Nano World Journal

https://doi.org/10.17756/nwj.2025-142

Editorial

Open Access

From Empiricism to Inverse Design: How Artificial Intelligence is Reshaping Nanobiotechnology

Roberto Eggenhoffner

Department of Surgical Sciences and Integrated Diagnostics (DISC), University of Genoa, Italy

*Correspondence to:
Roberto Eggenhoffner
Department of Surgical Sciences and Integrated Diagnostics,
University of Genoa,
Genova, Italy.
E-mail: roberto.eggenhoffner@unige.it

Received: Jul 15, 2025
Accepted: July 18, 2025
Published: July 21, 2025

Citation: Eggenhoffner R. 2025. From Empiricism to Inverse Design: How Artificial Intelligence Is Reshaping Nanobiotechnology. *NanoWorld J* 11(1): 15-18.

Copyright: © 2025 Eggenhoffner. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY) (http://creativecommons.org/licenses/by/4.0/) which permits commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited.

Published by United Scientific Group

The convergence of nanobiotechnology and artificial intelligence (AI) is reshaping our discipline, moving it from trial-and-error experimentation toward predictive, systematically engineered approaches. AI has moved from a supporting analytic tool to a central driver of discovery and translation. At the same time, researchers are gaining finer control over nanoscale architectures and biological interfaces, a trend reflected in routine laboratory practice. Laboratory experiments are increasingly structured around machine-learning (ML) requirements, with descriptors formatted for algorithmic use and closed-loop experimentation used to shorten validation cycles. The impact is not only faster progress but also a redefinition of what counts as evidence, reproducibility, and translational readiness. A clear illustration of this transformation is the inverse design of lipid nanoparticles (LNPs) for nucleic acid delivery, where ML models start from desired biological outcomes and work backward to propose nanoparticle structures most likely to achieve them.

This shift is already evident in primary research. Although further experimental confirmation is needed, studies show that embedding biological constraints, such as endosomal escape or tissue tropism, into model training can shape nanocarrier performance. Wang et al. [1] used AI-guided rational design to explore the vast chemical space of ionizable lipids and confirmed novel candidates that matched or outperformed existing standards *in vivo* [1]. A parallel effort at MIT trained deep models on thousands of LNP activity profiles and identified new ionizable lipids that enabled efficient pulmonary mRNA delivery in mice and ferrets, two particularly stringent preclinical models for respiratory delivery [2]. Together, these studies illustrate how biological constraints can be converted into machine-readable design spaces, while also underscoring the need for broader validation across systems and applications. This direction is consistent with recent reviews reporting the promise of AI-assisted LNPs design for nucleic acid delivery and its implications for clinical translation [3].

Realizing these advances depends on a concrete and evolving stack of computational tools that is foundational to nanomaterials informatics and predictive nanotoxicology, even if adoption varies across research groups. Python has become the dominant programming language in these domains, supporting rapid prototyping and reproducibility, while proprietary environments still persist in some specialized settings. For tabular, descriptor-rich datasets, gradient-boosting methods such as XGBoost remain highly effective, with Scikit-learn widely used for rapid baselines and algorithm selection. TensorFlow and PyTorch provide flexible frameworks for representation learning, well-suited to capturing nonlinear structure–function relationships and accommodating diverse input types, including sequences and graphs. Hyperparameter optimization is automated by libraries such as Optuna, which use Bayesian strategies to efficiently identify optimal configurations. Versioning and traceability of models and datasets are facilitated by platforms like DVC and ML flow, though their systematic adoption is not yet widespread. Finally, interpretability tools such as SHAP and integrated

gradients help researchers identify the features driving predictions and interrogate model logic, with insights that become meaningful only when coupled with domain expertise and experimental validation. Together, these resources form the backbone of AI-enabled nanobiotechnology.

In parallel with advances in model interpretability, the physical infrastructure of discovery is also adapting to the needs of AI-driven science. Microfluidic and lab-on-chip platforms, coupled with automated imaging, generate dense and standardized datasets well suited for ML. Image analysis methods such as U-Net segmentation and transformer-based tracking, often implemented in PyTorch, convert raw microscopy into reliable single-particle and single-cell trajectories, while active-learning frameworks steer experiments toward the most informative regions of design space. These advances underpin feedback systems in which computational models act as experimental partners, guiding data acquisition under human oversight. A clear example comes from vaccine-oriented research, where datasets on transfection efficiency across multiple cell types have enabled ML models to extract generalizable design rules. This logic underlies recent AI-guided development of LNPs for immunization [4]. Here, generative models such as VAEs or diffusion architectures in PyTorch are paired with property predictors in Scikit-learn or XGBoost to create closed-loop pipelines that propose, evaluate, and experimentally validate candidate materials, sometimes using automated microfluidic platforms. Although still concentrated in specialized laboratories, these approaches demonstrate how computational and experimental methods can be integrated to accelerate improvements.

The same computational infrastructure is increasingly applied to predictive nanotoxicology, where approaches are moving from illustrative case studies toward decision-support tools for early safety assessment. Supervised learning models trained on complex descriptors—such as zeta potential, hydrodynamic size, aspect ratio, shell composition, ligand density and protein corona signatures—have shown promise in flagging high-risk nanomaterials before costly or ethically sensitive assays are required. More recent work incorporates uncertainty estimation, helping experimentalists prioritize the most informative conditions rather than simply the most likely outcomes. In practice, Scikit-learn and XGBoost remain central for multi-endpoint toxicity prediction, while PyTorch-based frameworks are being explored for integrating raw imaging or single-cell data with conventional descriptors.

This strategy of computational triage, which involves screening candidates before directing targeted assays, aligns with ongoing regulatory discussions around model explainability and traceable data lineage, though formal acceptance will depend on further validation across systems. ML tools will only gain traction if paired with transparent evaluation criteria and harmonized validation frameworks [5]. AI is also beginning to shape pharmacokinetic (PK) modeling in nanomedicine, a critical area where nanoscale carriers often behave unpredictably compared to small molecules. Early studies have applied multi-view learning frameworks that combine nanoparticle properties, formulation variables, and biological priors to map how structural features influence PK, with implications for dosing and safety. More recently, hybrid models that integrate deep neural representations with ensemble learners [6] have shown promise in predicting PK parameters such as circulation half-life, organ distribution, and clearance. These approaches suggest ways to narrow the gap between *in vitro* observations and *in vivo* performance. However, broader validation and alignment with regulatory standards will be essential before such methods can reliably inform first-in-human planning.

Integrating multi-omics data is now central to tailoring nanomedicine. Single-cell and spatial analyses in transcriptomics, proteomics, and metabolomics can resolve biological heterogeneity at the same scale where nanocarriers act, offering clues for more precise targeting. Translating these high-dimensional measurements into actionable design rules remains a major challenge, but emerging opportunities include identifying receptor landscapes that favour particular ligands, tumour microenvironments responsive to specific stiffness or surface properties, and immune states requiring adjuvant co-delivery. Early-stage pipelines have begun to address this by training deep learning models on multi-modal data to generate representations of clinically relevant phenotypes, which can then be linked to property predictors for formulation design. Although still rare, proof-of-concept studies have used omics-derived signatures to guide LNP composition and dosing strategies in micro physiological systems. The long-term vision is adaptive therapy, with delivery systems adjusting to a patient's molecular profile, but progress depends on better data integration, validation, and translation. This design logic is now being extended to environmental challenges such as water purification. As in drug delivery, the process begins by defining key constraints such as permeability, selectivity, antifouling properties, and catalytic stability, and by representing them as descriptors.

Structure–function relationships can then be learned from experimental and historical datasets, while inverse design methods propose optimized nanocomposites and membrane architectures. The computational toolkit is similar: Scikit-learn and XGBoost for adsorption and permeability prediction, graph neural networks in PyTorch or TensorFlow for catalyst optimization, and global optimization methods to balance competing objectives. Environmental AI applications remain at an early stage, with smaller and less standardized datasets than in biomedicine. Nevertheless, pilot-scale studies have already demonstrated AI-optimized sorbents and photocatalysts, for example AI-assisted nanomaterials for clean energy and carbon capture [7] and magnetic nanocomposites for antibiotic degradation in wastewater [8]. Further, AI-driven models applied to wastewater treatment have shown high predictive accuracy for key effluent parameters—chemical oxygen demand, biochemical oxygen demand, total nitrogen and phosphorus—especially when using ensemble methods such as XGBoost and Gradient Boosting [9]. These results suggest that design-for-deployment is technically feasible when data collection is structured for learning, paralleling the closed-loop strategies currently emerging in personalized nanomedicine. Concurrently, nanoparticle and nanocomposite-based systems—using metals, metal oxides, carbon-based or hybrid

structures—continue to demonstrate promising performance in pollutant adsorption, catalysis, and membrane filtration [10].

To consolidate progress in both biomedical and environmental nanotechnology, AI systems must place priority on interpretability, robustness, and governance. Explainable AI should be embedded into decision-making, for example, using SHAP values, counterfactual analyses, and saliency maps, while ensuring that outputs are interpreted with domain expertise and validated experimentally. Robustness should be reinforced through cross-laboratory stress testing, drift detection, and auditable data versioning. These are not abstract ideals but practical quality-assurance measures already achievable with current Python-based toolchains.

Effective collaboration is strengthened when computational, experimental, and clinical communities work in alignment. Computational scientists can contribute by developing code that is reproducible and straightforward to use, while experimentalists provide model-ready descriptors that include information on variance and provenance. Reproducibility is further enhanced when preprocessing pipelines, training protocols, and model parameters are shared openly, allowing other groups to replicate results under comparable conditions. Involving clinicians and regulators early helps ensure that research outputs remain consistent with translational pathways. Notable advances, such as AI-guided LNP design [1, 2] and PK prediction incorporating biological priors [4], have emerged when experimental and computational strategies evolved together. Predictive modeling and data standards will be decisive for turning proof-of-concept AI applications into reproducible practice [3, 5]. Still, current approaches are not without limitations: AI models often face data sparsity and noise, poor transferability across cell lines or animal models, and domain shifts in PK prediction, where patterns learned in one context may fail in another. Recognizing these failure modes is critical to avoid overinterpretation and to design validation strategies that ensure robustness.

Collectively, advances in standards, predictive modeling, and awareness of limitations position AI as a practical driver of nanobiotechnology. Realizing this transition will require data standards that support cross-study learning, descriptors that integrate physics, chemistry, and biology, experimental platforms responsive to model queries, and regulatory frameworks that formally incorporate explainable and auditable AI into evidence standards. With such a foundation, design can be guided early by generative models constrained by biology, safety anticipated through predictive toxicology, and PK forecast at the project stage rather than measured only later. Personalization will ultimately rely on adaptive pipelines that integrate patient-specific multi-omics data, including comprehensive molecular profiles spanning DNA, RNA, proteins, metabolites, and their spatial organization, so that nanoparticle composition and dosing strategies can be tailored to each patient's biology.

AI in nanobiotechnology is entering a phase defined by clearer strategic priorities. Inverse-design generators will need to incorporate hard constraints from the outset to ensure outputs are experimentally realizable. Predictive models that inform decisions should embed uncertainty estimation and domain shift detection. Full publication of training and preprocessing pipelines will be essential for reproducibility across institutions, while user-friendly interfaces can broaden access for researchers outside computational domains.

Key challenges remain: AI-predicted nanotoxicology requires validation across multiple laboratories, materials, and biological systems. Integrating omics data with AI models in real time will also demand standardized procedures and seamless interfaces between experimental platforms and computational pipelines. At the same time, nanomedicine as a field needs publicly available, standardized datasets with comprehensive metadata to enable fair model comparison and transparent progress tracking, which would provide the foundation for reproducibility and broader translational impact. Addressing these priorities will require coordination among experimental consortia, data science teams, and regulatory stakeholders, but together they form the foundation for realizing the full potential of AI in nanobiotechnology. If successful, these efforts could accelerate discovery, strengthen clinical translation, and broaden the role of nanoscale solutions across medicine, industry, and environmental applications (Figure 1).

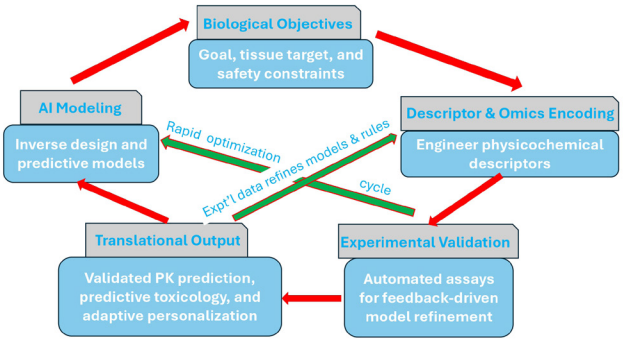

Figure 1: AI is reshaping nanobiotechnology.

References

1. Wang W, Chen K, Jiang T, Wu Y, Wu Z, et al. 2024. Artificial intelligence–driven rational design of ionizable lipids for mRNA delivery. *Nat Commun* 15(1): 1-17. https://doi.org/10.1038/s41467-024-55072-6
2. Witten J, Raji I, Manan RS, Beyer E, Bartlett S, et al. 2024. AI-guided design of lipid nanoparticles for pulmonary mRNA delivery identifies FO-32 and FO-35 with superior performance in preclinical models. *Nat Biotechnol* 3(2): 100123. https://doi.org/10.1038/s41587-024-02490-y
3. Mohr AE, Ortega-Santos CP, Whisner CM, Klein-Seetharaman J, Jasbi, P. 2024. Navigating challenges and opportunities in multi-omics integration for personalized healthcare. *Biomedicines* 12(7): 1496. https://doi.org/10.3390/biomedicines12071496
4. Khakpour A, Florescu L, Tilley R, Jiang H, Iyer KS, et al. 2025. AI-Powered prediction of nanoparticle pharmacokinetics: a multi-view learning approach. *arXiv Preprint*. https://arxiv.org/abs/2503.13798
5. Wang Y, Xiao Z, Wang Z, Lee D, Ma Y, et al. 2024. Multi-omics approaches to decipher the interactions of nanoparticles and biological systems. *Nat Rev Bioeng* 3(4): 333-348. https://doi.org/10.1038/s44222-024-00264-4
6. Bae SH, Choi H, Lee J, Kang MH, Ahn SH, et al. 2025. Rational design of lipid nanoparticles for enhanced mRNA vaccine delivery via machine learning. *Small* 21(8): e2405618. https://doi.org/10.1002/smll.202405618

7. Chen H, Zheng Y, Li J, Li L, Wang X. 2023. AI for nanomaterials development in clean energy and carbon capture, utilization and storage (CCUS). *ACS Nano* 17(11): 9763-9792. https://doi.org/10.1021/acsnano.3c01062

8. Zulfiqar N, Nadeem R, Musaimi OA. 2024. Photocatalytic degradation of antibiotics via exploitation of a magnetic nanocomposite: a green nanotechnology approach toward drug-contaminated wastewater reclamation. *ACS Omega* 9(7): 7986-8004. https://doi.org/10.1021/acsomega.3c08116

9. Dikmen F, Demir A, Özkaya B, Raza MO, Rasheed J, et al. 2025. AI-driven wastewater management through comparative analysis of feature selection techniques and predictive models. *Sci Rep* 15(1): 25347. https://doi.org/10.1038/s41598-025-07124-0

10. Tripathy J, Mishra A, Pandey M, Thakur RR, Chand S, et al. 2024. Advances in nanoparticles and nanocomposites for water and wastewater treatment: a review. *Water* 16(11): 1481. https://doi.org/10.3390/w16111481