NiviScholar Insights

BIOLINK
NSI-BL-2026-001

Can Artificial Intelligence Really Discover Drugs? The Evidence Beyond the Hype

AI is changing target identification, molecular design, and trial operations, but clinical success remains the decisive test.

Shaikh Sumayya Raheman[1*], **Adkine Angad Sharad**[1], **Rohini Prabhu Ghuge**[1]

1 Department of Biotechnology and Bioinformatics, Mahatma Gandhi Mission's College of Computer Science and Information Technology, Nanded, Maharashtra, India

*** Corresponding author:** sumayyashaikh8060@gmail.com

PUBLICATION	COLLECTION	REVIEW	LICENSE
22 July 2026	BioFrontiers Launch Collection 2026	Editorially reviewed	CC BY 4.0

ABSTRACT

Artificial intelligence is now used across drug discovery, from analysing disease biology and predicting protein structures to generating molecules, estimating toxicity, and supporting clinical-trial operations. Public discussion, however, often compresses this complicated pipeline into the claim that "AI discovers drugs." That phrase hides the difference between proposing a candidate and proving that a medicine is safe and effective. This Article examines what the strongest evidence currently supports. AI has demonstrated practical value in prioritising targets, navigating chemical space, and accelerating selected preclinical tasks. A small number of AI-originated programmes have also entered human trials, providing the first opportunity to test whether computational acceleration translates into clinical value. Yet biological uncertainty, biased data, weak prospective validation, and high late-stage attrition remain unresolved. The central argument is that AI should be judged not by how quickly it generates molecules, but by whether it improves the quality of decisions throughout the discovery pipeline. The near-term revolution is therefore likely to be disciplined, measurable, and hybrid: algorithms narrowing possibilities, scientists designing decisive experiments, and clinical evidence determining what survives.

KEYWORDS artificial intelligence; drug discovery; generative chemistry; clinical trials; target identification; validation

ARTICLE AT A GLANCE

- "AI-discovered drug" can refer to target selection, molecular generation, preclinical optimisation, or clinical operations; each requires different evidence.
- The strongest current value is in narrowing search spaces and improving decisions within iterative experimental workflows.
- Clinical-stage AI-originated programmes are important tests, but a few cases cannot establish portfolio-wide superiority.

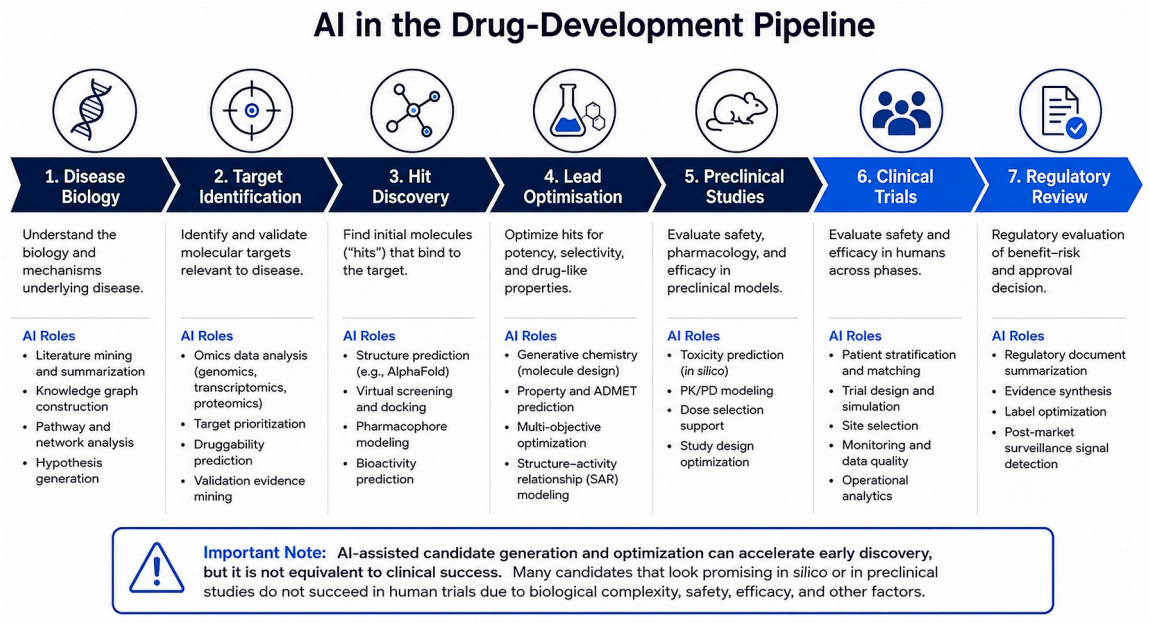

Figure 1. The drug-development pipeline and the points at which artificial intelligence can support decisions. AI does not replace experimental, clinical, or regulatory validation.

1. "AI-discovered" is not a single scientific event

Drug discovery is a chain of decisions rather than one invention. Researchers must identify a disease mechanism, select a tractable target, find or design a molecule, optimise potency and selectivity, assess absorption and toxicity, manufacture the candidate, and then test it in people. Machine learning can support nearly every step, but success at one step does not guarantee success at the next. A model may predict that a compound binds a protein while missing metabolism, off-target effects, tissue exposure, or the fact that the selected target is not causally important in patients. Reviews of the field therefore distinguish computational performance from therapeutic performance (Bender & Cortés-Ciriano, 2021a, 2021b; Vamathevan et al., 2019).

The phrase "AI-designed drug" can describe very different contributions. In one programme, AI may rank known targets. In another, a generative model may propose a novel chemical structure. Elsewhere, machine learning may optimise trial recruitment or predict which patients are most likely to respond. These are all meaningful uses, but they carry different evidentiary burdens. A target-selection model should be judged by prospective biological validation; a molecular generator by synthesis, potency, selectivity, and developability; and a clinical model by performance on representative patients. Treating all of them as one category makes both praise and criticism scientifically imprecise.

The most useful question is therefore not whether AI "works" in drug discovery. It is where the technology improves a decision compared with a strong non-AI baseline. This framing matters for students because spectacular demonstrations can distract from ordinary sources of failure: poorly defined labels, narrow datasets, information leakage, and optimisation against endpoints that are convenient rather than clinically meaningful. Explainability does not automatically solve these problems, but interpretable evidence can help researchers identify whether a model has learned chemistry and biology or merely repeated correlations in its training data (Jiménez-Luna et al., 2020).

2. Where the evidence is already strong

AI is particularly effective when a problem has abundant, well-curated data and a measurable objective. Molecular property prediction, virtual screening, image analysis, literature mining, and protein-structure prediction fit this pattern better than open-ended questions about complex disease causality. AlphaFold 3, for example, extended structure prediction to biomolecular complexes containing proteins, nucleic acids, ligands, ions, and modified residues, creating richer starting hypotheses for medicinal chemistry and structural biology (Abramson et al., 2024). A predicted complex is not proof of binding, but it can guide mutagenesis, assay design, and prioritisation.

Generative chemistry provides another concrete example. In 2019, researchers reported a deep-learning workflow that identified potent inhibitors of discoidin domain receptor 1 within a compressed design-and-test cycle (Zhavoronkov et al., 2019). The result did not establish a medicine; it demonstrated that generative methods could propose synthesizable

compounds with relevant activity. This distinction is important. Early discovery is dominated by search, and algorithms can reduce the number of compounds that must be made. The value is real even when the final molecule still requires extensive human-led optimisation.

AI is also becoming useful in less visible parts of development. Models can help identify trial sites, match participants to eligibility criteria, monitor safety signals, and standardise regulatory documents. These applications may lack the drama of de novo molecule generation, yet they address delays that directly affect development cost and duration. The U.S. Food and Drug Administration has accordingly emphasised a risk-based approach to AI and machine learning in drug and biological product development, with attention to context of use, data governance, model performance, and lifecycle management (U.S. Food and Drug Administration, 2023).

3. Human trials are the necessary reality check

The field entered a more demanding phase when AI-originated drug programmes reached clinical testing. The TNIK inhibitor rentosertib, developed for idiopathic pulmonary fibrosis, became a prominent case because AI contributed to both target identification and molecular design. Preclinical and early clinical work reported a coherent antifibrotic rationale, and a randomised phase 2a study provided evidence about tolerability and exploratory efficacy in patients (Ren et al., 2025; Xu et al., 2025). This is more informative than a retrospective benchmark because the programme had to survive synthesis, toxicology, manufacturing, regulatory review, and human dosing.

The correct interpretation is neither triumph nor dismissal. A phase 2a signal does not equal regulatory approval, and a single programme cannot establish that AI broadly increases success rates. Drug development is affected by indication, trial design, dose selection, patient heterogeneity, and chance. What the programme does show is that an AI-supported pipeline can produce a clinically testable candidate on a compressed timetable. The next question is whether multiple programmes demonstrate reproducible advantages in probability of success, not merely speed to nomination.

This is why prospective evidence matters more than headline counts of generated molecules. A model can create millions of structures, but the laboratory can synthesise and test only a small fraction. The bottleneck shifts from generation to selection. Strong platforms therefore integrate computation with repeated experimental feedback, sometimes described as a design-make-test-analyse loop. The algorithm proposes; chemistry and biology measure; the new data update the next proposal. AI becomes most valuable when embedded in this cycle rather than treated as an oracle.

4. The unresolved problems are biological and institutional

Training data in drug discovery are not neutral. Published datasets overrepresent successful experiments, well-studied protein families, and chemical series that companies are willing to disclose. Negative results are often missing. Assay conditions vary, labels can be noisy, and compounds that appear inactive may simply have been tested in unsuitable systems. A model trained on these records can reproduce historical attention rather than reveal new biology. Data diversity is therefore not an administrative detail; it determines which diseases, populations, and molecular mechanisms are visible to the system.

Generalisation is equally difficult. Random train-test splits can place closely related compounds in both sets and produce optimistic results. More realistic evaluations separate chemical scaffolds, time periods, laboratories, or therapeutic programmes. Prospective testing is stronger still because predictions are frozen before experiments are conducted. Bender and Cortés-Ciriano (2021a, 2021b) argue that unrealistic benchmarks and poorly matched baselines are major sources of inflated expectations. A useful model must outperform the decision process it is intended to replace, not merely a weak comparator.

Institutional incentives can also distort evaluation. Start-ups benefit from describing platforms as revolutionary, while pharmaceutical companies may publicise speed but withhold failures and cost data. Academic groups are rewarded for novel architectures more than for careful replication. As a result, the evidence base contains impressive case studies but limited portfolio-level comparison. Transparent reporting should include how targets were chosen, how many compounds were generated and synthesised, which candidates failed, what experiments changed the model, and whether timelines or attrition improved against historical controls.

5. A more scientific way to judge the AI-drug revolution

The first criterion should be decision quality. Did the model identify a target with stronger human genetic or mechanistic support? Did it reduce unnecessary synthesis while maintaining chemical diversity? Did it detect toxicity earlier? Did it improve patient selection without excluding underrepresented groups? These questions connect algorithmic outputs to the actual reasons drug programmes fail. They also make it possible to recognise partial value. An AI tool may be worthwhile even if it does not create the final drug, provided it reliably removes bad options sooner.

The second criterion should be reproducibility. Models should be evaluated on temporally separated data and, where possible, in prospective studies. Reporting must describe datasets, uncertainty, failure modes, and human interventions. The third criterion is clinical translation. Ultimately, medicines are judged by benefits and harms in patients. A faster route to a weak candidate is not progress; a faster route to a better-supported candidate is. Clinical outcomes will therefore remain the final audit of every claim about AI-driven discovery.

The emerging evidence supports a balanced conclusion. AI is already changing how scientists search chemical and biological space, and some programmes have progressed far enough to challenge the idea that the field is only hype. At the same time, no algorithm can remove the uncertainty of living systems. The most credible future is not autonomous drug discovery but a tighter scientific loop in which computational models prioritise experiments, experimental data correct the models, and human judgement defines the questions worth asking. That is less cinematic than a machine inventing a cure, but it is a more plausible route to better medicines.

Key takeaways

- "AI-discovered drug" can refer to target selection, molecular generation, preclinical optimisation, or clinical operations; each requires different evidence.
- The strongest current value is in narrowing search spaces and improving decisions within iterative experimental workflows.
- Clinical-stage AI-originated programmes are important tests, but a few cases cannot establish portfolio-wide superiority.
- Prospective validation, realistic data splits, transparent failures, and clinically meaningful endpoints are essential.
- AI should be judged by improved decision quality and patient outcomes, not by the number of molecules generated.

Declarations

Author contributions: All listed authors contributed to literature synthesis, manuscript development, revision, and approval of the publication version.

Funding: No specific funding was reported for this article.

Competing interests: The authors declare no competing interests.

Ethics: Not applicable. This article does not report new research involving human participants, animals, or identifiable personal data.

AI-assisted tools: AI-assisted tools were used for language refinement, structural review, and preparation of selected conceptual visuals. All scientific claims, citations, interpretations, and final text were reviewed and approved by the authors and the NiviScholar editorial team.

How to cite this article

Shaikh Sumayya Raheman; Adkine Angad Sharad; Rohini Prabhu Ghuge (2026). Can Artificial Intelligence Really Discover Drugs? The Evidence Beyond the Hype. NiviScholar Insights, NSI-BL-2026-001. https://insights.nivischolar.com/articles/can-artificial-intelligence-really-discover-drugs

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024).

Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630, 493–500. Accurate structure prediction of biomolecular interactions with AlphaFold 3 | Nature

Bender, A., & Cortés-Ciriano, I. (2021a). Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discovery Today, 26(2), 511–524.

Bender, A., & Cortés-Ciriano, I. (2021b). Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: A discussion of chemical and biological data. Drug Discovery Today, 26(4), 1040–1052.

Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2, 573–584. https://doi.org/10.1038/s42256-020-00236-4

Ren, F., Aliper, A., Chen, J., Zhao, H., Rao, S., Li, X., ... Zhavoronkov, A. (2025). A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. Nature Biotechnology, 43, 63–75. https://doi.org/10.1038/s41587-024-02143-0

Schneider, P., Walters, W. P., Plowright, A. T., Sieroka, N., Listgarten, J., Goodnow, R. A., Fisher, J., Jansen, J. M., Duca, J. S., Rush, T. S., Zentgraf, M., Hill, J. E., Krutoholow, E., Kohler, M., Blaney, J., Funatsu, K., Luebkemann, C., & Schneider, G. (2020). Rethinking drug design in the artificial intelligence era. Nature Reviews Drug Discovery, 19, 353–364. https://doi.org/10.1038/s41573-019-0050-3

U.S. Food and Drug Administration. (2023). Using artificial intelligence and machine learning in the development of drug and biological products: Discussion paper and request for feedback. https://www.fda.gov/media/167973/download

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18, 463–477. https://doi.org/10.1038/s41573-019-0024-5

Xu, Z., Ren, F., Wang, P., Cao, J., Tan, C., ... Zhavoronkov, A. (2025). A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: A randomized phase 2a trial. Nature Medicine, 31, 2602–2610. https://doi.org/10.1038/s41591-025-03743-2

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., ... Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology, 37, 1038–1040. https://doi.org/10.1038/s41587-019-0224-x