurate predictions for underrepresented groups.[20]

Some recommended approaches that could be used to resolve the challenge are:

a. Using diverse and representative datasets for training AI/ML models, ensuring that all demographic groups are adequately represented.
b. Performing regular audits of algorithms to identify and mitigate biases.
c. Utilising fairness-aware ML techniques that are designed to reduce bias during model development.

Promoting transparency and interpretability in AI/ML models enhancing transparency and interpretability in AI/ML models

The lack of transparency in many AI systems – particularly those based on deep learning – poses a significant concern in healthcare, where it's essential to understand how decisions are made. Clinicians, patients, and regulators require insight into the reasoning behind model outputs, but this is often hindered by the so-called "black box" nature of complex models.

A few approaches could foster transparency and interpretability in AI/ML models:

a. **Implementing explainable AI (XAI) approaches to shed light on model decision-making processes.** Techniques like feature importance analysis and decision trees can offer valuable insight into how predictions are generated, helping stakeholders grasp the underlying rationale.
b. **Designing intuitive interfaces that clearly convey model outputs** to healthcare providers and patients, making complex results easier to interpret and use in decision-making.
c. **Offering clear documentation and training resources** to educate users about the assumptions, limitations, and appropriate use of AI/ML models.
d. **Foster collaboration among data scientists, healthcare professionals, and regulators** to ensure models are both technically robust and aligned with clinical needs.

The role of medical writers in interpretation and dissemination of RWE insights derived from AI/ML

Simplifying AI/ML insights for clinical decision-making

Medical writers act as a bridge between the technical world of AI/ML and the practical needs of clinicians. They take complex outputs – such as predictive models, risk scores, or pattern recognition – and translate them into clear, actionable insights. By focusing on how these insights can improve patient outcomes or streamline clinical workflows, medical writers ensure that AI/ML findings are not just understood but also applied effectively in real-world healthcare settings.

Communicating RWE with precision and transparency

When medical writers interpret RWE and generate insights, they must take several precautions to ensure accuracy, relevance, and scientific integrity. RWE can be powerful but also complex and prone to misinterpretation if not handled rigorously. Medical writers play a key role in presenting these findings accurately and clearly. They explain the data sources, methodologies, and limitations in a way that is both scientifically rigorous and easy to understand. By providing context and highlighting the most relevant takeaways, medical writers help clinicians and stakeholders trust and utilise AI/ML-driven RWE in their decision-making.

Delivering actionable insights without compromising integrity

Medical writers ensure that AI/ML-driven insights are presented in a way that is both practical and ethical. They focus on how the findings can be used to improve patient care or inform policy, while also addressing the limitations of the data and models. By avoiding overhyped claims and emphasising the need for responsible interpretation, medical writers help maintain the credibility and scientific integrity of AI/ML applications in healthcare.

Conclusion

The integration of AI/ML into RWE is revolutionising healthcare decision-making by transforming vast amounts of data into actionable insights. The application of AI/ML methods to RWD enables the identification of patterns, prediction of outcomes, and simulation of interventions at scale. However, the complexity of these methodologies poses significant communication challenges. Medical writers are key to ensuring that the resulting insights are both accessible and meaningful to diverse stakeholders. As AI/ML-derived RWE becomes increasingly central to healthcare innovation, the role of the medical writer is evolving from content developer to strategic partner. By ensuring transparency, accuracy, and accessibility, medical writers enable stakeholders to act on complex insights with confidence.

Disclaimers

The views expressed in this article are the personal opinions of the authors and do not necessarily reflect the position of the authors' affiliated organisations.

Disclosures and conflicts of interest

The authors declare no conflict of interest.

References

1. Real world evidence solutions market: growth, size, share, and trends. [AW1] [VW2] Markets and Markets. 2024 [cited 2025 Feb 17]. Available from: https://www.marketsandmarkets.com/Market-Reports/real-world-evidence-solution-market-76173991.html
2. Kaatje B, Chloe WT, Miller E, Izurieta HS, Black S, Andrews N, et al. The role of real-world evidence for regulatory and public health decision-making for Accelerated Vaccine Deployment- a meeting report. Biologicals. 2024;85:101750. doi:101016/j.biologicals.2024.101750
3. Dang A. Real-world evidence: A primer. Pharm Med. 2023;37(1):25–36. doi:10.1007/s40290-022-00456-6
4. Shaikh J. AI Revolution in Evidence Synthesis and Research: NICE Insights. Axtriacom. 2024 Sep 23 [cited 2025 Feb 17]. Available from: https://insights.axtria.com/articles/revolutionizing-evidence-synthesis-with-ai-insights-from-nices-latest-position-statement
5. Chen C, Feng X, Li Y, et al. Integration of large language models and federated learning. Patterns (N Y). 2024;5(12):101098. doi:10.1016/j.patter.2024.101098
6. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–58. doi:10.1056/NEJMra1814259
7. Arora A, Wagner SK, Carpenter R, et al. The urgent need to accelerate synthetic data privacy frameworks for medical research. Lancet Digit Health. 2025;7(2):e157-e160. doi:10.1016/s2589-7500(24)00196-1
8. Froling E, Rajaeean N, Hinrichsmeyer KS, et al. Artificial intelligence in medical affairs: A new paradigm with novel opportunities. Pharmaceut Med. 2024;38(5):331–342 doi:10.1007/s40290-024-00536-9
9. Kragelund SH, Kjarsgaard M, Jensen-Fangel S, et al. Research electronic data capture (REDCap®) used as an audit tool with a built-in database. J Biomed Inform. 2018;81:112–8. doi:10.1016/j.jbi.2018.04.005
10. Gunasekeran DV, Ting DSW, Tan GSW, et al. Artificial intelligence for diabetic retinopathy screening, prediction and management. Curr Opin Ophthalmol. 2020;31(5):357–65. doi:10.1097/ICU.0000000000000693
11. Thomas R, Chalkidou K. Cost–effectiveness analysis. In: Cylus J, Papanicolas I, Smith PC, editors. Health system efficiency: How to make measurement matter for policy and management. . Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2016. (Health Policy Series, No. 46.) 6.
12. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NJP Digit Med. 2018;1(1):18. doi:10.1038/s41746-018-0029-1
13. IbrahimM, Al Khalil Y, Amirrajab S, et al. Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. Comput Biol Med. 2025;189:109834. doi:10.1016/j.compbiomed.2025.109834
14. Beam AL, Kohane, IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317–18. doi:10.1001/jama.2017.18391
15. Kalidindi VR, Rehana S, Seethamraju SM, et al. Revolutionizing medicine: Unleashing the power of real-world data and AI in advancing clinical trials. Braz J Pharm Sci. 2024:60:e23980. doi:10.1590/s2175-97902024e23980

16. Sachdeva S, Kaneria J, Malik R, et al. MSR40 leveraging artificial intelligence (AI) and generative AI (GenAI) for transforming real-world evidence (RWE)across the product value chain and industry functions. Value Health. 2024;27(12):S445. doi:10.1016/j.jval.2024.10.2274
17. Chopra H, Shin DK, Munjal K, et al. Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs. Int J Surg. 2023;109(12):4211-–20. doi:10.1097/JS9.0000000000000705
18. Yoon J, Jarrett D, Van der Schaar M. Time-series generative adversarial networks. Adv Neural Inf Process Syst. 2019:32.
19. Chatterjee S, Fruhling A, Kotiadis K, et al. Towards new frontiers of healthcare systems research using artificial intelligence and generative AI. Health Syst (Basingstoke). 2024;13(4):263–73. doi:10.1080/20476965.2024.2402128
20. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. doi:10.1016/j.jbi.2020.103655

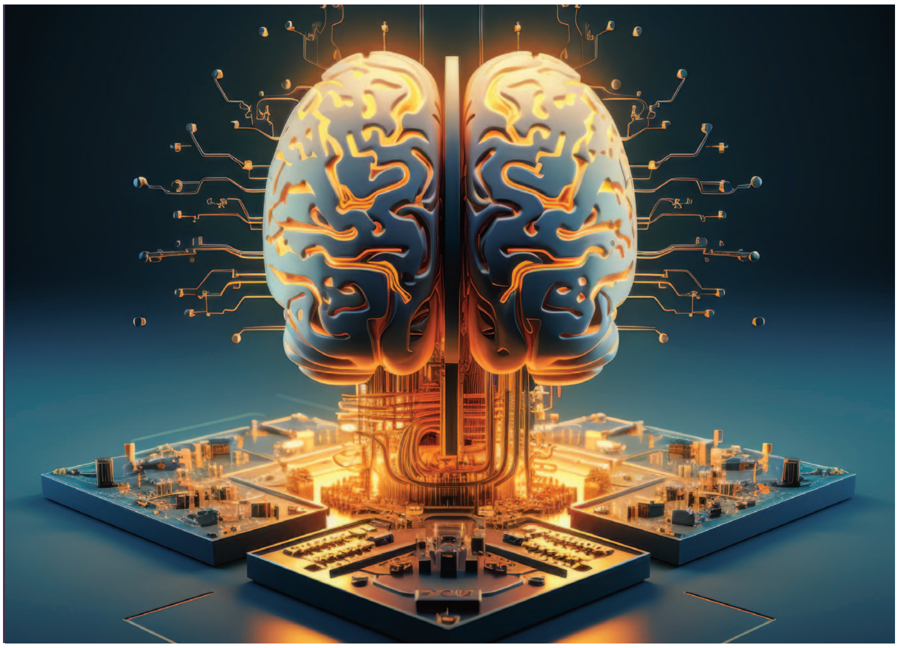

Image: Freepik

Author information

Pattabhi Machiraju, PhD, is a Senior Director at RWD & Biostats department, Indegene. He leads all AI/ML-driven RWE projects globally and brings over two decades of leadership in biostatistics, real-world data analytics, innovative study designs, regulatory submissions, and cross-functional collaboration with global stakeholders and research partners.

Manjari Deshmukh, PG in Bioinformatics, serves as a Lead in Analytics at RWD and Biostatistics. With over 11 years of experience in healthcare, pharmaceuticals, and genomics data, she specialises in advanced data analytics. Her expertise encompasses both qualitative and quantitative studies, focusing on real-world data and RWE.

Dr Hemalatha Jayapalan, BDS, PGDCR, is a Senior Manager at RWE & HEOR, Indegene. She is a Subject Matter Expert leading the RWE & HEOR operations. With over 15 years of experience in the Healthcare and Pharmaceutical industry, she possesses extensive expertise in designing, implementing, and executing evidence generation strategies that support market access, reimbursement, and regulatory submissions.

Giles Devasahayam, M. Pharm, PhD, is a Lead in Analytics – Bioinformatics at Indegene, bringing over 15 years of expertise at the intersection of drug discovery, translational bioinformatics, and real-world data. As a subject matter expert, he leverages advanced analytics techniques to generate actionable insights that drive clinical development and commercial strategy.