

Review

Not peer-reviewed version

From Prediction to Synthesis: Generative AI Architectures and Digital Twins for the Future of Vaccines and Immuno-Oncology

[Vrinda Gote](#)*

Posted Date: 30 January 2026

doi: 10.20944/preprints202601.2335.v1

Keywords: inverse vaccinology; generative artificial intelligence; immuno-oncology; neoantigen discovery; digital twins; de novo immunogen design; graph neural networks (GNN); machine learning in vaccinology; precision medicine



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

From Prediction to Synthesis: Generative AI Architectures and Digital Twins for the Future of Vaccines and Immuno-Oncology

Vrinda Gote

Division of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Missouri, 2464 Charlotte Street, Kansas City, MO 64108, USA; vrindagote9694@gmail.com

Abstract

Artificial intelligence (AI) and machine learning (ML) have progressively reshaped vaccinology, enabling the transition from empirical antigen discovery toward computationally guided reverse vaccinology. As the field enters 2026, a further conceptual shift is emerging: the use of generative AI not only to predict immune targets from existing pathogens, but to design immunogens de novo to satisfy predefined immunological objectives. This evolution is particularly relevant at the interface of prophylactic vaccines and therapeutic immuno-oncology, where antigen heterogeneity and patient specificity challenge conventional development paradigms. This review critically examines the transition from predictive to generative AI in vaccinology, a framework we refer to as inverse vaccinology, and evaluates its implications across antigen discovery, delivery system optimization, and early clinical development. I synthesized recent advances in deep learning architectures—including graph neural networks, protein language models, and diffusion-based generative systems—alongside emerging applications of digital immune modeling, Bayesian optimization, and AI-guided formulation design. Emphasis is placed on evidence derived from structural biology, immunopeptidomics, and translational vaccine research. Current evidence suggests that AI-enabled integration of antigen design with delivery and pharmacokinetic modeling can reduce attrition during preclinical development, particularly for mRNA-based vaccines and personalized neoantigen strategies. The convergence of immunogen design, lipid nanoparticle engineering, and in-silico immune modeling highlights a nascent immuno-pharmacology axis that links molecular optimization to biological exposure and immune activation. While generative AI offers a powerful extension of computational vaccinology, its successful translation depends on rigorous validation, transparent modeling assumptions, and realistic assessments of biological uncertainty. Rather than replacing experimental vaccinology, inverse vaccinology should be viewed as a design-acceleration framework that narrows the experimental search space and enables more rational, patient-aware vaccine development.

Keywords: inverse vaccinology; generative artificial intelligence; immuno-oncology; neoantigen discovery; digital twins; de novo immunogen design; graph neural networks (GNN); machine learning in vaccinology; precision medicine

1. Introduction

Vaccine and immuno-oncology drug development remain constrained by long development timelines, high financial risk, and substantial clinical attrition. Despite advances in molecular biology and manufacturing technologies, a majority of vaccine and cancer immunotherapy candidates fail during clinical translation, most commonly due to insufficient efficacy or unanticipated safety signals in humans. These challenges reflect the intrinsic complexity of the immune system and the limitations of empirical, population-averaged development strategies when applied to heterogeneous pathogens and genetically diverse patient populations [1,2].

Over the past decade, computational approaches have begun to reshape this landscape. Reverse vaccinology, originally based on genomic screening and sequence-level heuristics, demonstrated that in-silico prioritization could accelerate antigen identification for infectious diseases. More recently, machine learning (ML) models trained on expanding immunological datasets—spanning immunopeptidomics, structural biology, and multi-omics profiling—have improved predictions of epitope presentation, immunogenicity, and immune escape. These developments have been particularly impactful for T-cell-mediated immunity, where accurate modeling of peptide–HLA interactions is essential for both prophylactic and therapeutic vaccine design [3]. As we enter 2026, a fundamental paradigm shift is occurring; the convergence of high-throughput multi-omics and Artificial Intelligence (AI) is collapsing these timelines and bridging the gap between prophylactic protection against infectious pathogens and therapeutic intervention in immuno-oncology (IO) [4].

Early computational methods relied on linear scanning to identify immune motifs, but modern architectures like Graph Neural Networks (GNNs) and Transformers (Protein Language Models) now simulate the three-dimensional immune synapse with unprecedented fidelity [5]. By treating amino acids as nodes in a spatial graph or as characters in a biological language, AI can predict conformational epitopes in viruses and discover private neoantigens in cancers that were previously invisible to empirical screening, [6]. AI-driven shift focuses on the optimization of clinical development through the deployment of Digital Twins and Synthetic Control Arms [7]. By modeling the patient’s immune system in a virtual environment, researchers can run in-silico trials to predict dosage response and stratify responders before a single injection is administered [7]. This reduces the logistical burden of patient recruitment in oncology and streamlines the regulatory path for vaccines against emerging infectious diseases. [8].

However, most computational vaccinology efforts to date remain fundamentally *predative*, they identify candidate epitopes or antigens from naturally occurring sequences and rank them according to learned patterns. As the field moves into 2026, an important conceptual transition is underway. Advances in generative deep learning—such as protein language models, graph-based neural networks, and diffusion-based structure generators—now permit the *de novo* design of protein sequences and scaffolds that are optimized for predefined structural or immunological constraints. This shift enables a move from asking *which antigens exist* to *which antigens should be constructed* to elicit a desired immune response [9].

The relevance of this paradigm is especially evident in immuno-oncology. Unlike infectious disease vaccines, cancer vaccines must contend with private, patient-specific neoantigens, immune tolerance, and a suppressive tumor microenvironment. AI-driven integration of tumor genomics, transcriptomics, immunopeptidomics, and T-cell receptor modeling has enabled increasingly precise prioritization of neoantigen targets, but clinical success remains inconsistent. Emerging evidence suggests that failures often arise not from antigen selection alone, but from mismatches between antigen design, delivery kinetics, and immune context [10]. Table 1 explains the strategic evolution of computational vaccinology paradigms.

Accordingly, this review adopts a systems-level perspective on AI in vaccinology and immuno-oncology. We examine how modern ML architectures support epitope discovery, generative immunogen design, and delivery optimization, while also critically assessing current limitations, sources of bias, and translational barriers. Rather than presenting AI/ML as a supplementary for experimental immunology, we argue that its most immediate value lies in constraining biological uncertainty, accelerating hypothesis testing, and enabling more rational, data-informed vaccine development pipelines.

Table 1. Strategic Evolution of Computational Vaccinology Paradigms.

Feature	Reverse Vaccinology (2010s)	AI/ML Integration (2020–2024)	AI/ML Integration (2024–onwards)
Primary Goal	Identification of known antigens	Prediction of epitope binding	De novo design of immunogens

Data Source	Reference genomes	Large-scale multi-omics	Generative de novo synthesis
Vaccine Type	Prophylactic (Viral/Bacterial)	Viral & General Cancer	Precision Immunology & Digital Twins
Pharmacology	Antigen-only focus	Basic delivery scaffolds	Integrated Antigen-delivery kinetics

2. The Target: Unlocking B and T Cell Epitopes with Machine Learning

To design a truly effective vaccine, scientists must pinpoint the specific molecular targets, known as epitopes, that the immune system's T and B cells recognize [11]. Traditionally, machine learning (ML) approaches have tackled this challenge through supervised training, essentially teaching a model to discriminate between genuine epitope sites and inert protein [12]. This predictive power is especially vital for B-cell epitopes, which trigger the production of antibodies. B-cell epitopes exist in two primary forms: linear, which are continuous segments of an antigen's amino acid sequence, and conformational, where residues are spatially proximal in the three-dimensional (3D) structure but distal in the linear sequence [13,14]. Because these critical conformational epitopes rely entirely on the protein's folded shape, a much more accurate prediction is achieved when ML models are trained directly on protein structures, allowing them to truly recognise the surface topography and biochemical composition that the immune system targets [15–17]. Figure 1 demonstrates the evolutionary shift in epitope discovery.

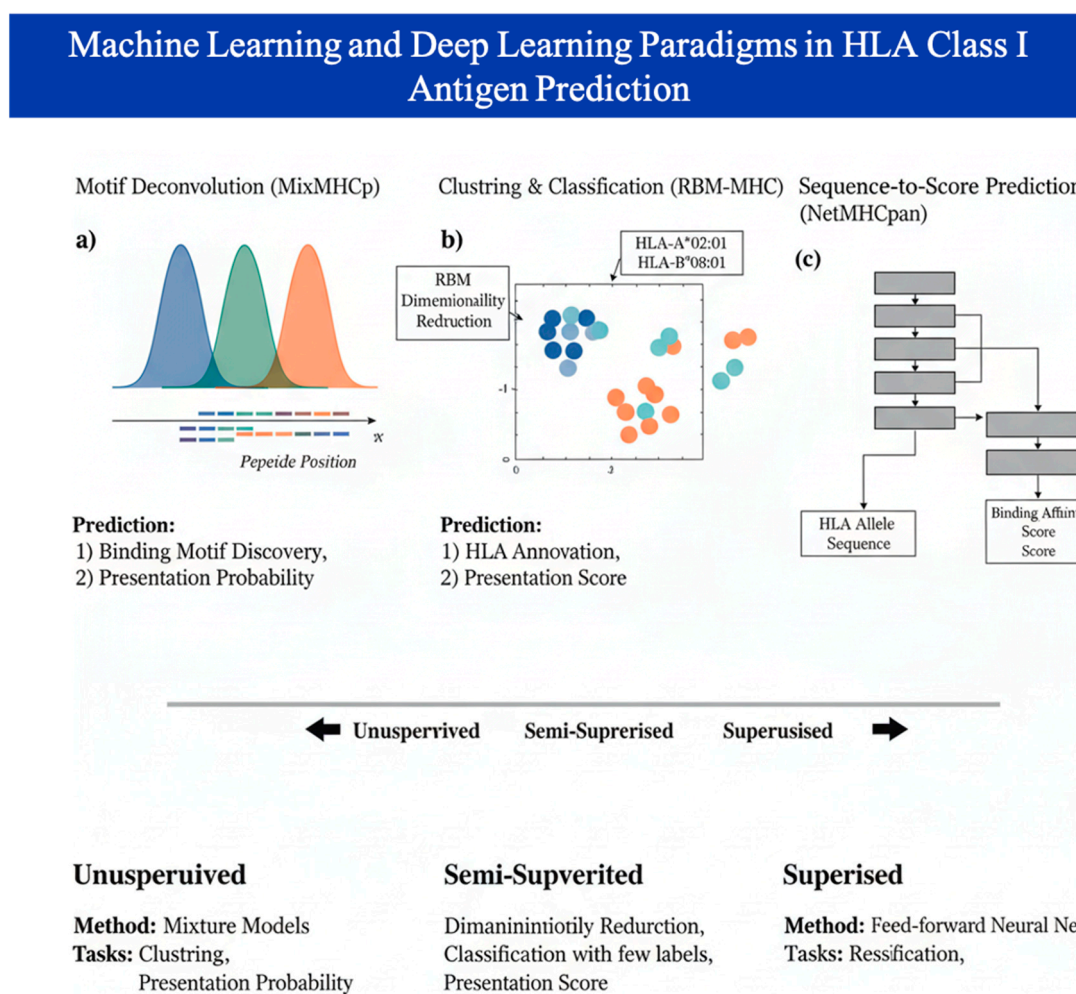


Figure 1. Machine Learning Paradigms for HLA Class I Antigen Presentation Prediction. This figure illustrates three distinct machine learning paradigms—Unsupervised, Semi-Supervised, and Supervised—used for

predicting the presentation of T-cell epitopes by Major Histocompatibility Complex (MHC) Class I molecules (HLA-I). **(A) Unsupervised (e.g., MixMHCp)**: Utilizes a mixture of probabilistic models to perform peptide clustering and binding motif deconvolution without relying on labeled affinity data **(B) Semi-Supervised (e.g., RBM-MHC)**: Employs dimensionality reduction (via a Restricted Boltzmann Machine, RBM) to classify peptides by HLA type, leveraging a small amount of labeled data to enhance prediction accuracy **(C) Supervised (e.g., NetMHCpan)**: Based on a fully trained feed-forward neural network that uses large datasets of known antigen and HLA sequences to predict peptide-HLA binding affinity and peptide elution scores.

2.1. Structural Vaccinology: From Sequence Motifs to 3D Graphs

The transition from handcrafted features to automated discovery marks a significant leap in the field's maturity. Before deep learning became the standard, traditional ML models like Support Vector Machines (SVMs) [18], and Random Forests [19] relied on meticulous feature selection and engineering—a process where researchers manually created variables to describe residue properties. These features typically fell into three categories: physicochemical properties (hydrophobicity, charge), geometric properties (solvent accessibility, surface curvature), and evolutionary information (residue conservation). However, these models were fundamentally limited by human intuition.

The advent of GNNs has revolutionized structural vaccinology by representing antigens as 3D graphs where amino acids are nodes and their spatial interactions are edges [20]. This allows for the capture of deep structural signatures that sequence-only models miss. For example, the GraphBepi and EpiGraph models utilize GNNs to calculate conformational B-cell epitopes with unprecedented precision [21]. Complementing this is the rise of Protein Language Models (PLMs) like ESM-2 and ProtBERT, which treat protein sequences as a language to learn rich, context-aware embeddings that reflect 3D proximity. Utilizing these embeddings in models such as BepiPred-3.0 has boosted performance metrics, achieving AUROC scores that allow for reliable zero-shot predictions of variant escape.

2.2. T-Cell Vaccinology: Decoding HLA Presentation

The T-cell immune response is entirely dependent on antigen presentation, where short linear peptides are displayed on the cell surface by Human Leukocyte Antigen (HLA) complexes [22]. This presentation step acts as a critical filter, and accurately predicting it is essential for narrowing down millions of potential targets [23]. For HLA Class I presentation, models are typically categorized into three paradigms: Unsupervised models like MixMHCp which use clustering to discover binding motifs from unlabeled mass spectrometry data [24]. Supervised feed-forward neural networks such as the NetMHCpan and MHCflurry [25] suites; and Semi-Supervised models like RBM-MHC, which leverage dimensionality reduction to classify peptides by HLA type with limited data [26]. Figure 1 demonstrates ML Paradigms for HLA Class I antigen presentation prediction.

While HLA Class I prediction is highly advanced, HLA Class II presentation remains more challenging due to the greater length variability of peptides and diverse binding cores, [27]. Modern tools like NetMHCIIpan implement dynamic searches for the optimal binding core using Convolutional Neural Networks (CNNs) to achieve AUROC scores around 0.85 [28] The current trend is shifting toward integrating immunopetidomic data (from mass spectrometry), which captures the entire processing and presentation pathway rather than just the final binding affinity [29].

2.3. The Immuno-Oncology Frontier: AI-Driven Neoantigen Discovery

As the field of vaccinology expands into therapeutic applications, the identification of tumor-specific neoantigens has emerged as the Holy Grail of personalized medicine. Unlike traditional viral antigens, neoantigens arise from somatic mutations exclusive to a patient's tumor and are absent in healthy tissue, making them ideal targets that bypass central immune tolerance. Identifying these targets requires the integration of high-dimensional multi-omics data, including DNA sequencing for mutation calling and RNA-seq to confirm expression [30]. The most advanced pipelines, such as

NeoDisc, utilize deep learning to prioritize neoantigens by integrating transcriptomic and immunopeptidomic profiles into a single predictive suite [31]. This approach is vital for solid tumors, where driver mutations must be distinguished from passenger mutations. AI architectures are now being designed to model the TCR-pMHC triad [32], predicting the precise structural interaction between the T-cell receptor (TCR) and the peptide-MHC complex. Models like TABR-BERT and NetTCR-struct use bimodal attention networks to capture the specificity of unseen TCRs, enabling the design of personalized mRNA vaccines that can compell the immune system to recognize and attack a patient's specific tumor landscape [32–34].

2.4. Generative Design and the Future of Target Optimization

The ultimate goal in modern target discovery has moved from mere prediction to De Novo design. Leveraging tools like RFDiffusion, AlphaFold2, and ProteinMPNN [35], researchers can now generate synthetic scaffolds that present epitopes with higher stability than their natural counterparts. For instance, structural modeling identified the two proline (2P) mutations used to stabilize the SARS-CoV-2 spike protein in mRNA vaccines. In immuno-oncology, this generative capacity allows for the creation of multi-neoantigen constructs that maximize immunogenicity while minimizing the risk of autoimmunity [36]. By defining the desired immune outcome first—such as a specific antibody neutralization profile or cytotoxic T-cell activation—AI-driven platforms can reverse-engineer the exact therapeutic molecule required, ushering in the era of Inverse Vaccinology [37]. Figure 2 demonstrates the challenges of traditional methods with the enhanced capabilities offered by Machine Learning (ML) in T-cell and B-cell epitope prediction. Table 2 summarizes representative computational methods across the major epitope prediction tasks, focusing on the model architecture and documented performance metrics derived from high-quality external or blinded validation datasets.

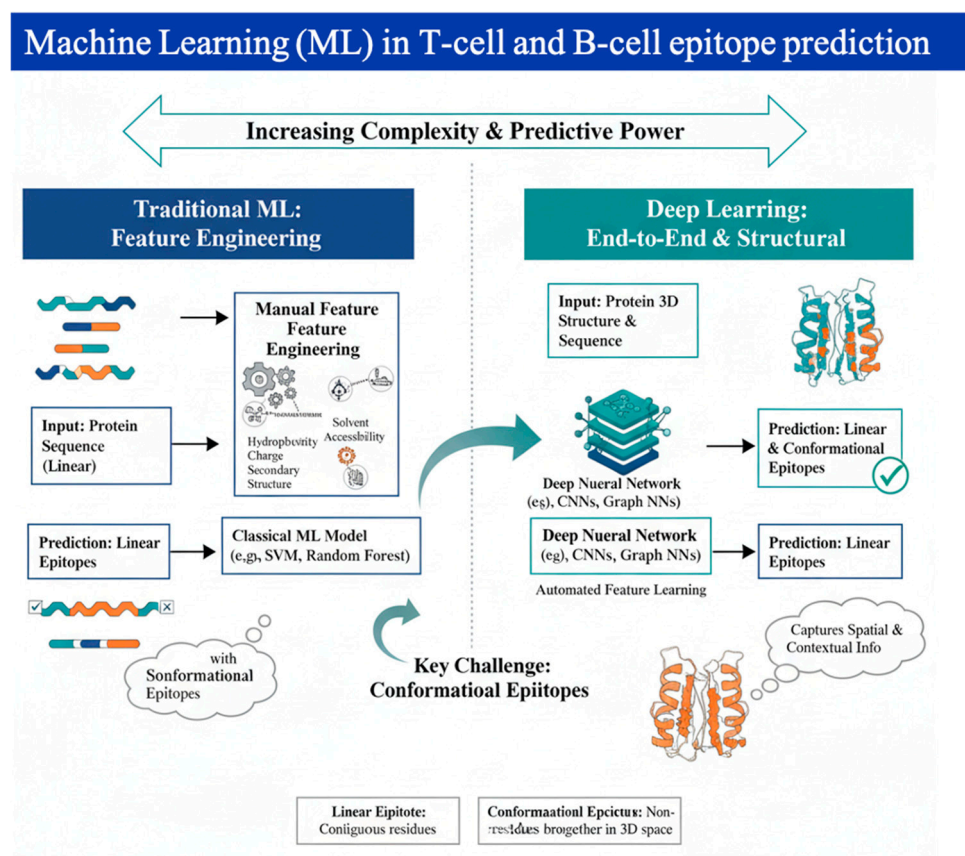


Figure 2. Evolutionary Shift in Epitope Discovery. The figure demonstrates the challenges of traditional methods with the enhanced capabilities offered by Machine Learning (ML) in T-cell and B-cell epitope

prediction. It highlights the crucial difference between the simpler task of predicting linear epitopes and the complex challenge of accurately identifying conformational (spatial) B-cell epitopes, which necessitate advanced structural modeling.

Table 2. Comparison of Leading Epitope Prediction Models by Type and Architecture.

Epitope Task	Model Example	Architecture Type	Key Features / Input	Peak Performance Metric	Reference Type	Reference
HLA Class I Presentation	NetMHCpan	Supervised Feed-Forward NN	Peptide + HLA Pseudosequence	AUROC ~ 0.96	Binding Affinity/Elution	[38,39]
HLA Class I Presentation	MixMHCpred	Unsupervised/Generative	Eluted Ligand Data Motifs	Performance Score (Motif Deconvolution)	Ligand Likelihood	[40]
HLA Class II Presentation	NetMHCIIpan	CNN-based Feed-Forward NN	Peptide Core Motif Search + HLA Sequence	AUROC ~ 0.85	Binding Affinity/Elution	[41]
Linear B-cell Epitope	BepiPred-2.0	Random Forest	Propensity Scales & Physicochemical Features	AUC ~ 0.75	Sequence Accessibility	[42]
Conformational B-cell Epitope	EpiGraph	Graph Neural Network (GNN)	3D Protein Graph Residue Contacts	AUC-PR ~ 0.24	Structural Proximity/Features	[43]
TCR-Epitope Specificity	TITAN	Bimodal Attention Network	Paired TCR CDR3 + Peptide Sequence	AUROC ~ 0.87 (Unseen TCRs)	Paired T-cell Specificity	[44]
Immunogenicity	Immunogenicity Predictor (e.g., from PMID: 106)	Supervised ML/Statistical	Amino Acid Enrichment (Immunogenic vs. Presented)	AUROC ~ 0.70	T-cell Activation	[45]

3. AI in Immuno-Pharmacology: Intelligent Delivery Systems and Adjuvant Optimization

The pharmacological efficacy of a vaccine is not solely determined by the antigen's sequence but is heavily dependent on the delivery-adjuvant axis. Even the most computationally perfect neoantigen or viral epitope will fail if it cannot reach the secondary lymphoid organs or if it lacks the necessary co-stimulatory signals to break immune tolerance. Traditional pharmacology relied on empirical formulation of known lipids and salts [46]. In contrast, AI-driven immuno-pharmacology uses predictive modeling to engineer the delivery vehicle alongside the cargo, ensuring that the vaccine's pharmacokinetic (PK) profile is optimized for maximum immunogenicity with minimal systemic toxicity [47].

3.1. Adjuvant Discovery: Breaking Tolerance with GNNs and Bayesian Optimization

Adjuvants are the accelerators of the immune response, yet their discovery has historically been a bottleneck due to the complex, non-linear interactions between chemical structures and innate immune receptors like Toll-like receptors (TLRs). Modern ML tools are now being used to navigate the vast chemical space of small-molecule adjuvants. VaxjoGNN, a specialized Graph Neural Network framework, treats adjuvant molecules as spatial graphs to predict their binding affinity to specific immune receptors [48]. By simulating these interactions in-silico, researchers can identify synergistic combinations—such as pairing a TLR4 agonist with a TLR7/8 ligand—that provoke a balanced Th1/Th2 response [49]. Furthermore, Bayesian Optimization is being deployed to conduct Active Learning loops, where the AI suggests a small set of adjuvant candidates for lab testing, learns from the experimental results, and refines its next prediction, reducing the search time for potent immunostimulants from years to months [50].

3.2. Lipid Nanoparticle (LNP) Engineering: Deep Learning for mRNA Delivery

The success of mRNA vaccines in both infectious diseases and immuno-oncology hinges on the Lipid Nanoparticle (LNP). Engineering an LNP involves optimizing a complex four-component system: ionizable lipids, helper lipids, cholesterol, and PEG-lipids. The environment of these formulations is highly sensitive; a minor change in the nitrogen-to-phosphate (N/P) ratio can lead to cargo degradation or liver toxicity [51]. To solve this, deep learning models like DeepLNP and LNP-Predict utilize multi-task neural networks trained on high-throughput screening data to predict LNP stability, cellular uptake, and endosomal escape efficiency. In the context of cancer vaccines, AI is specifically used to design organ-selective LNPs [52]. By analyzing the physicochemical properties of lipid libraries, ML models can predict which formulations will naturally accumulate in the spleen or lymph nodes rather than the liver, thereby concentrating the therapeutic effect where T-cell priming actually occurs [53].

The 2026 landscape has moved beyond simple multi-task networks to the use of AlphaFold and RoseTTAFold to simulate the physical 'docking' of the mRNA-LNP complex with the dendritic cell membrane. By modeling the atomic-scale interactions between ionizable lipids and the cellular lipid bilayer, researchers can now predict endosomal escape efficiency—the primary bottleneck in mRNA delivery—with over 90% accuracy, significantly reducing the empirical 'trial-and-error' phase of formulation.

3.3. Pharmacokinetics and Predictive Toxicology: Reducing the Attrition Rate

A significant portion of vaccine candidates fail in Phase I trials due to reactogenicity—excessive inflammation or off-target effects. AI-driven Quantitative Structure-Activity Relationship (QSAR) models are now being integrated with physiological-based pharmacokinetic (PBPK) modeling to create In-silico Safety Profiles [54]. These models predict the biodistribution of the vaccine components, identifying potential accumulation in non-target organs like the heart or brain before animal studies even begin [55]. For personalized oncology vaccines, this predictive pharmacology is crucial, as the AI must ensure that the patient-specific neoantigen delivery system does not trigger a cytokine storm or exacerbate existing autoimmune conditions [56].

3.4. The Convergence: Formulation-as-a-Service (FaaS)

The ultimate vision for AI in vaccine pharmacology is the transition toward Formulation-as-a-Service [57]. In this model, the AI not only designs the antigen; it outputs a map for complete candidate vaccine that includes the optimal mRNA sequence, the specific LNP composition for that patient's HLA type, and the predicted adjuvant synergy [58]. This holistic approach ensures that the vaccine is not just a biological product, but a precision-engineered pharmacological system. This is particularly transformative for mRNA-based immuno-oncology, where the speed of formulation is a matter of survival for patients with late-stage tumors [59]. By automating the pharmacology, AI allows for the rapid transition from biopsy to bedside, collapsing the formulation phase from months to a matter of days [60].

4. Deep Learning Paradigms in Epitope Discovery

The integration of advanced Deep Learning (DL) architectures marks the latest leap in epitope prediction, moving beyond traditional feature-based methods by automatically extracting complex, hidden patterns from biological sequences and structures. Convolutional Neural Networks (CNNs) excel at identifying local sequence motifs. They have been effectively applied to both T-cell and B-cell epitope prediction, with models like DeepImmuno-CNN explicitly integrating HLA context to improve precision and recall across diverse datasets, including neoantigens [61]. For B-cell epitopes, CNN-based models, often combined with attention mechanisms and Long Short-Term Memory (LSTM)s(e.g., NetBCE), substantially outperform classic tools, achieving ROC AUC values around 0.85 [62]. This ability to detect short, conserved patterns makes CNNs highly effective for finding

linear epitopes and MHC-binding cores. In contrast, Recurrent Neural Networks (RNNs), particularly LSTM units, are essential for handling the long-range dependencies inherent in biological sequences, mitigating the vanishing gradient problem common in standard RNNs [63]. LSTM-based predictors like MHCnuggets have drastically improved peptide-MHC affinity prediction, validated by mass spectrometry, [64] while hybrid Attention-BiLSTM-CNN models achieve state-of-the-art accuracy (AUC approximately 0.974) in predicting specific T-cell Receptor (TCR)–epitope specificity [65]. Modern systems often merge RNNs with other paradigms; for instance, GraphBepi couples a BiLSTM sequence encoder with a GNN to leverage both sequence context and 3D structural information [66]. Figure 3 illustrates the shift from scanning pathogens to a target-driven design workflow. The figure demonstrates the concept of Inverse Vaccinology.

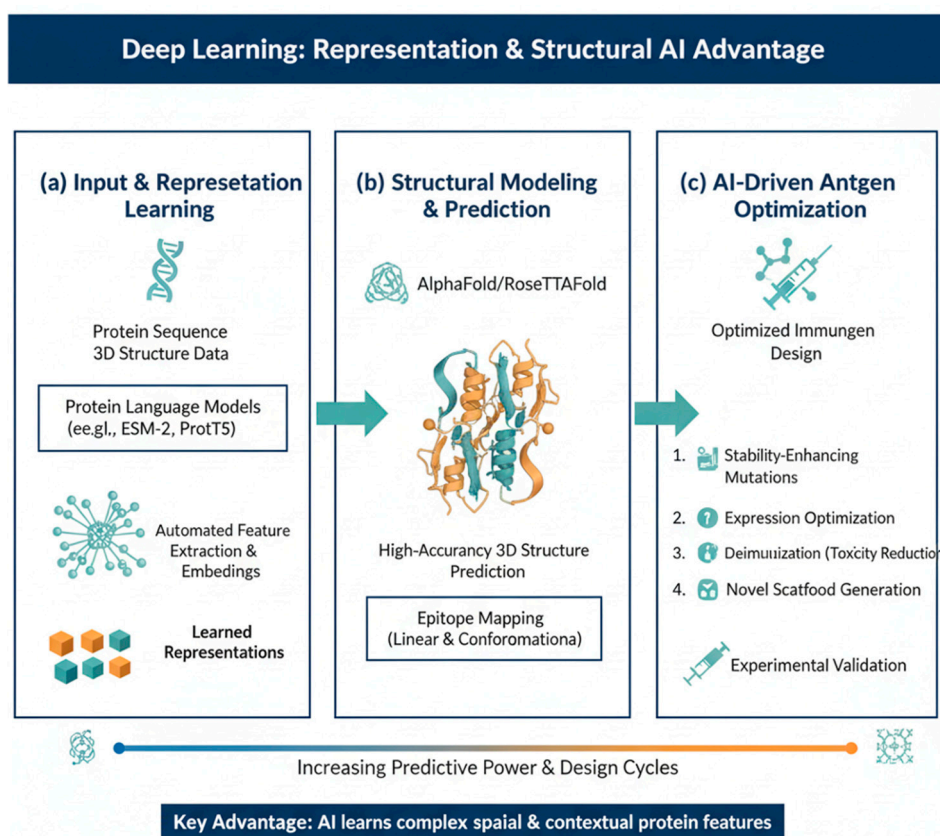


Figure 3. The Generative Inverse Vaccinology Pipeline. This figure illustrates the shift from scanning pathogens to a target-driven design workflow. (A) Target Specification: Defining the optimal immune synapse and TCR-pMHC interaction requirements. (B) Generative de novo synthesis: Utilizing diffusion-based models (e.g., RFDiffusion) to create de novo protein scaffolds that stabilize the target epitope. (C) In-Silico Validation: Testing the scaffold against Digital Twins of the human immune system to predict reactogenicity and memory cell differentiation before clinical synthesis.

The most disruptive architectural shift comes from Attention-Based Models, namely Transformers, which have established the new state-of-the-art for both B-cell and T-cell targets [67]. Models like BERTMHC and those leveraging the ESM (Evolutionary Scale Modeling) family use self-attention mechanisms to globally assess antigen sequences, identifying critical residues regardless of their physical distance in the sequence [68,69]. This context-aware representation improves on earlier methods, leading to higher accuracy (e.g., an AUC approx. 0.882 for MHC Class II prediction with BERTMHC) and better generalization [70,71]. Crucially, Transformers offer high interpretability, with attention weights often highlighting known immunodominant regions like the SARS-CoV-2 spike protein's receptor-binding domain. This permits researchers to move from modest prediction toward mechanistic insight. Furthermore, GNNs are fundamentally reshaping structure-based

prediction. GNN models like GraphBepi and EpiGraph operate on 3D protein graphs to learn complex spatial patterns, consistently outperforming traditional geometric predictors like DiscoTope by significant margins [72–74]. This structural modeling is also applied to T-cell targets, with tools like GraphMHC simulating MHC–peptide complexes as 3D atomic graphs to achieve high accuracy (0.92) in binding prediction [75]. These advancements, particularly the fusion of GNNs with attention and sequence encoders, provide both high predictive performance and clear rationales for antigen selection [76,77].

4.1. Navigating the Challenges of Translational Science

While the move to graph-based ML offers unprecedented power, it introduces new challenges that must be addressed to ensure these predictions translate into real-world vaccines. Unlike traditional feature-based methods, the design of GNNs requires careful consideration of what defines the graph—for instance, should it be based on the atomic level or the residue level—and how to effectively embed the structural and contextual information [78]. Furthermore, there is a risk that models might become prone to bias, over-relying on superficial properties (like simple surface exposure) associated with the functional behavior of an epitope, rather than truly learning the deep immunological mechanism [79]. Even the fundamental question of which features are best to combine with ML algorithms to achieve superior accuracy remains an open debate, lacking clear, universally accepted guidelines [80]. Therefore, while the tools are more powerful than ever, the next frontier requires researchers to move beyond systematic screening and focus on establishing robust, community-wide best practices for feature integration, systematic validation, and ultimately, translating these elegant computational predictions into reliable, protective vaccine candidates [81,82].

At its core, the aim of predicting B cell epitopes—the molecular triggers for antibody production—is to train ML models in a supervised manner to differentiate between true antibody binding sites and the many generic regions of an antigen that are ignored by the immune system [83,84]. The model assigns an epitope likelihood score to each site, allowing researchers to prioritize candidates. Since most B cell epitopes are conformational (meaning they rely on the protein's 3D folded shape, not just a continuous sequence), prediction accuracy is significantly higher when using models trained on full protein structures. These structure-based methods can leverage information about the antigen's surface topology and spatial arrangement, providing far more detail than sequence alone [85,86].

4.2. The Evolution of Feature-Based and Graph-Based Models

Early ML for epitope discovery centered on feature-based learning, requiring a preliminary, laborious step of manually engineering and selecting features. The underlying biological intuition was that only a few key sequence and structural properties dictate whether a residue is a high-affinity antibody binding site. Researchers hypothesized that exposed, flexible, and biochemically distinct residues were ideal targets [87]. To quantify this, the features fed into ML models typically included: physicochemical attributes (e.g., hydrophobicity and electrostatic charge, high-level geometric properties (e.g., secondary structure, surface curvature, and solvent accessibility, evolutionary data e.g., residue conservation, and specific amino acid combinations [88]. This dimensionality reduction, from thousands of atomic coordinates to a focused set of features, significantly improved computational efficiency [89].

The field has since moved toward more sophisticated techniques like Graph-based representations, which model the complex 3D relationships between residues as a network. This approach, adapted from successful applications in protein design and interaction site identification [90–92], is ideal for epitopes because they possess unique signatures related to residue packing and bond arrangements that are elegantly summarized by a graph [93]. However, constructing these graphs presents a challenge: determining the appropriate scale (atom vs. residue) and deciding which geometric information should be embedded into the graph connections remains complex. This

reliance on *a priori* choices in feature and graph design can inadvertently introduce bias—for example, favoring protrusive over flatter epitope surfaces—highlighting an ongoing challenge in optimally combining features for accurate prediction [94].

5. AI-Driven Clinical Optimization: From In-Silico Simulation to Personalized Trials

The clinical development phase remains the most significant hurdle in the vaccine and immuno-oncology (IO) lifecycle, traditionally accounting for nearly 75% of total R&D costs and serving as the primary site of candidate attrition [104]. As we enter the era of precision medicine, AI is fundamentally restructuring this phase by transitioning from rigid, large-scale empirical studies to agile, data-driven intelligent trials. By integrating multi-modal data—ranging from electronic health records (EHRs) to high-resolution immunopeptidomics—AI allows drug developers to simulate biological outcomes before a single patient is dosed, ensuring that only the most viable candidates progress to human testing [105]. Table 3 summarizes representative AI/ML methodologies applied across the vaccine and immuno-oncology development pipeline. Applications span antigen and epitope discovery, multi-epitope and neoantigen design, clinical trial optimization, and downstream manufacturing and distribution. While many approaches were initially developed for prophylactic vaccines, their extension to immuno-oncology highlights shared computational challenges, including immune heterogeneity, data sparsity, and translational uncertainty.

Table 3. AI Methodologies Across the Vaccine and Immuno-oncology Development Pipeline.

AI Approach	Application Stage	Core Function	Advantages	Limitations/Challenges	Representative Studies
EpiBERTope (Transformer-based)	Antigen & Epitope Prediction (Vaccines & Tumor Antigens)	Predicts linear and conformational B-cell epitopes from pathogen- or tumor-derived antigens.	Captures long-range sequence dependencies; adaptable to structurally complex antigens.	Requires large, high-quality labeled datasets; limited direct validation in tumor antigens. Risk of overfitting in	[95]
Ensemble ML (Vaxign-ML)	Antigen & Epitope Prioritization	Integrates antigenicity, host-pathogen, or tumor-specific features to prioritize vaccine or neoantigen candidates.	Robust to noisy inputs; flexible integration of heterogeneous biological features.	small neoantigen datasets; performance depends on feature engineering quality.	[96]
NetMHCpan (MHC Binding Predictor)	T-cell Epitope & Neoantigen Prediction	Predicts peptide binding affinity to MHC class I and II molecules for infectious or tumor-derived peptides.	Widely validated; foundational for both prophylactic vaccines and personalized cancer vaccines.	Reduced accuracy for rare HLA alleles; binding does not guarantee T-cell immunogenicity.	[97]
Vaxijen	Antigen Prediction	Identifies protective antigens or tumor-associated antigens without sequence alignment.	Rapid screening; alignment-free and computationally efficient.	Limited performance for multi-domain proteins and highly heterogeneous tumor antigens.	[98]
IntegralVac (Machine Learning)	Multi-epitope Vaccine & Neoantigen Construct Design	Designs multivalent constructs integrating antigenicity, immunogenicity, and allergenicity features.	Supports rational assembly of CD4+, CD8+, and B-cell epitopes; applicable to cancer vaccines.	Generalizability limited by epitope coverage and experimental validation availability.	[99]
Causal Inference Models	CoPs & Immune Response Modeling	Identifies correlates of protection or response from complex vaccine or immunotherapy trial datasets.	Addresses confounding and bias; supports uncertainty-aware inference in heterogeneous populations.	Requires rigorous trial design and expert statistical interpretation; sensitive to missing data.	[100]

Predictive Analytics / Regression	Clinical Trial Optimization (Vaccines & IO)	Optimizes patient recruitment, site selection, and enrollment forecasting for vaccine and immunotherapy trials.	Reduces trial timelines and cost; enables stratification by biomarker or immune phenotype.	Dependent on access to harmonized EHR and biomarker data; regulatory and privacy constraints.	[101]
Deep Learning	Manufacturing, Logistics & Supply Chain	Predicts demand and optimizes production scheduling for vaccines and cell- or RNA-based immunotherapies.	Integrates epidemiological, clinical, and operational signals for improved forecasting.	Vulnerable to unmodeled shocks; limited by historical representativeness.	[102]
IoT & Real-time Monitoring	Cold Chain & Advanced Therapy Logistics	Monitors storage and transport conditions for vaccines and temperature-sensitive immunotherapies.	Preserves product integrity; supports compliance for complex biologics and personalized therapies.	High infrastructure cost; cybersecurity and interoperability challenges.	[103]

5.1. Digital Twins and In-Silico Immune Modeling

The most transformative application for clinicians and biotech companies is the deployment of Digital Twins—computational duplicates of individual patients' immune systems [106,107]. Utilizing high-fidelity mechanistic models and Bayesian Neural Networks, platforms such as Novartis's Jinkō or GNS Healthcare's Gemini can simulate the interaction between a vaccine candidate and a specific patient's T-cell repertoire [108,109]. This capability allows for In-Silico Dose Escalation, where AI predicts the optimal therapeutic window for a neoantigen vaccine, potentially bypassing months of traditional Phase I dose-finding studies. For immuno-oncology, these models are critical for predicting the efficacy of combination therapies—such as pairing a personalized mRNA vaccine with a PD-1 inhibitor—by simulating the complex, non-linear dynamics of the tumor microenvironment (TME) [110, [111].

5.2. Adaptive Trial Designs and Synthetic Control Arms (SCA)

To address the ethical and logistical challenges of recruiting placebo groups in life-threatening conditions, AI-generated Synthetic Control Arms (SCA) are becoming a regulatory reality [112]. SCAs leverage historical trial data and real-world evidence (RWE) to create a virtual cohort that mirrors the treatment group's baseline characteristics, thereby reducing the number of human subjects required by up to 30% without sacrificing statistical power [113]. Furthermore, Bayesian Adaptive Designs allow for real-time adjustments to trial parameters, such as sample size re-estimation or the early termination of futile arms based on emerging immunological correlates of protection (CoPs). This fail fast approach is essential for biotech companies to manage risk and reallocate capital toward successful candidates [114].

CoPs are instrumental in enabling computational analyses that facilitate *in silico* clinical trials, allowing developers to address critical questions such as the extrapolation of efficacy data from animal models to human populations. A common application, immunobridging, utilizes established CoPs to predict the efficacy of an existing vaccine against a known pathogen, or to estimate the performance of a novel candidate in a different host population using a distinct CoP [115–117]. While initial analyses were often performed using ad-hoc statistical techniques [118], recent theoretical advances in causal inference have introduced robust, generic frameworks. These new methodologies can be systematically applied to identify and precisely estimate CoPs, even when confronted with significant sources of uncertainty, including unobserved confounding, sample selection bias, external validity issues, missing data, measurement error, and inter-individual variability [119]. The evolution of these methods, which can establish tighter confidence bounds on vaccine efficacy estimates despite various uncertainties [120], is poised to revolutionize the design of both novel vaccines and their associated clinical trials. Furthermore, the adoption of rigorous statistical methods for assessing CoPs from randomized, controlled efficacy trials—underscoring the necessity of meticulous experimental

design, pre-registration, and standardized statistical analysis plans—will significantly enhance the reliability of predictive data [121].

5.3. Biomarker Intelligence and Patient Stratification

The success of therapeutic vaccines in oncology hinges on identifying the right patient for the right antigen. AI algorithms are proficient at analyzing vast, unstructured datasets—including spatial transcriptomics and digital pathology—to identify complex multiplex biomarkers that predict response [122]. Unlike single-marker tests (e.g., PD-L1 expression), AI-driven stratification models like DeepPatient or MHCnuggets integrate a patient's mutational burden, HLA genotype and T-cell exhaustion markers to calculate a responder score [123,124]. For clinicians, this means a shift toward personalized enrollment, ensuring that patients most likely to benefit are prioritized, which has been shown to increase the trial enrollment efficiency [125].

5.4. Regulatory Landscape: The FDA Modernization Act 2.0

The integration of AI into clinical workflows is further supported by the FDA Modernization Act 2.0, which allows for the replacement of traditional animal models with human-relevant AI/ML-based approaches and organoids. This regulatory shift encourages biotech companies to adopt In-Silico Safety Assessments, utilizing deep learning to predict cardiotoxicity or cytokine release syndrome (CRS) with higher accuracy than murine models. By aligning computational evidence with clinical outcomes, AI is not just a tool for acceleration but an important requirement for the next-generation of safe, effective, and ethically sound precision therapies. Table 4 summarizes the AI Frameworks for Clinical Trial Transformation in vaccines and IO. Figure 4 demonstrates comprehensive schematic illustrates the heterogeneous yet interconnected applications of AI across three major vaccine development modalities: Infectious Diseases, Cancer, and Pandemic Preparedness.

Table 4. AI Frameworks for Clinical Trial Transformation in Vaccines and IO.

Optimization Strategy	Key AI/ML Architecture	Translatable Clinical/Industrial Benefit	Technical Impact / Metric
In-Silico Trials (Digital Twins)	Mechanistic Modeling + Bayesian NNs	Dose Optimization: Predicts individual safety/efficacy profiles before Phase I.	Reduces dose-finding duration by ~50%.
Synthetic Control Arms (SCA)	Generative Adversarial Networks (GANs)	Ethical Compliance: Replaces placebo arms in rare disease or oncology trials.	Reduces required human subjects by 20–30%.
Patient Stratification	Transformer-based Multi-modal Fusion	Responder Selection: Identifies high-likelihood responders via multi-omic biomarkers.	Improves Phase II/III success rates by 15%.
Site Selection & Recruitment	Geospatial AI + NLP (EHR Mining)	Enrollment Speed: Finds optimal global sites based on pathogen prevalence or tumor type.	Accelerates recruitment timelines by 3–6 months.
Adaptive Bayesian Design	Reinforcement Learning (RL)	Dynamic Optimization: Real-time trial adjustments (dosage/sample size).	Minimizes capital loss via early-exit of failing arms.
Regulatory Compliance	Explainable AI (XAI) / SHAP Values	FDA/EMA Trust: Provides white-box reasoning for AI-driven clinical decisions.	Standardizes AI-supported regulatory submissions.

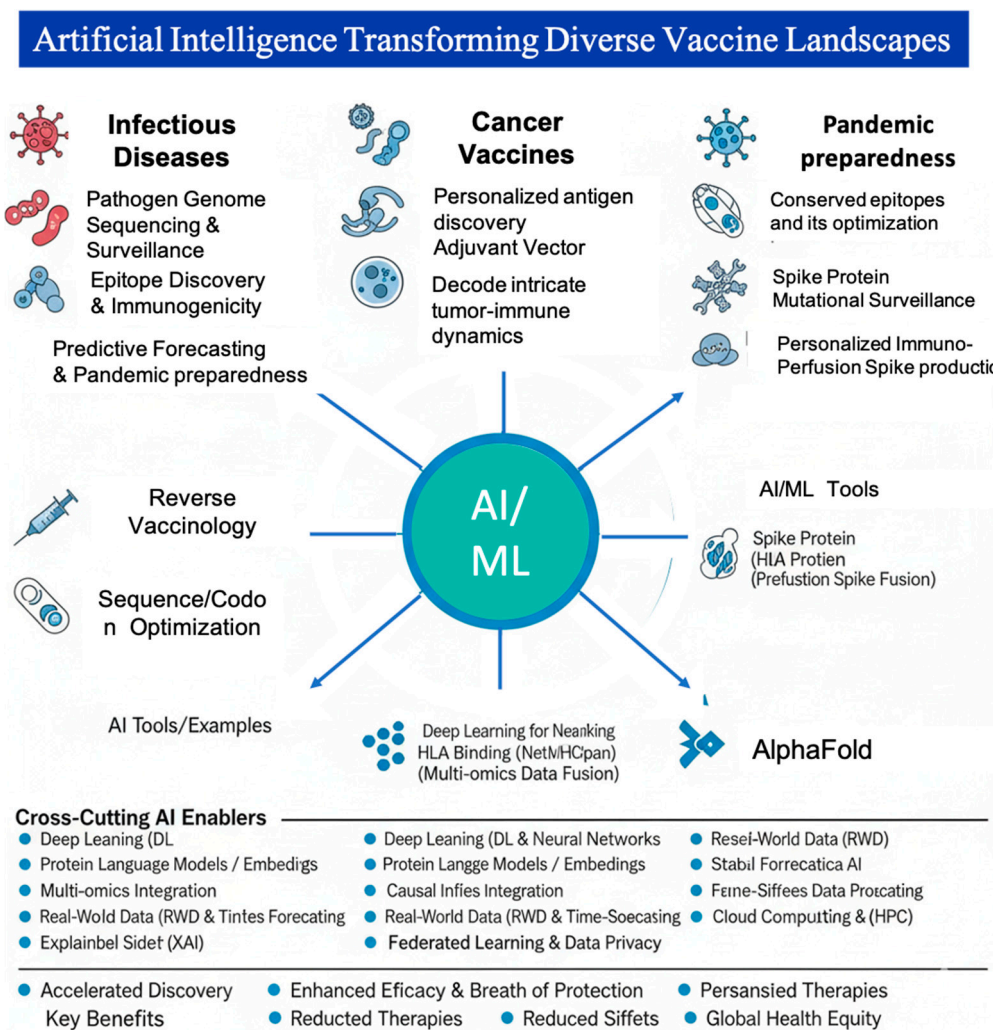


Figure 4. AI Integration Across Diverse Vaccine Development Modalities. The figure highlights the specific computational contributions of AI tailored to the unique challenges of each field, demonstrating its versatility as a cross-cutting technology (A) **Infectious Diseases:** AI accelerates the identification of key pathogen targets, predicts epidemiological patterns for strain prioritization (e.g., in annual influenza updates), and optimizes vaccine formulation components like adjuvants for enhanced safety and efficacy. (B) **Personalized oncology vaccines:** Here, AI is essential for managing patient-specific tumor heterogeneity. It leverages multi-omics data (genomics, transcriptomics) to accurately predict patient-specific tumor neoantigens, model T-Cell Receptor (TCR) binding, and select personalized targets for therapeutic vaccine design.

6. Future Directions, Ethical Considerations, and Global Deployment

As we move into 2026, the integration of AI into vaccinology and immuno-oncology (IO) is transitioning from a proof-of-concept phase to a integration phase. The future of the field is no longer defined by whether AI can predict an epitope, but by whether those predictions can be translated into global health equity and regulatory-approved therapies. This final phase focuses on bridging the gap between high-tech computational discovery and real-world clinical application, particularly through the lens of explainability and global accessibility.

6.1. Explainable AI (XAI) and the Regulatory White-Box Paradigm

The primary hurdle for FDA and EMA approval of AI-designed vaccines is the black-box nature of deep learning [126]. To achieve regulatory trust, researchers are adopting Explainable AI (XAI) scaffolds like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) [127]. These tools decompose complex neural network predictions into quantifiable

feature contributions, allowing clinicians to see why a specific neoantigen was prioritized—for instance, by highlighting the specific amino acid residues contributing to MHC-binding affinity [128, [129]. In Jan 2026, the FDA released updated Guiding Principles for AI in Drug Development, which explicitly emphasize Transparency and Lifecycle Monitoring as requirements for submissions. For biotech companies, implementing XAI is no longer an optional research add-on but a critical requirement for a successful IND application [130]

6.2. Global Health Equity: Frugal AI and Decentralized Manufacturing

One of the most profound impacts of AI in vaccinology is its potential to democratize access to life-saving preventatives. While mRNA platforms are highly effective, their dependence on a cold-chain (-80°C to -20°C) has historically limited their reach in low-resource settings [131]. Emerging Systems Biology-guided AI (SBg-AI) frameworks are now being used to design thermostable formulations by predicting the degradation trajectories of lipid nanoparticles (LNPs) under varying environmental conditions [132]. By optimizing lyophilization cycles (freeze-drying) through AI-driven digital twins, researchers can reduce mRNA wastage by up to 22% in Southeast Asian and African supply chains [133]. Furthermore, AI-driven localized manufacturing units—portable, micro-factory containers—utilize real-time AI monitoring to adjust manufacturing parameters on-the-fly, ensuring that high-quality personalized vaccines can be produced at the point of care in remote regions [134].

6.3. The Rise of Agentic AI and Inverse Vaccinology

Looking toward 2027, the field is preparing for the advent of Agentic AI autonomous systems capable of not just predicting data, but planning and executing multi-step laboratory workflows. This will fully realize the concept of Inverse Vaccinology, where the desired immune outcome (e.g., a specific neutralizing antibody titer against a broad range of coronavirus variants) is used as the starting prompt, and the AI autonomously designs the antigen, the delivery vehicle, and the clinical trial protocol. This outcome-first approach is expected to reduce the design-to-dose timeline for emerging pathogens from months to less than 30 days [135,136].

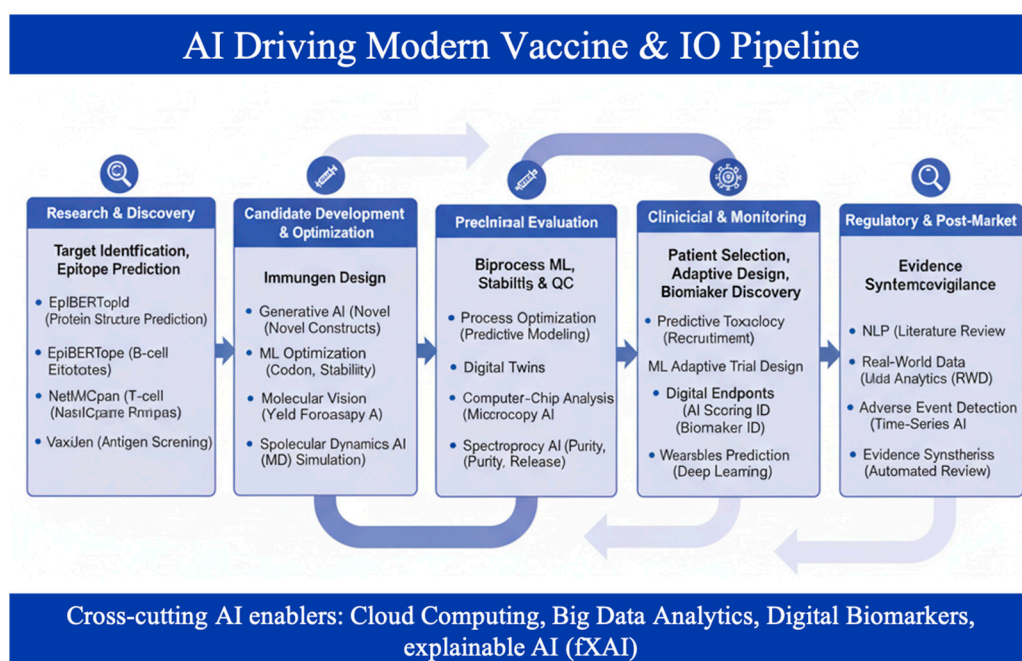


Figure 5. A schematic illustration of the intended use of AI/ML models across vaccine research, development, CMC manufacturing, preclinical evaluation, clinical evaluation and regulatory approval for AI assisted vaccine product development.

7. Limitations and Failure Modes of AI-Driven Vaccinology

Despite rapid methodological advances, the application of AI and ML to vaccinology and immuno-oncology is subject to fundamental limitations that constrain predictive reliability and clinical translation. Recognizing these failure modes is essential to avoid over-interpretation of in-silico results and to guide responsible deployment in vaccine development [137].

7.1. Data Bias and Incomplete Immunological Representation

Most AI models in vaccinology are trained on datasets that are unevenly distributed across pathogens, HLA alleles, and patient populations. Immunopeptidomic and epitope validation data are disproportionately derived from individuals of European ancestry and from a limited set of well-studied viruses and tumor types [138]. As a result the model presentation may degrade significantly when applied to underrepresented HLA haplotypes, rare pathogens, or genetically diverse populations, potentially exacerbating health inequities rather than alleviating them [139].

7.2. Prediction Does Not Guarantee Immunogenicity

Accurate prediction of peptide–HLA binding or structural epitope exposure does not ensure functional immunogenicity. T-cell activation and antibody maturation are emergent biological processes influenced by antigen processing, immune context, tolerance mechanisms, and host history. AI models may prioritize epitopes that are technically presented but biologically ignored, an issue that is particularly pronounced in cancer neoantigen prediction where immune editing and exhaustion dominate clinical outcomes [140].

7.3. Overfitting and Benchmark Inflation

Many reported performance metrics for epitope prediction and immunogenicity models are derived from retrospective or partially overlapping datasets. Without strict separation of training and evaluation data—or validation in independent experimental systems—models may overfit to dataset-specific artifacts rather than generalizable immunological principles. Benchmark inflation poses a risk of overstating readiness for translational or clinical application [141].

7.4. Structural Uncertainty and Model Assumptions

Structure-based models, including GNNs and diffusion-generated protein scaffolds, rely on predicted or static protein conformations that may not reflect dynamic immune synapse interactions in vivo. Errors in upstream structure prediction propagate through downstream epitope or immunogen design workflows, potentially leading to misleading confidence in structurally optimized candidates [142].

7.5. Delivery and Context Dependence

Even optimally designed antigens can fail clinically due to suboptimal delivery, biodistribution, or innate immune activation. While AI-guided lipid nanoparticle and adjuvant optimization is advancing, current models often simplify complex pharmacokinetic and immunostimulatory processes. As a result, predictions of endosomal escape, tissue targeting, or reactogenicity may not generalize across species or clinical settings [143].

7.6. Regulatory and Interpretability Constraints

Deep learning models frequently operate as black boxes, limiting their acceptability in regulatory submissions [144]. Although explainable AI methods are emerging, many remain descriptive rather than mechanistic, and their outputs may not satisfy regulatory expectations for causal understanding or risk assessment. Without transparent model governance and lifecycle monitoring, AI-driven vaccinology may face delays in regulatory adoption [145].

8. Conclusion: The AI-Driven Era of Precision Immunology

Artificial intelligence and machine learning have reached a level of technical maturity that allows them to contribute meaningfully across multiple stages of vaccinology and immuno-oncology, from antigen discovery to formulation optimization and early clinical strategy [146]. The most immediate value of these approaches lies not in replacing experimental immunology, but in narrowing the design space, prioritizing biologically plausible hypotheses, and reducing inefficiencies inherent in empirical vaccine development [147].

This review has outlined the emergence of inverse vaccinology as a conceptual extension of reverse vaccinology, enabled by generative deep learning architectures capable of designing immunogens de novo in response to predefined immunological objectives. Evidence from epitope modeling, neoantigen prioritization, and structure-informed protein design suggests that such approaches can improve early-stage decision-making, particularly in settings characterized by antigenic heterogeneity or patient specificity. However, the translation of computationally optimized candidates into durable and clinically meaningful immune responses remains contingent on biological context, delivery mechanisms, and host immune state.

Importantly, current limitations—including biased training data, incomplete representations of immune complexity, structural uncertainty, and challenges in model interpretability—constrain the predictive reliability of AI-driven vaccinology. Overestimation of model performance or clinical readiness risks undermining confidence in computational approaches and delaying their responsible adoption. As such, rigorous validation, transparent reporting of assumptions, and integration with orthogonal experimental data are essential to ensure that AI functions as a decision-support tool rather than a source of false precision [148]. Looking forward, progress in AI-enabled vaccinology will depend on tighter coupling between computational design and experimental feedback, improved representation of diverse immune populations, and harmonization with regulatory expectations. When embedded within multidisciplinary development pipelines, AI has the potential to accelerate iteration cycles and support more rational vaccine design strategies. Its long-term impact will be determined not by the sophistication of algorithms alone, but by their ability to reliably inform biological decision-making in complex, real-world clinical settings.

The convergence of AI/ML and synthetic biology has deeply altered the trajectory of both prophylactic vaccinology and therapeutic immuno-oncology [146]. We are moving away from the era of one-size-fits-all public health toward a future of Precision Immunology, where vaccines are as unique as the patients they treat [147]. For clinicians, AI provides the decision-support tools needed to navigate the complexity of the tumor microenvironment; for drug developers, it offers a fail-fast roadmap that reduces capital risk; and for the global community, it promises a more equitable and resilient defense against future pandemics [148]. As we refine the mathematical architectures of GNNs and Transformers, our goal remains singular: to ensure that the speed of our computational discovery is matched by the safety and accessibility of our biological solutions.

Author Contributions: V.G. contributed by conceptualizing and writing all the sections in this comprehensive review article.

Funding: This review article received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CNNs	Convolutional Neural Networks
RNNs	Recurrent Neural Networks
GNNs	Graph Neural Networks
CoP	Correlates of Protection
NLP	Natural Language Processing
VO	Vaccine Ontology
AUROC	Area Under the Receiver Operating Characteristic Curve
HLA	Human Leukocyte Antigen
MHC	Major Histocompatibility Complex

References:

1. Benn, C.S.; Fisker, A.B.; Rieckmann, A.; Sørup, S.; Aaby, P. Vaccinology: Time to Change the Paradigm? *Lancet Infect. Dis.* **2020**, *20*, e274–e283, doi:10.1016/S1473-3099(19)30742-X.
2. Light, D.W.; Lexchin, J. The Costs of Coronavirus Vaccines and Their Pricing. *J. R. Soc. Med.* **2021**, *114*, 502–504, doi:10.1177/014107682111053006.
3. Villanueva-Flores, F.; Sanchez-Villamil, J.I.; Garcia-Atutxa, I. AI-Driven Epitope Prediction: A System Review, Comparative Analysis, and Practical Guide for Vaccine Development. *NPJ Vaccines* **2025**, *10*, 207, doi:10.1038/s41541-025-01258-y.
4. Villanueva-Flores, F.; Sanchez-Villamil, J.I.; Garcia-Atutxa, I. AI-Driven Epitope Prediction: A System Review, Comparative Analysis, and Practical Guide for Vaccine Development. *NPJ Vaccines* **2025**, *10*, 207, doi:10.1038/s41541-025-01258-y.
5. Villanueva-Flores, F.; Sanchez-Villamil, J.I.; Garcia-Atutxa, I. AI-Driven Epitope Prediction: A System Review, Comparative Analysis, and Practical Guide for Vaccine Development. *NPJ Vaccines* **2025**, *10*, 207, doi:10.1038/s41541-025-01258-y.
6. Villanueva-Flores, F.; Sanchez-Villamil, J.I.; Garcia-Atutxa, I. Publisher Correction: AI-Driven Epitope Prediction: A Systematic Review, Comparative Analysis, and Practical Guide for Vaccine Development. *NPJ Vaccines* **2025**, *10*, 209, doi:10.1038/s41541-025-01273-z.
7. Bravi, B. Development and Use of Machine Learning Algorithms in Vaccine Target Selection. *NPJ Vaccines* **2024**, *9*, 15, doi:10.1038/s41541-023-00795-8.
8. da Silva, O.L.T.; da Silva, M.K.; Rodrigues-Neto, J.F.; Santos Lima, J.P.M.; Manzoni, V.; Akash, S.; Fulco, U.L.; Bourhia, M.; Dawoud, T.M.; Nafidi, H.-A.; et al. Advancing Molecular Modeling and Reverse Vaccinology in Broad-Spectrum Yellow Fever Virus Vaccine Development. *Sci. Rep.* **2024**, *14*, 10842, doi:10.1038/s41598-024-60680-9.
9. Niu, J.; Deng, R.; Dong, Z.; Yang, X.; Xing, Z.; Yu, Y.; Kang, J. Mapping the Landscape of AI and ML in Vaccine Innovation: A Bibliometric Study. *Hum. Vaccines Immunother.* **2025**, *21*, 2501358, doi:10.1080/21645515.2025.2501358.
10. Banday, A.H.; Manzoor, M.M.; Nissar, U.; Jaleel, S. From Neoantigen Discovery to Immune-Checkpoint Synergy: Peptide Cancer Vaccines as Precision Tools for Personalised Cancer Therapy. *Scand. J. Immunol.* **2026**, *103*, e70084, doi:10.1111/sji.70084.
11. Arunachalam, P.S.; Wimmers, F.; Mok, C.K.P.; Perera, R.A.P.M.; Scott, M.; Hagan, T.; Sigal, N.; Feng, Y.; Bristow, L.; Tak-Yin Tsang, O.; et al. Systems Biological Assessment of Immunity to Mild versus Severe COVID-19 Infection in Humans. *Science* **2020**, *369*, 1210–1220, doi:10.1126/science.abc6261.
12. Ciani, O.; Manyara, A.M.; Davies, P.; Stewart, D.; Weir, C.J.; Young, A.E.; Blazeby, J.; Butcher, N.J.; Bujkiewicz, S.; Chan, A.-W.; et al. A Framework for the Definition and Interpretation of the Use of Surrogate Endpoints in Interventional Trials. *EclinicalMedicine* **2023**, *65*, 102283, doi:10.1016/j.eclinm.2023.102283.
13. Rubinstein, N.D.; Mayrose, I.; Pupko, T. A Machine-Learning Approach for Predicting B-Cell Epitopes. *Mol. Immunol.* **2009**, *46*, 840–847, doi:10.1016/j.molimm.2008.09.009.

14. Jespersen, M.C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Res.* **2017**, *45*, W24–W29, doi:10.1093/nar/gkx346.
15. Saha, S.; Raghava, G.P.S. Prediction of Continuous B-Cell Epitopes in an Antigen Using Recurrent Neural Network. *Proteins* **2006**, *65*, 40–48, doi:10.1002/prot.21078.
16. da Silva, B.M.; Myung, Y.; Ascher, D.B.; Pires, D.E.V. epitope3D: A Machine Learning Method for Conformational B-Cell Epitope Prediction. *Brief. Bioinform.* **2022**, *23*, bbab423, doi:10.1093/bib/bbab423.
17. Jespersen, M.C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Res.* **2017**, *45*, W24–W29, doi:10.1093/nar/gkx346.
18. Saha, S.; Raghava, G.P.S. Prediction of Continuous B-Cell Epitopes in an Antigen Using Recurrent Neural Network. *Proteins* **2006**, *65*, 40–48, doi:10.1002/prot.21078.
19. Rubinstein, N.D.; Mayrose, I.; Pupko, T. A Machine-Learning Approach for Predicting B-Cell Epitopes. *Mol. Immunol.* **2009**, *46*, 840–847, doi:10.1016/j.molimm.2008.09.009.
20. Villanueva-Flores, F.; Sanchez-Villamil, J.I.; Garcia-Atutxa, I. AI-Driven Epitope Prediction: A System Review, Comparative Analysis, and Practical Guide for Vaccine Development. *NPJ Vaccines* **2025**, *10*, 207, doi:10.1038/s41541-025-01258-y.
21. Villanueva-Flores, F.; Sanchez-Villamil, J.I.; Garcia-Atutxa, I. AI-Driven Epitope Prediction: A System Review, Comparative Analysis, and Practical Guide for Vaccine Development. *NPJ Vaccines* **2025**, *10*, 207, doi:10.1038/s41541-025-01258-y.
22. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449–W454, doi:10.1093/nar/gkaa379.
23. Andreatta, M.; Nielsen, M. Gapped Sequence Alignment Using Artificial Neural Networks: Application to the MHC Class I System. *Bioinforma. Oxf. Engl.* **2016**, *32*, 511–517, doi:10.1093/bioinformatics/btv639.
24. O'Donnell, T.J.; Rubinsteyn, A.; Laserson, U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst.* **2020**, *11*, 42-48.e7, doi:10.1016/j.cels.2020.06.010.
25. O'Donnell, T.J.; Rubinsteyn, A.; Bonsack, M.; Riemer, A.B.; Laserson, U.; Hammerbacher, J. MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Syst.* **2018**, *7*, 129-132.e4, doi:10.1016/j.cels.2018.05.014.
26. Nielsen, M.; Lundegaard, C.; Blicher, T.; Lamberth, K.; Harndahl, M.; Justesen, S.; Røder, G.; Peters, B.; Sette, A.; Lund, O.; et al. NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence. *PLoS One* **2007**, *2*, e796, doi:10.1371/journal.pone.0000796.
27. Bassani-Sternberg, M.; Chong, C.; Guillaume, P.; Solleder, M.; Pak, H.; Gannon, P.O.; Kandalaft, L.E.; Coukos, G.; Gfeller, D. Deciphering HLA-I Motifs across HLA Peptidomes Improves Neo-Antigen Predictions and Identifies Allosteric Regulating HLA Specificity. *PLoS Comput. Biol.* **2017**, *13*, e1005725, doi:10.1371/journal.pcbi.1005725.
28. Vitiello, A.; Zanetti, M. Neoantigen Prediction and the Need for Validation. *Nat. Biotechnol.* **2017**, *35*, 815–817, doi:10.1038/nbt.3932.
29. Karosiene, E.; Rasmussen, M.; Blicher, T.; Lund, O.; Buus, S.; Nielsen, M. NetMHCIIpan-3.0, a Common Pan-Specific MHC Class II Prediction Method Including All Three Human MHC Class II Isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* **2013**, *65*, 711–724, doi:10.1007/s00251-013-0720-y.
30. Wirth, T.C.; Kühnel, F. Neoantigen Targeting—Dawn of a New Era in Cancer Immunotherapy? *Front. Immunol.* **2017**, *8*, 1848, doi:10.3389/fimmu.2017.01848.
31. Gumpert, N.; Sagie, S.; Arnedo-Pac, C.; Babu, T.; Weller, C.; Gonzalez-Perez, A.; Wang, Y.; Michel Todó, L.; Levy, R.; Chen, X.; et al. Recurrent Immunogenic Neoantigens and Their Cognate T-Cell Receptors in Treatment-Resistant Metastatic Prostate Cancer. *Cancer Discov.* **2025**, doi:10.1158/2159-8290.CD-24-1213.
32. Zhang, X.; Xu, Z.; Dai, X.; Zhang, X.; Wang, X. Research Progress of Neoantigen-Based Dendritic Cell Vaccines in Pancreatic Cancer. *Front. Immunol.* **2023**, *14*, 1104860, doi:10.3389/fimmu.2023.1104860.

33. Deleuran, S.N.; Nielsen, M. NetTCR-Struc, a Structure Driven Approach for Prediction of TCR-pMHC Interactions. *Front. Immunol.* **2025**, *16*, 1616328, doi:10.3389/fimmu.2025.1616328.
34. Hashempour, A.; Khodadad, N.; Akbarinia, S.; Ghasabi, F.; Ghasemi, Y.; Nazar, M.M.K.A.; Falahi, S. Reverse Vaccinology Approaches to Design a Potent Multiepitope Vaccine against the HIV Whole Genome: Immunoinformatic, Bioinformatics, and Molecular Dynamics Approaches. *BMC Infect. Dis.* **2024**, *24*, 873, doi:10.1186/s12879-024-09775-2.
35. Watson, J.L.; Juergens, D.; Bennett, N.R.; Trippe, B.L.; Yim, J.; Eisenach, H.E.; Ahern, W.; Borst, A.J.; Ragotte, R.J.; Milles, L.F.; et al. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620*, 1089–1100, doi:10.1038/s41586-023-06415-8.
36. Hajnik, R.L.; Plante, J.A.; Liang, Y.; Alameh, M.-G.; Tang, J.; Bonam, S.R.; Zhong, C.; Adam, A.; Scharton, D.; Rafael, G.H.; et al. Dual Spike and Nucleocapsid mRNA Vaccination Confer Protection against SARS-CoV-2 Omicron and Delta Variants in Preclinical Models. *Sci. Transl. Med.* **2022**, *14*, eabq1945, doi:10.1126/scitranslmed.abq1945.
37. Xiao, W.; Jiang, W.; Chen, Z.; Huang, Y.; Mao, J.; Zheng, W.; Hu, Y.; Shi, J. Advance in Peptide-Based Drug Development: Delivery Platforms, Therapeutics and Vaccines. *Signal Transduct. Target. Ther.* **2025**, *10*, 74, doi:10.1038/s41392-024-02107-5.
38. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449–W454, doi:10.1093/nar/gkaa379.
39. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449–W454, doi:10.1093/nar/gkaa379.
40. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449–W454, doi:10.1093/nar/gkaa379.
41. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449–W454, doi:10.1093/nar/gkaa379.
42. Clifford, J.N.; Høie, M.H.; Deleuran, S.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-3.0: Improved B-Cell Epitope Prediction Using Protein Language Models. *Protein Sci. Publ. Protein Soc.* **2022**, *31*, e4497, doi:10.1002/pro.4497.
43. Choi, S.; Kim, D. B Cell Epitope Prediction by Capturing Spatial Clustering Property of the Epitopes Using Graph Attention Network. *Sci. Rep.* **2024**, *14*, 27496, doi:10.1038/s41598-024-78506-z.
44. Weber, A.; Born, J.; Rodriguez Martínez, M. TITAN: T-Cell Receptor Specificity Prediction with Bimodal Attention Networks. *Bioinforma. Oxf. Engl.* **2021**, *37*, i237–i244, doi:10.1093/bioinformatics/btab294.
45. Armstrong, K.M.; Piepenbrink, K.H.; Baker, B.M. Conformational Changes and Flexibility in T-Cell Receptor Recognition of Peptide-MHC Complexes. *Biochem. J.* **2008**, *415*, 183–196, doi:10.1042/BJ20080850.
46. Shahin, M.H.; Desai, P.; Terranova, N.; Guan, Y.; Helikar, T.; Lobentanzer, S.; Liu, Q.; Lu, J.; Madhavan, S.; Mo, G.; et al. AI-Driven Applications in Clinical Pharmacology and Translational Science: Insights From the ASCPT 2024 AI Preconference. *Clin. Transl. Sci.* **2025**, *18*, e70203, doi:10.1111/cts.70203.
47. Gu, D.; Feng, Y.; Li, H. AI-Driven Immunotherapy: Synergizing with Radiotherapy to Reconfigure the Tumor Microenvironment and Treatment Landscape. *Front. Pharmacol.* **2025**, *16*, 1747638, doi:10.3389/fphar.2025.1747638.
48. Zheng, Y.; He, Y. VaxjoGNN: A Graph Neural Network for Ontology-Grounded Vaccine Adjuvant Recommendation. *BioRxiv Prepr. Serv. Biol.* **2025**, 2025.11.27.690985, doi:10.1101/2025.11.27.690985.
49. Muhammad, S.; Li, B.; Zhong, W.; Siddiqui, H.; Zhao, Y.; Wang, J. AI-Enhanced Synergistic Chemo-Immunotherapy: From Mechanistic Insights to Clinical Translation. *Crit. Rev. Oncol. Hematol.* **2026**, *217*, 105064, doi:10.1016/j.critrevonc.2025.105064.
50. Wages, N.A.; Slingluff, C.L. Flexible Phase I-II Design for Partially Ordered Regimens with Application to Therapeutic Cancer Vaccines. *Stat. Biosci.* **2020**, *12*, 104–123, doi:10.1007/s12561-019-09245-3.

51. Xu, Y.; Ma, S.; Cui, H.; Chen, J.; Xu, S.; Gong, F.; Golubovic, A.; Zhou, M.; Wang, K.C.; Varley, A.; et al. AGILE Platform: A Deep Learning Powered Approach to Accelerate LNP Development for mRNA Delivery. *Nat. Commun.* **2024**, *15*, 6305, doi:10.1038/s41467-024-50619-z.
52. Zeng, J.; Chen, R.; He, S.; Zhang, C.; Liu, W.-C.; Song, H.; Liu, S.; Nan, K. High-Throughput, AI-Assisted Design and Optimization of Lipid Nanoparticles for Drug Delivery. *J. Control. Release Off. J. Control. Release Soc.* **2025**, *391*, 114573, doi:10.1016/j.jconrel.2025.114573.
53. Witten, J.; Raji, I.; Manan, R.S.; Beyler, E.; Bartlett, S.; Tang, Y.; Ebadi, M.; Lei, J.; Nguyen, D.; Oladimeji, F.; et al. Artificial Intelligence-Guided Design of Lipid Nanoparticles for Pulmonary Gene Therapy. *Nat. Biotechnol.* **2025**, *43*, 1790–1799, doi:10.1038/s41587-024-02490-y.
54. Gautam, V.; Gupta, R.; Gupta, D.; Ruhela, A.; Mittal, A.; Mohanty, S.K.; Arora, S.; Gupta, R.; Saini, C.; Sengupta, D.; et al. deepGraphh: AI-Driven Web Service for Graph-Based Quantitative Structure-Activity Relationship Analysis. *Brief. Bioinform.* **2022**, *23*, bbac288, doi:10.1093/bib/bbac288.
55. Koirala, M.; Yan, L.; Mohamed, Z.; DiPaola, M. AI-Integrated QSAR Modeling for Enhanced Drug Discovery: From Classical Approaches to Deep Learning and Structural Insight. *Int. J. Mol. Sci.* **2025**, *26*, 9384, doi:10.3390/ijms26199384.
56. Mao, J.; Akhtar, J.; Zhang, X.; Sun, L.; Guan, S.; Li, X.; Chen, G.; Liu, J.; Jeon, H.-N.; Kim, M.S.; et al. Comprehensive Strategies of Machine-Learning-Based Quantitative Structure-Activity Relationship Models. *iScience* **2021**, *24*, 103052, doi:10.1016/j.isci.2021.103052.
57. Monroig-Bosque, P. del C.; Rivera, C.A.; Calin, G.A. MicroRNAs in Cancer Therapeutics: From the Bench to the Bedside. *Expert Opin. Biol. Ther.* **2015**, *15*, 1381–1385, doi:10.1517/14712598.2015.1074999.
58. Vora, L.K.; Gholap, A.D.; Jetha, K.; Thakur, R.R.S.; Solanki, H.K.; Chavda, V.P. Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. *Pharmaceutics* **2023**, *15*, 1916, doi:10.3390/pharmaceutics15071916.
59. Olawade, D.B.; Teke, J.; Fapohunda, O.; Weerasinghe, K.; Usman, S.O.; Ige, A.O.; Clement David-Olawade, A. Leveraging Artificial Intelligence in Vaccine Development: A Narrative Review. *J. Microbiol. Methods* **2024**, *224*, 106998, doi:10.1016/j.mimet.2024.106998.
60. Lu, L.; Lu, X.; Luo, W. Personalized Cancer Vaccines: Current Advances and Emerging Horizons. *Vaccines* **2025**, *13*, 1231, doi:10.3390/vaccines13121231.
61. Kawashima, S.; Pokarowski, P.; Pokarowska, M.; Kolinski, A.; Katayama, T.; Kanehisa, M. AAindex: Amino Acid Index Database, Progress Report 2008. *Nucleic Acids Res.* **2008**, *36*, D202-205, doi:10.1093/nar/gkm998.
62. Qi, Y.; Zheng, P.; Huang, G. DeepLBCEPred: A Bi-LSTM and Multi-Scale CNN-Based Deep Learning Method for Predicting Linear B-Cell Epitopes. *Front. Microbiol.* **2023**, *14*, 1117027, doi:10.3389/fmicb.2023.1117027.
63. Olawade, D.B.; Teke, J.; Fapohunda, O.; Weerasinghe, K.; Usman, S.O.; Ige, A.O.; Clement David-Olawade, A. Leveraging Artificial Intelligence in Vaccine Development: A Narrative Review. *J. Microbiol. Methods* **2024**, *224*, 106998, doi:10.1016/j.mimet.2024.106998.
64. Shao, X.M.; Bhattacharya, R.; Huang, J.; Sivakumar, I.K.A.; Tokheim, C.; Zheng, L.; Hirsch, D.; Kaminow, B.; Omdahl, A.; Bonsack, M.; et al. High-Throughput Prediction of MHC Class I and II Neoantigens with MHCnuggets. *Cancer Immunol. Res.* **2020**, *8*, 396–408, doi:10.1158/2326-6066.CIR-19-0464.
65. Bi, J.; Zheng, Y.; Wang, C.; Ding, Y. An Attention Based Bidirectional LSTM Method to Predict the Binding of TCR and Epitope. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 3272–3280, doi:10.1109/TCBB.2021.3115353.
66. Li, J.; Li, X.; Dong, J.; Wei, J.; Guo, X.; Wang, G.; Xu, M.; Zhao, A. Enhanced Immune Responses in Mice by Combining the Mpox Virus B6R-Protein and Aluminum Hydroxide-CpG Vaccine Adjuvants. *Vaccines* **2024**, *12*, 776, doi:10.3390/vaccines12070776.
67. Zeng, Y.; Wei, Z.; Yuan, Q.; Chen, S.; Yu, W.; Lu, Y.; Gao, J.; Yang, Y. Identifying B-Cell Epitopes Using AlphaFold2 Predicted Structures and Pretrained Language Model. *Bioinforma. Oxf. Engl.* **2023**, *39*, btad187, doi:10.1093/bioinformatics/btad187.
68. Hu, R.-S.; Gu, K.; Ehsan, M.; Abbas Raza, S.H.; Wang, C.-R. Transformer-Based Deep Learning Enables Improved B-Cell Epitope Prediction in Parasitic Pathogens: A Proof-of-Concept Study on *Fasciola Hepatica*. *PLoS Negl. Trop. Dis.* **2025**, *19*, e0012985, doi:10.1371/journal.pntd.0012985.

69. Machaca, V.; Goyzueta, V.; Cruz, M.G.; Seije, E.; Pilco, L.M.; López, J.; Túpac, Y. Transformers Meets Neoantigen Detection: A Systematic Literature Review. *J. Integr. Bioinforma.* **2024**, *21*, 20230043, doi:10.1515/jib-2023-0043.
70. Tokar, T.; Pastrello, C.; Abovsky, M.; Rahmati, S.; Jurisica, I. miRAnno-Network-Based Functional microRNA Annotation. *Bioinforma. Oxf. Engl.* **2022**, *38*, 592–593, doi:10.1093/bioinformatics/btab527.
71. Darmawan, J.T.; Leu, J.-S.; Avian, C.; Ratnasari, N.R.P. MITNet: A Fusion Transformer and Convolutional Neural Network Architecture Approach for T-Cell Epitope Prediction. *Brief. Bioinform.* **2023**, *24*, bbad202, doi:10.1093/bib/bbad202.
72. Choi, S.; Kim, D. B Cell Epitope Prediction by Capturing Spatial Clustering Property of the Epitopes Using Graph Attention Network. *Sci. Rep.* **2024**, *14*, 27496, doi:10.1038/s41598-024-78506-z.
73. Theiler, J.; Yoon, H.; Yusim, K.; Picker, L.J.; Fruh, K.; Korber, B. Epigraph: A Vaccine Design Tool Applied to an HIV Therapeutic Vaccine and a Pan-Filovirus Vaccine. *Sci. Rep.* **2016**, *6*, 33987, doi:10.1038/srep33987.
74. Zeng, Y.; Wei, Z.; Yuan, Q.; Chen, S.; Yu, W.; Lu, Y.; Gao, J.; Yang, Y. Identifying B-Cell Epitopes Using AlphaFold2 Predicted Structures and Pretrained Language Model. *Bioinforma. Oxf. Engl.* **2023**, *39*, btad187, doi:10.1093/bioinformatics/btad187.
75. Thrift, W.J.; Perera, J.; Cohen, S.; Lounsbury, N.W.; Gurung, H.R.; Rose, C.M.; Chen, J.; Jhunjhunwala, S.; Liu, K. Graph-pMHC: Graph Neural Network Approach to MHC Class II Peptide Presentation and Antibody Immunogenicity. *Brief. Bioinform.* **2024**, *25*, bbae123, doi:10.1093/bib/bbae123.
76. Seyran, M. Artificial Intelligence and Clinical Data Suggest the T Cell-Mediated SARS-CoV-2 Nonstructural Protein Intranasal Vaccines for Global COVID-19 Immunity. *Vaccine* **2022**, *40*, 4296–4300, doi:10.1016/j.vaccine.2022.06.052.
77. Kovalchik, K.A.; Hamelin, D.J.; Kubiniok, P.; Bourdin, B.; Mostefai, F.; Poujol, R.; Paré, B.; Simpson, S.M.; Sidney, J.; Bonnell, É.; et al. Machine Learning-Enhanced Immunopeptidomics Applied to T-Cell Epitope Discovery for COVID-19 Vaccines. *Nat. Commun.* **2024**, *15*, 10316, doi:10.1038/s41467-024-54734-9.
78. He, L.; Zhu, J. Computational Tools for Epitope Vaccine Design and Evaluation. *Curr. Opin. Virol.* **2015**, *11*, 103–112, doi:10.1016/j.coviro.2015.03.013.
79. He, L.; Zhu, J. Computational Tools for Epitope Vaccine Design and Evaluation. *Curr. Opin. Virol.* **2015**, *11*, 103–112, doi:10.1016/j.coviro.2015.03.013.
80. He, L.; Zhu, J. Computational Tools for Epitope Vaccine Design and Evaluation. *Curr. Opin. Virol.* **2015**, *11*, 103–112, doi:10.1016/j.coviro.2015.03.013.
81. He, L.; Zhu, J. Computational Tools for Epitope Vaccine Design and Evaluation. *Curr. Opin. Virol.* **2015**, *11*, 103–112, doi:10.1016/j.coviro.2015.03.013.
82. Rubinstein, N.D.; Mayrose, I.; Pupko, T. A Machine-Learning Approach for Predicting B-Cell Epitopes. *Mol. Immunol.* **2009**, *46*, 840–847, doi:10.1016/j.molimm.2008.09.009.
83. Niu, J.; Deng, R.; Dong, Z.; Yang, X.; Xing, Z.; Yu, Y.; Kang, J. Mapping the Landscape of AI and ML in Vaccine Innovation: A Bibliometric Study. *Hum. Vaccines Immunother.* **2025**, *21*, 2501358, doi:10.1080/21645515.2025.2501358.
84. Chai, Z.-L.; Qi, X.-X.; Li, R.; Luo, J.-R.; Li, C.; Shi, H.-D.; Tian, T.-T.; Shang, K.-Y.; Zhu, Y.-J.; Zhang, F.-B. Reverse Vaccinology-Driven Construction and Bioinformatics Validation of a Multi-Epitope Vaccine against *Brucella* Spp. *Sci. Rep.* **2025**, *15*, 36663, doi:10.1038/s41598-025-20507-7.
85. Vaghasiya, J.; Khan, M.; Milan Bakhda, T. A Meta-Analysis of AI and Machine Learning in Project Management: Optimizing Vaccine Development for Emerging Viral Threats in Biotechnology. *Int. J. Med. Inf.* **2025**, *195*, 105768, doi:10.1016/j.ijmedinf.2024.105768.
86. Anderson, L.N.; Hoyt, C.T.; Zucker, J.D.; McNaughton, A.D.; Teuton, J.R.; Karis, K.; Arokium-Christian, N.N.; Warley, J.T.; Stromberg, Z.R.; Gyori, B.M.; et al. Computational Tools and Data Integration to Accelerate Vaccine Development: Challenges, Opportunities, and Future Directions. *Front. Immunol.* **2025**, *16*, 1502484, doi:10.3389/fimmu.2025.1502484.
87. Parker, J.M.; Guo, D.; Hodges, R.S. New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-Ray-Derived Accessible Sites. *Biochemistry* **1986**, *25*, 5425–5432, doi:10.1021/bi00367a013.

88. Thornton, J.M.; Edwards, M.S.; Taylor, W.R.; Barlow, D.J. Location of continuous Antigenic Determinants in the Protruding Regions of Proteins. *EMBO J.* **1986**, *5*, 409–413, doi:10.1002/j.1460-2075.1986.tb04226.x.
89. Emini, E.A.; Hughes, J.V.; Perlow, D.S.; Boger, J. Induction of Hepatitis A Virus-Neutralizing Antibody by a Virus-Specific Synthetic Peptide. *J. Virol.* **1985**, *55*, 836–839, doi:10.1128/JVI.55.3.836-839.1985.
90. Roche, R.; Moussad, B.; Shuvo, M.H.; Bhattacharya, D. E(3) Equivariant Graph Neural Networks for Robust and Accurate Protein-Protein Interaction Site Prediction. *PLoS Comput. Biol.* **2023**, *19*, e1011435, doi:10.1371/journal.pcbi.1011435.
91. Cha, M.; Emre, E.S.T.; Xiao, X.; Kim, J.-Y.; Bogdan, P.; VanEpps, J.S.; Violi, A.; Kotov, N.A. Unifying Structural Descriptors for Biological and Bioinspired Nanoscale Complexes. *Nat. Comput. Sci.* **2022**, *2*, 243–252, doi:10.1038/s43588-022-00229-w.
92. Yuan, Q.; Chen, J.; Zhao, H.; Zhou, Y.; Yang, Y. Structure-Aware Protein-Protein Interaction Site Prediction Using Deep Graph Convolutional Network. *Bioinforma. Oxf. Engl.* **2021**, *38*, 125–132, doi:10.1093/bioinformatics/btab643.
93. Muhammed, M.T.; Aki-Yalcin, E. Homology Modeling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chem. Biol. Drug Des.* **2019**, *93*, 12–20, doi:10.1111/cbdd.13388.
94. Dormitzer, P.R.; Ulmer, J.B.; Rappuoli, R. Structure-Based Antigen Design: A Strategy for next Generation Vaccines. *Trends Biotechnol.* **2008**, *26*, 659–667, doi:10.1016/j.tibtech.2008.08.002.
95. ValizadehAslani, T.; Shi, Y.; Ren, P.; Wang, J.; Zhang, Y.; Hu, M.; Zhao, L.; Liang, H. PharmBERT: A Domain-Specific BERT Model for Drug Labels. *Brief. Bioinform.* **2023**, *24*, bbad226, doi:10.1093/bib/bbad226.
96. Ong, E.; Wong, M.U.; Huffman, A.; He, Y. COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning. *Front. Immunol.* **2020**, *11*, 1581, doi:10.3389/fimmu.2020.01581.
97. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449–W454, doi:10.1093/nar/gkaa379.
98. Doytchinova, I.A.; Flower, D.R. VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinformatics* **2007**, *8*, 4, doi:10.1186/1471-2105-8-4.
99. Suri, S.; Dakshanamurthy, S. IntegralVac: A Machine Learning-Based Comprehensive Multivalent Epitope Vaccine Design Method. *Vaccines* **2022**, *10*, 1678, doi:10.3390/vaccines10101678.
100. Tennant, P.W.G.; Murray, E.J.; Arnold, K.F.; Berrie, L.; Fox, M.P.; Gadd, S.C.; Harrison, W.J.; Keeble, C.; Ranker, L.R.; Textor, J.; et al. Use of Directed Acyclic Graphs (DAGs) to Identify Confounders in Applied Health Research: Review and Recommendations. *Int. J. Epidemiol.* **2021**, *50*, 620–632, doi:10.1093/ije/dyaa213.
101. Alomair, G. Predictive Performance of Count Regression Models versus Machine Learning Techniques: A Comparative Analysis Using an Automobile Insurance Claims Frequency Dataset. *PloS One* **2024**, *19*, e0314975, doi:10.1371/journal.pone.0314975.
102. Kimura, H.; Nishikawa, M.; Kutsuzawa, N.; Tokito, F.; Kobayashi, T.; Kurniawan, D.A.; Shioda, H.; Cao, W.; Shinha, K.; Nakamura, H.; et al. Advancements in Microphysiological Systems: Exploring Organoids and Organ-on-a-Chip Technologies in Drug Development -Focus on Pharmacokinetics Related Organs. *Drug Metab. Pharmacokinet.* **2025**, *60*, 101046, doi:10.1016/j.dmpk.2024.101046.
103. Qorib, M.; Oladunni, T.; Denis, M.; Ososanya, E.; Cotae, P. COVID-19 Vaccine Hesitancy: A Global Public Health and Risk Modelling Framework Using an Environmental Deep Neural Network, Sentiment Classification with Text Mining and Emotional Reactions from COVID-19 Vaccination Tweets. *Int. J. Environ. Res. Public Health* **2023**, *20*, 5803, doi:10.3390/ijerph20105803.
104. Vo, T.H.; McNeela, E.; O'Donovan, O.; Rani, S.; Mehta, J.P. Artificial Intelligence and the Evolving Landscape of Immunoepitidomics. *Proteomics Clin. Appl.* **2025**, *19*, e70018, doi:10.1002/prca.70018.
105. Shahbazy, M.; Ramarathinam, S.H.; Li, C.; Illing, P.T.; Faridi, P.; Croft, N.P.; Purcell, A.W. MHCpLogics: An Interactive Machine Learning-Based Tool for Unsupervised Data Visualization and Cluster Analysis of Immunoepitidomes. *Brief. Bioinform.* **2024**, *25*, bbae087, doi:10.1093/bib/bbae087.
106. Esposito, S.; Campana, B.R.; Seferi, H.; Cinti, E.; Argentiero, A. Digital Twins in Pediatric Infectious Diseases: Virtual Models for Personalized Management. *J. Pers. Med.* **2025**, *15*, 514, doi:10.3390/jpm15110514.

107. Wang, H.; Arulraj, T.; Ippolito, A.; Popel, A.S. From Virtual Patients to Digital Twins in Immuno-Oncology: Lessons Learned from Mechanistic Quantitative Systems Pharmacology Modeling. *NPJ Digit. Med.* **2024**, *7*, 189, doi:10.1038/s41746-024-01188-4.
108. Li, L.; Back, S.-I.; Ma, J.; Guo, Y.; Galeandro-Diamant, T.; Clénet, D. Bayesian Optimization and Machine Learning for Vaccine Formulation Development. *PloS One* **2025**, *20*, e0324205, doi:10.1371/journal.pone.0324205.
109. Somogyi, E.; Csiszovszki, Z.; Molnár, L.; Lőrincz, O.; Tóth, J.; Pattijn, S.; Schockaert, J.; Mazy, A.; Miklós, I.; Pántya, K.; et al. A Peptide Vaccine Candidate Tailored to Individuals' Genetics Mimics the Multi-Targeted T Cell Immunity of COVID-19 Convalescent Subjects. *Front. Genet.* **2021**, *12*, 684152, doi:10.3389/fgene.2021.684152.
110. James, C.A.; Ronning, P.; Cullinan, D.; Cotto, K.C.; Barnell, E.K.; Campbell, K.M.; Skidmore, Z.L.; Sanford, D.E.; Goedegebuure, S.P.; Gillanders, W.E.; et al. In Silico Epitope Prediction Analyses Highlight the Potential for Distracting Antigen Immunodominance with Allogeneic Cancer Vaccines. *Cancer Res. Commun.* **2021**, *1*, 115–126, doi:10.1158/2767-9764.crc-21-0029.
111. Ringeval, M.; Etindele Sosso, F.A.; Cousineau, M.; Paré, G. Advancing Health Care With Digital Twins: Meta-Review of Applications and Implementation Challenges. *J. Med. Internet Res.* **2025**, *27*, e69544, doi:10.2196/69544.
112. Lasky, T.; McMahon, A.W.; Hua, W.; Forshee, R. Methodologic Approaches in Studies Using Real-World Data (RWD) to Measure Pediatric Safety and Effectiveness of Vaccines Administered to Pregnant Women: A Scoping Review. *Vaccine* **2021**, *39*, 3814–3824, doi:10.1016/j.vaccine.2021.05.071.
113. Batech, M.; Madsen, A.; Gatto, N.; Zhang, T.C.; Ricci, D.; Harvey, R.; Khan, N.; Jain, S. Combining Real-World and Clinical Trial Data Through Privacy-Preserving Record Linkage: Opportunities and Challenges—A Narrative Review. *Health Sci. Rep.* **2025**, *8*, e71272, doi:10.1002/hsr2.71272.
114. Arunachalam, P.S.; Wimmers, F.; Mok, C.K.P.; Perera, R.A.P.M.; Scott, M.; Hagan, T.; Sigal, N.; Feng, Y.; Bristow, L.; Tak-Yin Tsang, O.; et al. Systems Biological Assessment of Immunity to Mild versus Severe COVID-19 Infection in Humans. *Science* **2020**, *369*, 1210–1220, doi:10.1126/science.abc6261.
115. Ramasamy, M.N.; Kelly, E.J.; Seegobin, S.; Dargan, P.I.; Payne, R.; Libri, V.; Adam, M.; Aley, P.K.; Martinez-Alier, N.; Church, A.; et al. Immunogenicity and Safety of AZD2816, a Beta (B.1.351) Variant COVID-19 Vaccine, and AZD1222 (ChAdOx1 nCoV-19) as Third-Dose Boosters for Previously Vaccinated Adults: A Multicentre, Randomised, Partly Double-Blinded, Phase 2/3 Non-Inferiority Immunobridging Study in the UK and Poland. *Lancet Microbe* **2023**, *4*, e863–e874, doi:10.1016/S2666-5247(23)00177-5.
116. Bockstal, V.; Leyssen, M.; Heerwegh, D.; Spiessens, B.; Robinson, C.; Stoop, J.N.; Roozendaal, R.; Van Effelterre, T.; Gaddah, A.; Van Roey, G.A.; et al. Non-Human Primate to Human Immunobridging Demonstrates a Protective Effect of Ad26.ZEBOV, MVA-BN-Filo Vaccine against Ebola. *NPJ Vaccines* **2022**, *7*, 156, doi:10.1038/s41541-022-00564-z.
117. Khoury, D.S.; Schlub, T.E.; Cromer, D.; Steain, M.; Fong, Y.; Gilbert, P.B.; Subbarao, K.; Triccas, J.A.; Kent, S.J.; Davenport, M.P. Correlates of Protection, Thresholds of Protection, and Immunobridging among Persons with SARS-CoV-2 Infection. *Emerg. Infect. Dis.* **2023**, *29*, 381–388, doi:10.3201/eid2902.221422.
118. Joffe, M. Principal Stratification and Attribution Prohibition: Good Ideas Taken Too Far. *Int. J. Biostat.* **2011**, *7*, Article 35, doi:10.2202/1557-4679.1367.
119. Sherman, E.; Shpitser, I. Identification and Estimation Of Causal Effects from Dependent Data. *Adv. Neural Inf. Process. Syst.* **2018**, *2018*, 9446–9457.
120. Duarte, G.; Finkelstein, N.; Knox, D.; Mummolo, J.; Shpitser, I. An Automated Approach to Causal Inference in Discrete Settings. *J. Am. Stat. Assoc.* **2024**, *119*, 1778–1793, doi:10.1080/01621459.2023.2216909.
121. Gilbert, P.B.; Fong, Y.; Hejazi, N.S.; Kenny, A.; Huang, Y.; Carone, M.; Benkeser, D.; Follmann, D. Four Statistical Frameworks for Assessing an Immune Correlate of Protection (Surrogate Endpoint) from a Randomized, Controlled, Vaccine Efficacy Trial. *Vaccine* **2024**, *42*, 2181–2190, doi:10.1016/j.vaccine.2024.02.071.
122. Zhang, X.-M.; Gao, T.-H.; Cai, Q.-Y.; Xia, J.-B.; Sun, Y.-N.; Yang, J.; Li, W.-H.; Zhang, S.-X.-M.; Lou, H.-R.; Yu, X.-T.; et al. Artificial Intelligence in Digital Pathology Diagnosis and Analysis: Technologies, Challenges, and Future Prospects. *Mil. Med. Res.* **2026**, *12*, 93, doi:10.1186/s40779-025-00680-6.

123. Mistretta, B.; Rankothgedera, S.; Castillo, M.; Rao, M.; Holloway, K.; Bhardwaj, A.; El Noafal, M.; Albarracin, C.; El-Zein, R.; Rezaei, H.; et al. Chimeric RNAs Reveal Putative Neoantigen Peptides for Developing Tumor Vaccines for Breast Cancer. *Front. Immunol.* **2023**, *14*, 1188831, doi:10.3389/fimmu.2023.1188831.
124. Shao, X.M.; Huang, J.; Niknafs, N.; Balan, A.; Cherry, C.; White, J.; Velculescu, V.E.; Anagnostou, V.; Karchin, R. HLA Class II Immunogenic Mutation Burden Predicts Response to Immune Checkpoint Blockade. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2022**, *33*, 728–738, doi:10.1016/j.annonc.2022.03.013.
125. Shao, X.M.; Bhattacharya, R.; Huang, J.; Sivakumar, I.K.A.; Tokheim, C.; Zheng, L.; Hirsch, D.; Kaminow, B.; Omdahl, A.; Bonsack, M.; et al. High-Throughput Prediction of MHC Class I and II Neoantigens with MHCnuggets. *Cancer Immunol. Res.* **2020**, *8*, 396–408, doi:10.1158/2326-6066.CIR-19-0464.
126. Jiménez-Luna, J.; Grisoni, F.; Weskamp, N.; Schneider, G. Artificial Intelligence in Drug Discovery: Recent Advances and Future Perspectives. *Expert Opin. Drug Discov.* **2021**, *16*, 949–959, doi:10.1080/17460441.2021.1909567.
127. Zushin, P.-J.H.; Mukherjee, S.; Wu, J.C. FDA Modernization Act 2.0: Transitioning beyond Animal Models with Human Cells, Organoids, and AI/ML-Based Approaches. *J. Clin. Invest.* **2023**, *133*, e175824, doi:10.1172/JCI175824.
128. Singh, R.; Bapna, M.; Diab, A.R.; Ruiz, E.S.; Lotter, W. How AI Is Used in FDA-Authorized Medical Devices: A Taxonomy across 1,016 Authorizations. *NPJ Digit. Med.* **2025**, *8*, 388, doi:10.1038/s41746-025-01800-1.
129. Niazi, S.K. The Coming of Age of AI/ML in Drug Discovery, Development, Clinical Testing, and Manufacturing: The FDA Perspectives. *Drug Des. Devel. Ther.* **2023**, *17*, 2691–2725, doi:10.2147/DDDT.S424991.
130. Garg, P.; Pareek, S.; Kulkarni, P.; Horne, D.; Salgia, R.; Singhal, S.S. Next-Generation Immunotherapy: Advancing Clinical Applications in Cancer Treatment. *J. Clin. Med.* **2024**, *13*, 6537, doi:10.3390/jcm13216537.
131. Li, L.; Sun, M.; Wang, J.; Wan, S. Multi-Omics Based Artificial Intelligence for Cancer Research. *Adv. Cancer Res.* **2024**, *163*, 303–356, doi:10.1016/bs.acr.2024.06.005.
132. He, X.; Liu, X.; Zuo, F.; Shi, H.; Jing, J. Artificial Intelligence-Based Multi-Omics Analysis Fuels Cancer Precision Medicine. *Semin. Cancer Biol.* **2023**, *88*, 187–200, doi:10.1016/j.semcancer.2022.12.009.
133. Zadravec, M.; Metsi-Guckel, E.; Kamenik, B.; Remelgas, J.; Khinast, J.; Roscioli, N.; Flamm, M.; Renawala, H.; Najarian, J.; Karande, A.; et al. Towards a Digital Twin of Primary Drying in Lyophilization Using Coupled 3-D Equipment CFD and 1-D Vial-Scale Simulations. *Eur. J. Pharm. Biopharm. Off. J. Arbeitsgemeinschaft Pharm. Verfahrenstechnik EV* **2025**, *208*, 114662, doi:10.1016/j.ejpb.2025.114662.
134. Olawade, D.B.; Teke, J.; Fapohunda, O.; Weerasinghe, K.; Usman, S.O.; Ige, A.O.; Clement David-Olawade, A. Leveraging Artificial Intelligence in Vaccine Development: A Narrative Review. *J. Microbiol. Methods* **2024**, *224*, 106998, doi:10.1016/j.mimet.2024.106998.
135. Prakash, S.; Dhanushkodi, N.R.; Zayou, L.; Ibraim, I.C.; Quadiri, A.; Coulon, P.G.; Tifrea, D.F.; Suzer, B.; Shaik, A.M.; Chilukuri, A.; et al. Cross-Protection Induced by Highly Conserved Human B, CD4+, and CD8+ T-Cell Epitopes-Based Vaccine against Severe Infection, Disease, and Death Caused by Multiple SARS-CoV-2 Variants of Concern. *Front. Immunol.* **2024**, *15*, 1328905, doi:10.3389/fimmu.2024.1328905.
136. Vahed, H.; Prakash, S.; Quadiri, A.; Ibraim, I.C.; Omorogieva, E.; Patel, S.; Tadros, J.; Liao, E.J.; Lau, L.; Chentoufi, A.A.; et al. A Pan-Beta-Coronavirus Vaccine Bearing Conserved and Asymptomatic B- and T-Cell Epitopes Protects against Highly Pathogenic Delta and Highly Transmissible Omicron SARS-CoV-2 Variants. *Hum. Vaccines Immunother.* **2025**, *21*, 2527438, doi:10.1080/21645515.2025.2527438.
137. El Arab, R.A.; Alkhunaizi, M.; Alhashem, Y.N.; Al Khatib, A.; Bubsheet, M.; Hassanein, S. Artificial Intelligence in Vaccine Research and Development: An Umbrella Review. *Front. Immunol.* **2025**, *16*, 1567116, doi:10.3389/fimmu.2025.1567116.
138. Kovalchik, K.A.; Hamelin, D.J.; Kubiniok, P.; Bourdin, B.; Mostefai, F.; Poujol, R.; Paré, B.; Simpson, S.M.; Sidney, J.; Bonnell, É.; et al. Machine Learning-Enhanced Immunopeptidomics Applied to T-Cell Epitope Discovery for COVID-19 Vaccines. *Nat. Commun.* **2024**, *15*, 10316, doi:10.1038/s41467-024-54734-9.
139. Cho, M.K. Rising to the Challenge of Bias in Health Care AI. *Nat. Med.* **2021**, *27*, 2079–2081, doi:10.1038/s41591-021-01577-2.

140. Kanguane, P.; Sakharkar, M.K. T-Epitope Designer: A HLA-Peptide Binding Prediction Server. *Bioinformatics* **2005**, *1*, 21–24, doi:10.6026/97320630001021.
141. Gygi, J.P.; Kleinstein, S.H.; Guan, L. Predictive Overfitting in Immunological Applications: Pitfalls and Solutions. *Hum. Vaccines Immunother.* **2023**, *19*, 2251830, doi:10.1080/21645515.2023.2251830.
142. Şerban, M.; Toader, C.; Covache-Busuioc, R.-A. CRISPR and Artificial Intelligence in Neuroregeneration: Closed-Loop Strategies for Precision Medicine, Spinal Cord Repair, and Adaptive Neuro-Oncology. *Int. J. Mol. Sci.* **2025**, *26*, 9409, doi:10.3390/ijms26199409.
143. Witten, J.; Raji, I.; Manan, R.S.; Beyer, E.; Bartlett, S.; Tang, Y.; Ebadi, M.; Lei, J.; Nguyen, D.; Oladimeji, F.; et al. Artificial Intelligence-Guided Design of Lipid Nanoparticles for Pulmonary Gene Therapy. *Nat. Biotechnol.* **2025**, *43*, 1790–1799, doi:10.1038/s41587-024-02490-y.
144. Budhkar, A.; Song, Q.; Su, J.; Zhang, X. Demystifying the Black Box: A Survey on Explainable Artificial Intelligence (XAI) in Bioinformatics. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 346–359, doi:10.1016/j.csbj.2024.12.027.
145. Setegn, G.M.; Dejene, B.E. Explainable AI for Symptom-Based Detection of Monkeypox: A Machine Learning Approach. *BMC Infect. Dis.* **2025**, *25*, 419, doi:10.1186/s12879-025-10738-4.
146. Elfatimi, E.; Lekbach, Y.; Prakash, S.; BenMohamed, L. Artificial Intelligence and Machine Learning in the Development of Vaccines and Immunotherapeutics-Yesterday, Today, and Tomorrow. *Front. Artif. Intell.* **2025**, *8*, 1620572, doi:10.3389/frai.2025.1620572.
147. Jin, R.; Zhang, L. AI Applications in HIV Research: Advances and Future Directions. *Front. Microbiol.* **2025**, *16*, 1541942, doi:10.3389/fmicb.2025.1541942.
148. Gasperini, G.; Baylor, N.; Black, S.; Bloom, D.E.; Cramer, J.; de Lannoy, G.; Denoel, P.; Feinberg, M.; Helleputte, T.; Kang, G.; et al. Vaccinology in the Artificial Intelligence Era. *Sci. Transl. Med.* **2025**, *17*, eadu3791, doi:10.1126/scitranslmed.adu3791.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.