

Proceeding artificial intelligence technology towards vaccine designs

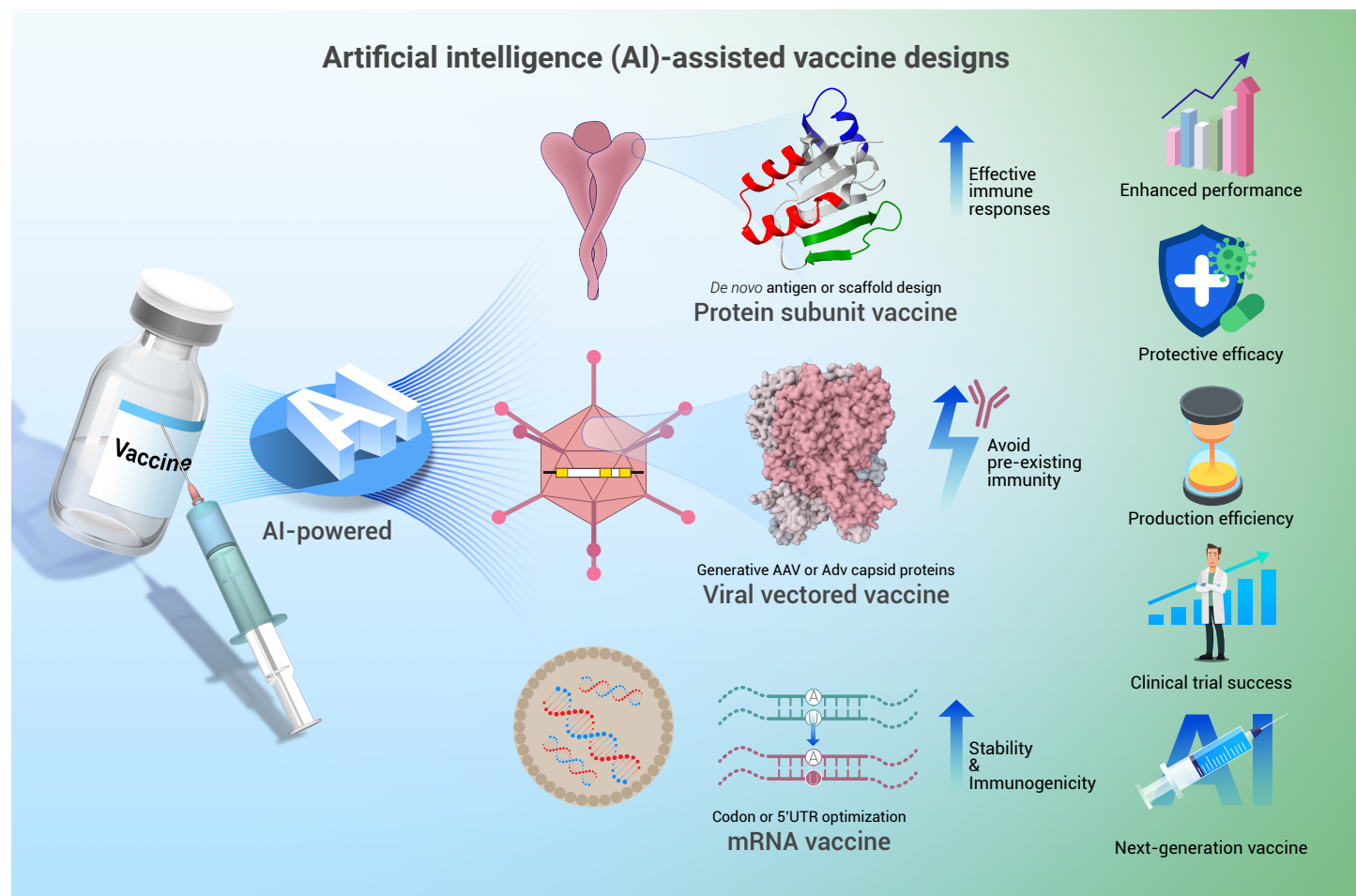
Fangxin Zhao^{1,2} and Xiaodong Zai^{1,*}

*Correspondence: zaxiaodong@163.com

Received: April 25, 2025; Accepted: September 30, 2025; Published Online: October 10, 2025; <https://doi.org/10.59717/j.xinn-life.2025.100162>

© 2025 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Artificial intelligence (AI) accelerates vaccine design for protein subunit, viral vector and mRNA platforms.
- AI revolutionizes vaccine development by modeling immune response and accelerating clinical trials.
- AI-driven strategies may enhance the efficacy of next-generation vaccines, improving pandemic preparedness.

Proceeding artificial intelligence technology towards vaccine designs

Fangxin Zhao^{1,2} and Xiaodong Zai^{1,*}

¹Laboratory of Advanced Biotechnology, Beijing Institute of Biotechnology, Beijing 100071, China

²School of Medicine, Zhejiang University, Hangzhou 310058, China

*Correspondence: zaxiaodong@163.com

Received: April 25, 2025; Accepted: September 30, 2025; Published Online: October 10, 2025; <https://doi.org/10.59717/j.xinn-life.2025.100162>

© 2025 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Zhao F. and Zai X. (2025). Proceeding artificial intelligence technology towards vaccine designs. *The Innovation Life* 3:100162.

Although vaccines have been designed to be safe and effective, it still takes years to move vaccine development from benchside to bedside. Novel strategies are needed to accelerate the vaccine design process. Artificial intelligence (AI) has started to play an important role in the life science industry. In this review, we summarized progress of AI-assisted vaccine designs in three areas, including *de novo* protein subunit design that focuses more on the immunogenic part of the antigen, viral vectored-based vaccine innovation that combats pre-existing immunity, and mRNA vaccine platform optimization that improves stability and translational efficiency. Additionally, we also address the multifaceted challenges associated with integrating AI into vaccine development, encompassing technical, ethical, and regulatory considerations, and offer strategic recommendations to navigate these complexities. AI-driven methodologies hold the promise of revolutionizing next-generation vaccine design by improving safety profiles, stability, immunogenicity, and overall efficacy, thereby strengthening global preparedness against emerging infectious threats.

INTRODUCTION

A new era of vaccine development using artificial intelligence (AI)

The outbreak of the coronavirus 2019 disease (COVID-19) pandemic has witnessed a surge in vaccine research and development, with numerous research teams and companies competing to develop effective vaccines. COVID-19 vaccination is estimated to prevent 14.4 million deaths within one year and has brought a huge impact on world health.¹ Vaccine research is constantly updated from formulations to delivery routes. First-generation vaccines include live-attenuated and inactivated vaccines that contain intact but weakened or dead pathogens.² As one of the most representative achievements in live-attenuated vaccines, the measles mumps and rubella (MMR) vaccine has saved millions of young lives worldwide.³⁻⁵ For inactivated vaccines, both inactivated vaccines from Sinovac and SinoPharm had been authorized for emergency use in fighting against COVID-19 by the World Health Organization (WHO).^{6,7} Other vaccines produced to protect the body from hepatitis B, influenza, and Japanese encephalitis virus infection were also inactivated. Although they are regarded as inducing the most biologically similar immunity, these techniques rely on large-scale live virus cultivation, isolation, propagation, and purification.

Protein subunit, vector-based, and nucleic acid vaccines are developed to guide our immune system in recognizing antigens more efficiently and thus stimulate a more robust immune response (Figure 1). Protein subunit vaccine that contains antigenic fragments of the pathogen elicits more direct and robust immune responses and eliminates adverse reactions caused by redundant components at the same time.^{8,9} However, the production of the protein subunit by recombinant technology may alter its original pathogen's secondary or tertiary protein conformation, directly influencing the antigen specificity. Besides, sometimes protein subunits alone are insufficient to elicit a strong immunogenic response, and thus both adjuvant and extra booster doses are required. Vector-based vaccines, including the adeno-associated virus (AAV) vaccine, lentiviral vaccine, and adenoviral vectors vaccine, effectively induce CD8⁺ T cell response for long-term immunity.¹⁰⁻¹³ Adenovirus type-5 (Ad5) (Convidecia developed by CanSino Biologics), Ad26 (Ad26.COV2.S developed by Johnson & Johnson), and ChAdOx1 (AZD1222 produced by AstraZeneca) were involved as vaccine platforms in SARS-CoV-2 vaccine development right after COVID-19 outbreak.¹⁴⁻¹⁶ They function by integrating target genes of pathogens into a harmless virus to mimic the infection process, triggering protective immunity against the targeted pathogen-associated molecular patterns (PAMPs). Ad5 has demonstrated its

efficacy as a delivery vector for vaccines against symptomatic or severe COVID-19 diseases in (i) homologous or heterologous booster vaccination strategy, (ii) monovalent or multivalent, and (iii) aerosolized or intramuscular administration way.¹⁶⁻¹⁸ However, pre-existing immunity sometimes dampens the efficacy of many viral-vectored vaccines. Nucleic acid vaccines, including DNA (as DNA plasmids) and RNA (as mRNA) vaccines, exhibit their abilities to induce dual (humoral and cellular) and durable immune responses via different mechanisms.¹⁹ Clinical trials of mRNA vaccine have been extensively reviewed, showcasing its potential as a promising alternative to conventional vaccine approaches.^{20,21} However, due to mRNA's low stability and easy degradation, storage and transportation of mRNA vaccines require fastidious, cold-chain conditions, creating insurmountable obstacles to preservation in less developed countries.²²

Vaccine development includes candidate vaccine design and evaluation, pre-clinical and clinical phases, large-scale manufacturing, and regulatory approval (Figure 1). For most infectious diseases, it takes more than a decade to go through the whole process, as each stage begins when the previous stage is completed, in which candidate vaccine design and evaluation typically consume five or more years. The manufacturing process has gradually transitioned from traditional platforms to genetically engineered technologies, making vaccine development towards reliance on computational modeling instead of tedious batch screening, offering a more systematic, efficient, and safe way to produce novel vaccines. Catalyzed by the COVID-19 pandemic, vaccines for infectious diseases were developed and produced at an unprecedented speed, with technical details encompassing strategies to combat SARS-CoV-2 variants and enhance the efficacy of novel delivery system.

Artificial intelligence (AI) refers to computer systems that are adept at executing intricate tasks, simulating the traditionally unique capabilities and intelligence of a human being (Figure 1).^{23,24} Machine learning (ML) is a subset of AI, while deep learning (DL) is a subfield of machine learning (Table 1, see details in Table S1).²⁵ ML refers to computer programs that automatically learn from and adapt to new data.²⁵ It could be further categorized into supervised, unsupervised learning, and reinforcement learning. The most significant difference between them is that supervised learning models have a baseline understanding of the correct output values, that is, the labels.²⁶ ML-guided B and T cell epitope discovery and antigen prediction have been extensively reviewed.²⁷ DL is a significant advancement over traditional machine learning that it autonomously determines weights. Emulating human neuron connection enables automatic learning using multilayered artificial neural networks (ANN) trained on vast amounts of data. Within hidden layers, DL algorithms iteratively extract sample information and perform calculations repeatedly, progressively improving the accuracy over time with increasing dataset size and generating predictions at the output layer.²⁸ AlphaFold, trRosetta, and RoseTTAFold are three representative DL-based protein structure prediction tools.²⁹⁻³¹ These protein prediction models pave the way for AI-based *de novo* vaccine candidate antigen generation. Large language models (LLMs), which is built upon the transformer architecture, are artificial neural networks operated by training extensive text datasets to learn patterns and relationships between words, grammar, and phrases (Table 1). LLMs have been successfully utilized in drug discovery and antibody and protein sequence generation.³²⁻³⁵ Generative protein language models disassemble functional protein features to amino acid sequences and learn to generate grammatically and semantically correct proteins, which holds the feasibility of expanding usage to variant effect prediction and novel drugs or vaccine immunogens creation.³⁶ Extensive studies have reviewed how generative LLMs contribute to medical research and healthcare.³⁷

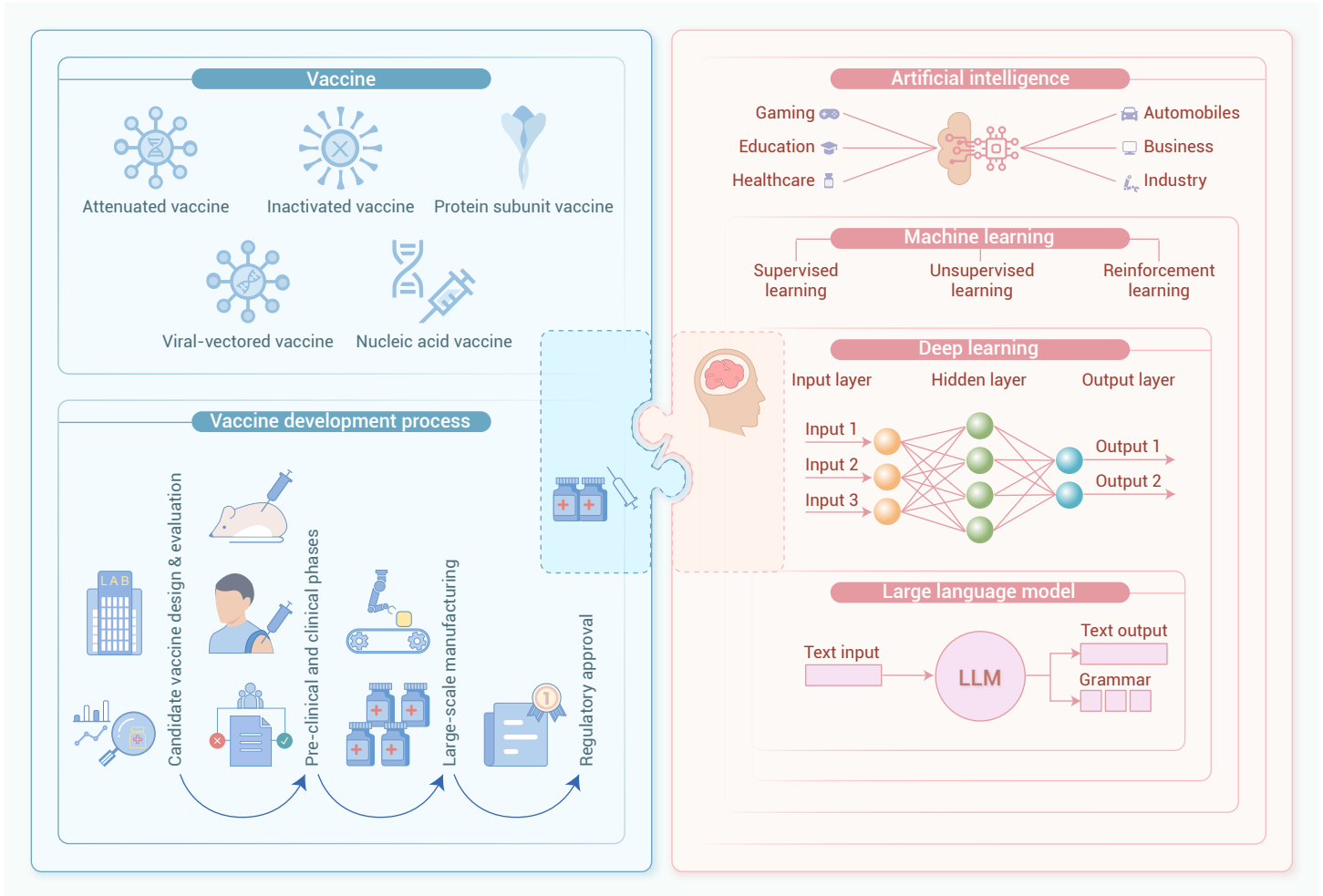


Figure 1. Brief introduction to vaccines and AI Left, currently available vaccines could be classified into five general types, and the vaccine development process contains four steps as shown in the lower left panel. Right, AI is involved in many areas. Machine learning is a subset of AI, which could be further divided into supervised learning, unsupervised learning, and reinforcement learning. Deep learning is an evolution of machine learning. For the most upgraded large language model, it is a computational model capable of comprehend and generate language after learning.

Table 1. AI-related technical terms

Concept	Terms (and examples)	References
Machine learning (ML): A subset of AI that involves algorithms allowing computers to learn from and make predictions or decisions based on data.	Supervised learning Unsupervised learning Semi-supervised learning Reinforcement learning	39,40
Deep learning (DL): An advanced subset of ML utilizing multilayered artificial neural networks to model complex patterns in large datasets.	Artificial neural networks (Autoencoder, RosettaFold2) Convolutional neural networks Recurrent neural networks (Long short-term memory networks, Self-attention mechanisms)	41
Generative models: AI models designed to generate new data instances by learning the underlying structure of input data.	Generative adversarial networks Variational autoencoders (Protein variational autoencoders) Transformer architecture Diffusion models (RFdiffusion, AlphaFold3) Flow-based	42-46
Large language models (LLMs): AI models based on transformer architectures that can understand and generate human-like text by learning from vast amounts of textual data.	Generative pre-trained transformer (ProtGPT2, Large language model Meta AI, ProLLaMa) Bidirectional encoder representations from transformers (PeptideBert) Text-to-text transfer transformer Protein language models (Evolutionary scale modeling, ProGen, UTR-LM, xTrimoPGLM, Evo/Evo2)	34,47-54

Empowered by AI technology, we are managing to go opposite the way “central dogma” does, which begins with protein function exploration, to structure arrangement, and then to amino acid or nucleic acid sequences.^{34,38} Currently, AI technology is increasingly integrated into biomedical develop-

ment pipelines such as facilitating earlier disease detection, advancing drug discovery, diversifying clinical trials, and evaluating drug-target interactions.^{55,56} Moreover, AI algorithms, profoundly with bioinformatics, now present researchers with the potential to expedite vaccine development,

including predicting potential antigens, identifying evolutionarily conserved regions, and assessing antigenic epitopes with high immunogenicity.^{27,57-59} In order to provide more detailed technical reference, we review the advancements of vaccine optimization in three fields under the support of AI technology, including protein subunit vaccine design, innovation of viral-vectored vaccines, and modification of mRNA vaccines. Large-scale data training instead of labor-intensive experiments position them as promising frontrunners for future vaccine design. These enumerated advancements would facilitate and accelerate the vaccine preclinical development process.

AI-based protein subunit vaccine design

Ideal protein subunit vaccines assist in redirecting the immune system to specific regions of the pathogens, thus triggering epitope-specific immune responses. Such vaccines may have significant advantages over conventional vaccine design routes, as the cautious assortment of the protein components eliminates the possibility of our body being exposed to inflammatory antigenic areas that do not elicit protective immunity while mimicking the original structure of epitope, a prerequisite in stimulating epitope-specific antibodies.

Respiratory syncytial virus (RSV) is a highly contagious virus that causes low respiratory tract inflammation with cold-like symptoms in children and the elderly. In the previous study, Correia et al. developed a computational method called Rosetta Fold From Loops (FFL) to design protein antigens with full backbone flexibility.⁶⁰ The small, thermally, and conformationally stable protein scaffolds can accurately maintain the tertiary structure of an epitope from RSV fusion glycoprotein, and more importantly, they successfully elicited neutralizing antibodies in non-human primates. Although this method is limited to single and linear epitopes, it highlights the potential for computer-assisted methods to generate effective vaccine immunogens. One major bottleneck for the above strategy is the lack of known, ready-to-use templates to fit the structural repertoire of the target epitope. With the development of a large language model, scientists gradually achieve the goal of protein antigen *de novo* design in which protein backbones are assembled *in silico*. Castro et al. described a *de novo* protein antigen design methodology called RFjoint2 Inpainting that can host up to three structurally distinct motifs of RSV simultaneously in a single scaffold (Figure 2A).⁶¹ The finding that these multi-epitope immunogens show high binding affinities to all three antibodies marks the success of AI stepping towards the next generation of *de novo* protein design with multiple epitopes. The generated immunogens with higher epitope-scaffold ratios are structurally accurate and show broader neutralization profiles than the single epitope.

Besides antigen *de novo* design, AI can also assist in the design of nanoparticles for multivalent antigen presentation.⁶² A milestone for AI-based COVID-19 vaccine design is that SKYCovione™ (GBP510 adjuvanted with AS03) won full approval in South Korea in 2022 and in the UK in 2023 under the collaboration of King's Lab at the Institute for Protein Design, SK Bioscience, and GlaxoSmithKline (GSK).^{63,64} Results from a Phase III clinical trial showed that SKYCovione™, the protein nanoparticle vaccine that contains 60 copies of the receptor binding domain (RBD) from the ancestral SARS-CoV-2 strain, is safe and more immunogenic to induce neutralizing antibody responses against the viral strain in comparison to AstraZeneca's Vaxzevria (NCT05007951).⁶⁴⁻⁶⁶ Besides, the GBP510-Beta variant booster dose demonstrates immunogenicity and protective efficacy against challenges with SARS-CoV-2 Delta subvariants in rhesus macaques.⁶⁷ The *de novo* nanoparticle tailored for self-assembly and multivalent antigen presentation was further developed by Ueda et al.⁶⁸ They firstly assembled monomeric repeat proteins into C3-symmetric trimers by precisely positioning C termini of the viral glycoproteins to the N-terminal of trimeric helices, and validated trimers called fusion components were then docked pairwise with *de novo* designed symmetric homo-oligomers, namely assembly component, to generate tetrahedral, octahedral, and icosahedral nanoparticle configurations (Figure 2B). Three types of viral glycoprotein that adopt trimeric conformation were utilized to confirm the design: influenza, HIV, and RSV. Results showed that in each case, the relevant viral glycoproteins could attach themselves to the scaffold and assemble themselves into nanoparticles with predicted geometrical shapes, showing high binding affinities to respective monoclonal antibodies during experimental characterization.

AI-assisted protein subunit vaccine design is not restricted to vaccines combating infectious diseases and scaffold construction. A personalized peptide-based cancer vaccine, EVX-01, that elicits immune responses targeting tumor neoantigens developed by AI has entered phase II clinical trial (NCT05309421).⁶⁹ More recently, a new breakthrough has been made in protein *de novo* design. ProteinGenerator is a sequence space diffusion model that generates both desired protein structures and sequences based on RoseTTAFold, which holds great promise to be further utilized in protein subunit vaccine development due to its flexibility in multi-state protein *de novo* design.⁷⁰

Beyond *de novo* immunogen designs, AI models can also help improve conformational stability, immunogenicity, and expression level of antigens. Liang et al. and Che et al. introduced amino acid point mutations aided by computational approaches, maintaining the RSV F protein in a highly immunogenic prefusion stable state, and enhancing protective immunity.^{71,72} For undesirable antibodies elicited by foldon trimeric tag which is commonly used in most RSV vaccines, the team at Janssen Vaccines and Prevention BV utilized tools such as the molecular dynamics engine and the Rosetta structural prediction to identify key unstable regions during trimerization process of the RSV F. By optimizing these regions, they developed the first stable RSV preF trimer vaccine without the need for heterologous sequences.⁷³ Bakkers et al. also developed an AI model named ReCaP to optimize human metapneumovirus fusion protein charged and solvent-mediated interactions.⁷⁴ They found that an additional cleavage event in the F protein enables the automatic and correct folding of the trimer into its prefusion conformation. Besides, by adding computational designed mutations, SARS-CoV-2 demonstrated higher yields and increased immunogenicity.⁷⁵

AI-based viral vectored vaccine innovation

Adeno-associated virus (AAV) is a single-stranded DNA dependoparvovirus. Due to its high thermostability and low pathogenicity, AAV has been proven to be safe and effective as a versatile vaccine vector.⁷⁶⁻⁷⁸ There are several serotypes of AAV with different receptors, among which AAV type 2 (AAV2) was first modified and approved by the US Food and Drug Administration as a powerful gene transfer toolkit for retinal disease treatment.^{79,80} However, pre-exposure to AAV may compromise vaccinal and therapeutic efficacy. Indeed, by design, epitopes available for neutralizing antibodies binding exist at many locations across the capsid surface. This strategic placement renders antibody neutralization impervious to random mutation, error-prone mutagenesis, or random shuffling between AAV serotypes of capsid protein as the resultant sequences remain a high similarity to the natural isolates.⁸¹⁻⁸³

Based on high-throughput DNA sequencing results, Ogden et al. trained ML models using dataset-architecture combinations to guide the design of functional and highly diverse capsid protein of AAV2 variants (Figure 3A).⁸⁴ The researchers first conducted saturation mutagenesis and next-generation sequencing on a 28-amino acid segment of the AAV2 VP1 capsid protein. This segment encompasses diverse structural domains, including buried regions, surface exposures, and interfacial zones, and overlaps both known heparin and antibody-binding sites. To generate large datasets for training ML models, they assembled various combinations: complete single mutations plus random double mutations ($C_1 + R_2$), complete single mutations plus random multiple mutations ($C_1 + R_{10}$), and random multiple mutations plus additively designed mutants ($R_{10} + A_{39}$). These datasets were then used as inputs for three different model architectures: logistic regression (LR) model, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) (Figure 3A, left panel). For the sequence generation workflow of each dataset-architecture combination, they trained an ensemble comprising 11 randomly initialized replica models. The average score derived from each ensemble was used to rank an extensive collection of 2.1 billion sequences. Model-selected sequences were then obtained by ranking candidates generated at random by model ensemble score. Top candidates underwent an intensive process of 20 iterative design cycles to obtain model-designed sequences (Figure 3A, middle panel). The research resulted in 201,426 resultant variants, of which 110,689 were viable and maintained the ability of assembly and genome packaging. Half of these viable proteins with 12–29 mutations across this region possessed a higher infection and delivery activ-

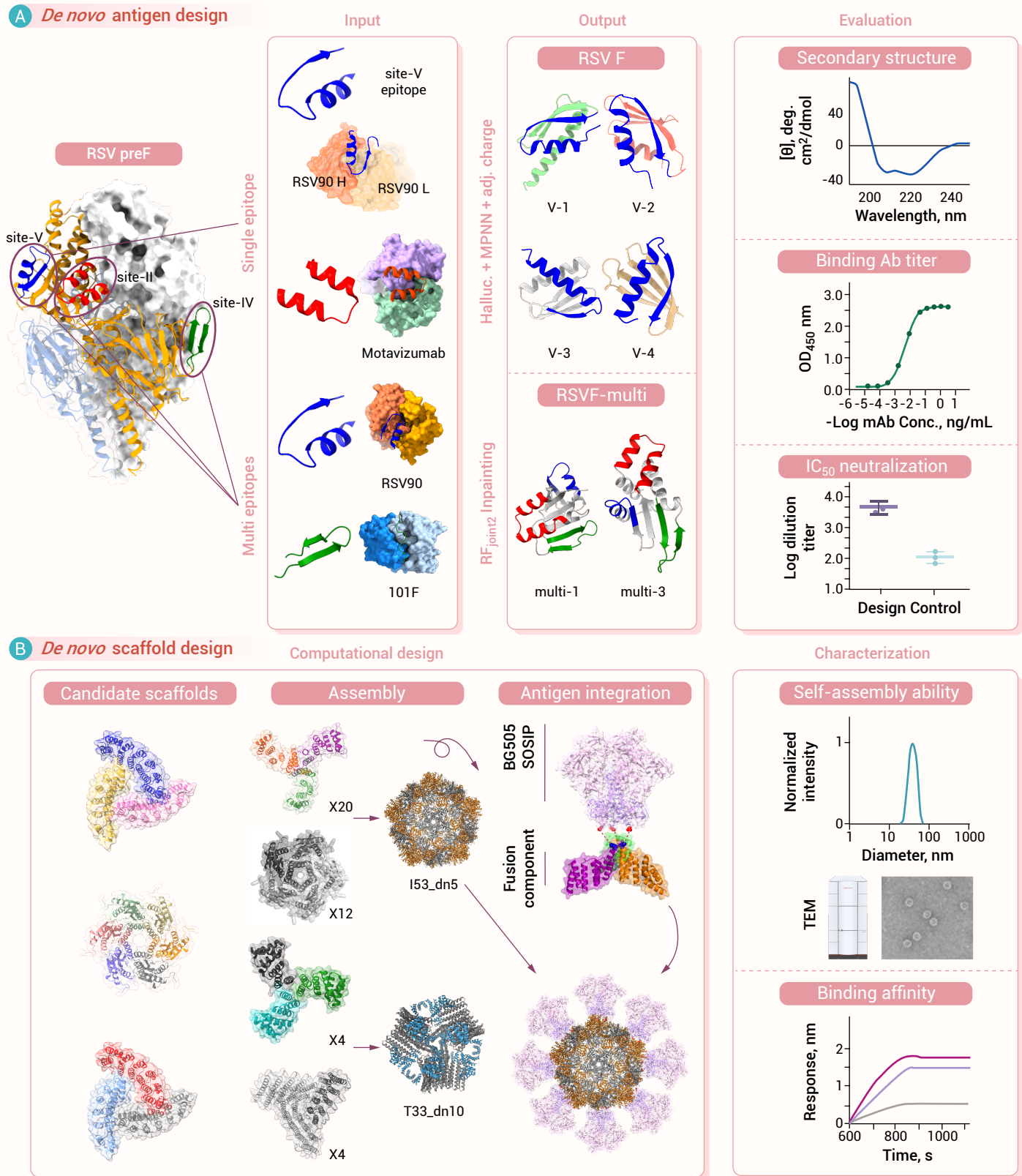
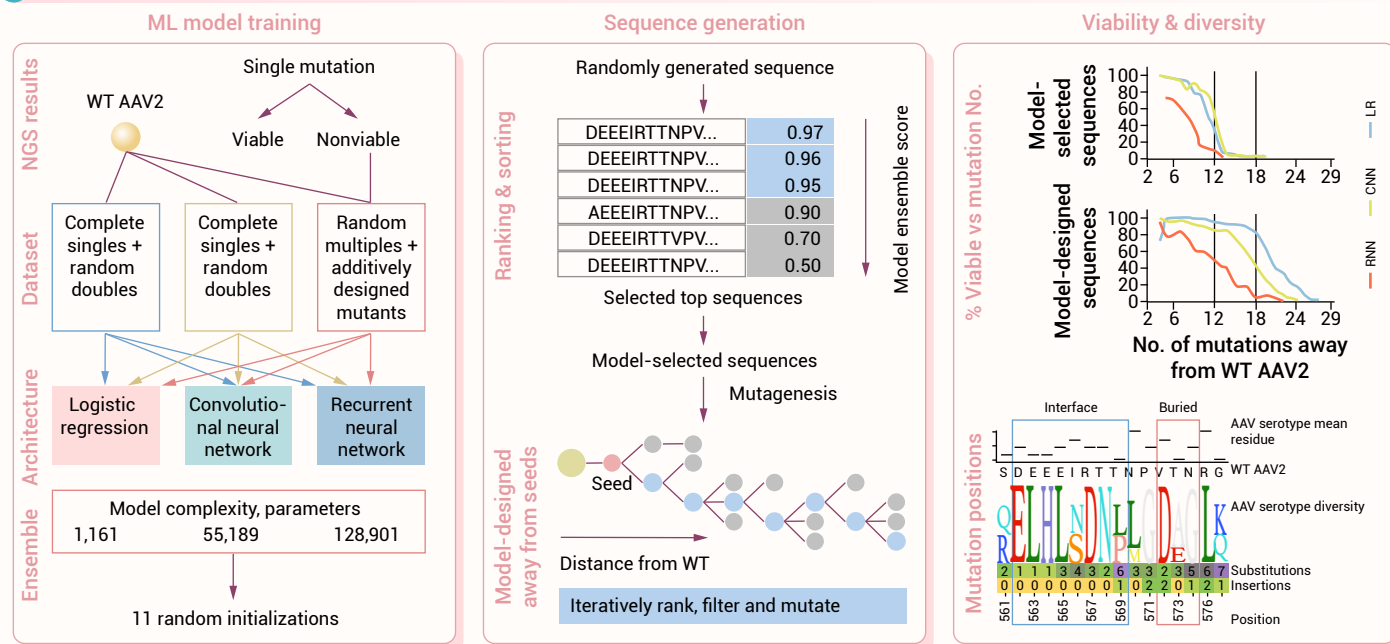
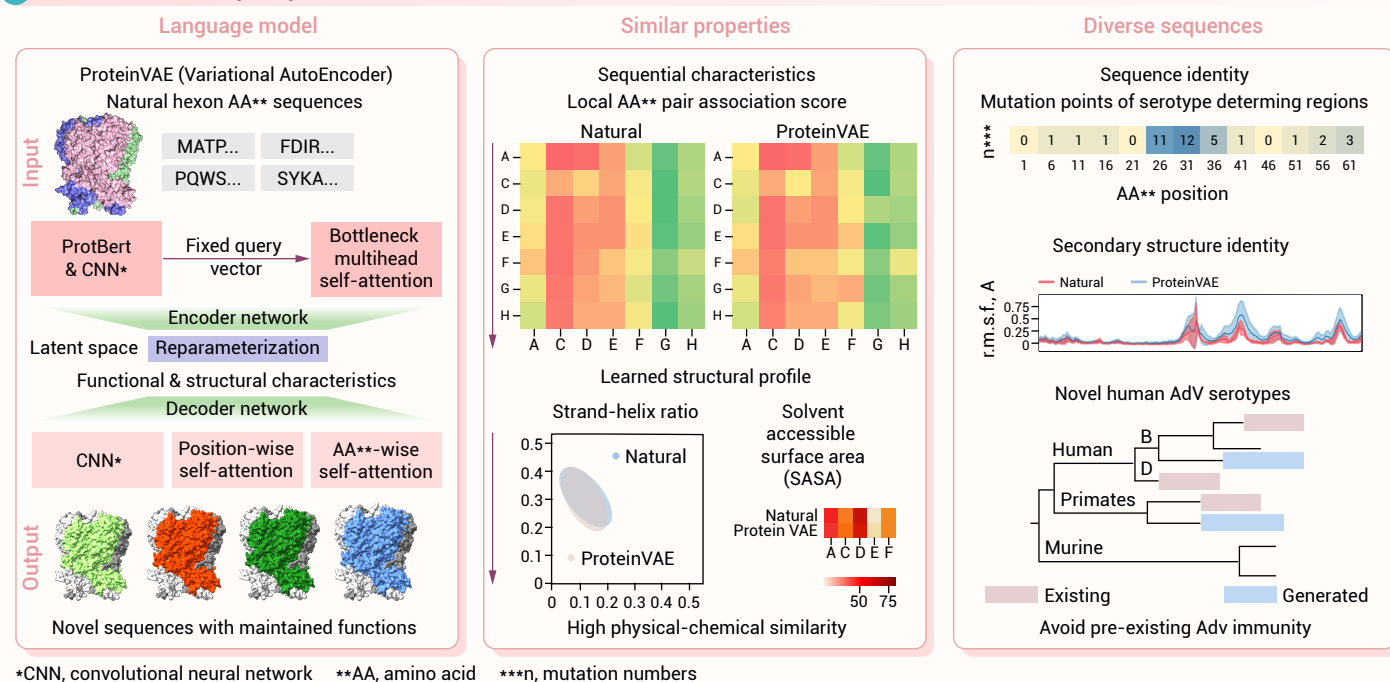


Figure 2. AI-assisted protein subunit vaccine design (A) Workflow and evaluation process of *de novo* RSV epitope-focused immunogen designs. Specific RSV epitope and its detailed binding information to neutralizing antibody was chosen as input for *de novo* vaccine immunogen design. *In silico* computational protein design strategies were used to generate novel backbones tailored for epitope display. The design was upgraded to involve three epitopes in a single immunogen. Secondary structure was evaluated to confirm its epitope display ability. Binding and neutralizing antibody titers were measured to evaluate the immunogenicity of the designed proteins. Halluc. Hallucination, MPNN Protein-MPNN, adj. adjusted. (B) *De novo* protein scaffold design to increase antigen multiply display efficiency. For the computational design process, several candidate scaffolds (fusion component and assembly component) were generated, and top designs with an estimated high chance of conducting self-assembly were experimentally verified. At last, the antigen of the model organism was docked onto the nanoparticle via the fusion component. For the biophysical characterization process, the diameter of the designed trimers incorporated in the designed nanoparticle was measured. Electron micrographs were randomly picked representing their actual morphology after self-assembly. The binding affinity of the targeted antigen was also measured.

A Generative AAV capsid proteins



B Generative Adv capsid proteins



*CNN, convolutional neural network **AA, amino acid ***n, mutation numbers

Figure 3. AI-assisted viral vectored vaccine innovation (A) *De novo* generation of AAV capsid protein. Left panel, training of machine learning model. Next-generation sequencing results and a complete single mutation dataset indicating viable or nonviable mutations were set as input. Input data was manipulated by complete (C), random (R), or additive (A) sampling strategies and were further allocated into three training datasets: complete singles + random doubles ($C_1 + R_2$), complete singles + random multiples ($C_1 + R_{10}$) and random multiples + additively designed mutants ($R_{10} + A_{30}$). The subscript number indicates the maximum number of mutations relative to WT. Each dataset was used to train three ML models: logistic regression (LR), convolutional neural network (CNN), and recurrent neural network (RNN). For each of the nine resultant dataset–ML model pairings, they implemented an ensemble of 11 randomly initialized model replicas and subsequently employed the average model score derived from each ensemble to rank 2.1 billion sequences. Middle panel, sequence generation part. Top-ranked sequences were selected as seeds and were used to generate new sequences that hold similar functions. The viability and diversity of synthetic sequences were validated experimentally. (B) *De novo* generation of adenovirus hexon sequences. Schematic illustration of protein variational autoencoder (ProteinVAE) model that is used to produce novel hexon sequences. For encoder network, the pre-trained language model ProtBert converts the sequence information to a linear layer. Decoder network takes the vector representation of the input, after going through the neural network and self-attention modules, the amino acid with the highest prediction probability at each position was selected and output sequences were created. Distinguishing from the traditional AE model, VAE model contains an extra part called latent space. Here, functional and structural characteristics of the target protein were involved, enabling a set of more precise outputs that maintain the function of natural hexons. To assess whether ProteinVAE has learned structure profiles of hexons during training, several parameters such as strand-helix ratio and solvent accessible surface area (SASA) were compared. At last, phylogenetic analysis was conducted to verify that the generated sequences belong to the identical phylogenetic clade but are divergent from existing serotypes.

ity than wild-type AAV2 virus, confirming the ability of ML models to generate functional protein sequences divergent from any natural homolog (Figure 3A, right panel).

Adenoviruses (Adv) are non-enveloped double-stranded DNA viruses that cause mild respiratory diseases. Adenoviral vectors, with their high gene transduction efficiency, large payload capacity, and broad tissue tropism, hold

significant promise as a vaccine antigen delivery vehicle.⁸⁵ However, a considerable portion of the human population possesses pre-existing anti-Ad immunity acquired from previous natural infections, especially in childhood, which is reported to affect the adenoviral vaccine immunogenicity.⁸⁶ Despite this challenge, the potential of adenoviral vectors offers hope for the future of vaccine development. The hexon is the most abundant protein that forms the icosahedral faces of the capsid and it is also the primary target of anti-Ad neutralizing antibodies.⁸⁷ The previous strategy to lower the global prevalence of pre-existing nAbs is to use rare serotype adenoviral vectors.⁸⁸ Lyu et al. from the University of Toronto presented a generative model called protein variational autoencoder (ProteinVAE) that can design synthetic hexons with highly distinguished sequences using only 12.4 million parameters (Figure 3B).⁴³ Distinguishing from traditional autoencoder language models, ProteinVAE holds an extra part, the latent space, between encoder and decoder networks. This 'latent space' is a conceptual space where the model learns physical-chemical features from discrete protein sequence space and converts it to a continuous space for ease of sampling and manipulation (Figure 3B, left panel). Although being divergent from existing clades in sequences, generated hexons are indistinguishable from natural vector counterparts without disturbing their protein folding, particle assembly, or cell transducing functions (Figure 3B, middle panel). They further predicted *in silico* that the accessible hexon surfaces, targeted by neutralizing antibodies, varied significantly from natural hexons to the point where they could be classified as novel serotypes, potentially evading pre-existing immune responses (Figure 3B, right panel).

The aforementioned instances of AI-driven viral vectored vaccine innovations underscore the potential of deep neural network models to precisely predict capsid viability across diverse variants, even with a constrained training dataset. Besides novel capsid proteins designed for combating pre-Ad/AAV immunity, ML models that predict the assembly abilities of capsid variants into virus-like structures thrive, which is expected to extend the applicability in augmenting both the packaging capacity and transduction efficacy of AAV-based vaccines.^{89,90}

AI-based mRNA vaccine modification

Since the discovery of RNA-based gene delivery platforms, it has gained rapid development in vaccines and therapeutics. RNA vaccinology works by instructing the host cells to produce vaccine antigens using their protein production machinery based on the information provided by RNA sequences.⁹¹ The first mRNA vaccine for infectious diseases was reported for influenza A prevention in 2012.⁹² Both COVID-19 mRNA vaccines produced by Pfizer-BioNTech (BNT162b2) and Moderna (mRNA-1273) showed high safety and efficacy in preventing severe diseases and were approved for emergency use soon after the COVID-19 outbreak.^{93,94} More importantly, the ability to simultaneously induce cellular and humoral immune responses makes mRNA a promising vaccine platform.⁹⁵

The secondary structure stability and translation efficiency of mRNA vaccines are closely related to the nucleic acid sequences; however, there is always a trade-off between them during codon optimization.⁹⁶ Leppek et al. raised a combinatorial optimization of RNA-based therapeutics using three models concerning mRNA secondary structure, in-solution stability, and translation efficiency.⁹⁷ Besides the above advancement, LinearDesign is developed to optimize codon usage for efficient protein expression and, at the same time, improve mRNA half-life by decreasing free energy.⁹⁸ LinearDesign adapts the classical concept of lattice parsing in computational linguistics, looking for the most likely "sentence" among many similar-sounding alternatives (Figure 4A, left panel). This AI algorithm profoundly increases mRNA vaccines' immunogenicity compared to the codon-optimization benchmark for SARS-CoV-2 and varicella-zoster virus (Figure 4A, right panel).⁹⁸ This model, with its potential to revolutionize mRNA vaccine development, provides a timely tool for the exploration of previously unreachable but highly stable and efficient codon arrangements in mRNA vaccines and mRNA-based medicines.

While the coding region of mRNA is the primary source of information about the designed antigen, the non-coding area also plays a crucial role in translation regulation and ultimate antigen protein expression.⁹⁹ The 5'UTR sequence is critical in ribosome recruitment, directly impacting protein

production.¹⁰⁰ In this context, Chu et al. introduced a semi-supervised language model (UTR-LM).⁵¹ This model was pre-trained on a transformer-based model using endogenous 5' UTRs from multiple species via nucleotide masking and reconstruction and combination of supervised information, including the secondary structure and minimum free energy, enhances model predictive capability. At last, the trained model was fine-tuned for downstream tasks such as mean ribosome loading (MRL) prediction, mRNA expression level (EL), and translational efficiency (TE) prediction (Figure 4B, left panel). The UTR-LM model's performance in MRL prediction was particularly impressive, outperforming the benchmark method MTtrans by 5% in Spearman R value.⁹⁹ The top designs showed superior performance in mRNA translational efficiency and protein expression level relative to the currently well-established 5' UTR optimization benchmark method Cao-RF (Figure 4B, middle panel).¹⁰¹ They next expanded the usage of the training model to accurately identify unannotated internal ribosome entry sites (IRESs) (Figure 4B, middle panel). Of greater significance, they generated new 5'UTRs and further confirmed their high performance in translational efficiency via wet-lab assays (Figure 4B, right panel).

In addition to optimizing the mRNA sequence to increase its stability and translation efficiency, AI technology also participates in mRNA delivery system optimization.¹⁰² One of the most promising platforms for mRNA delivery refers to mRNA-encapsulated in lipid nanoparticle (mRNA-LNP).¹⁰³ Xu et al. developed a deep learning-based platform called AGILE that is utilized to design and evaluate novel ionizable lipids for mRNA delivery.¹⁰⁴ Notably, the generated LNPs tailored for specific cell types enhance the effectiveness of mRNA vaccines in clinical therapies. In another study, Witten et al. introduced a deep learning-based LNP optimization and discovery network LiON that improves mRNA delivery potency.¹⁰⁵ AI was also used to optimize the microfluidic conditions and lipid mix ratio of mRNA-LNP vaccine, improving the efficiency of vaccine manufacturing.¹⁰⁶

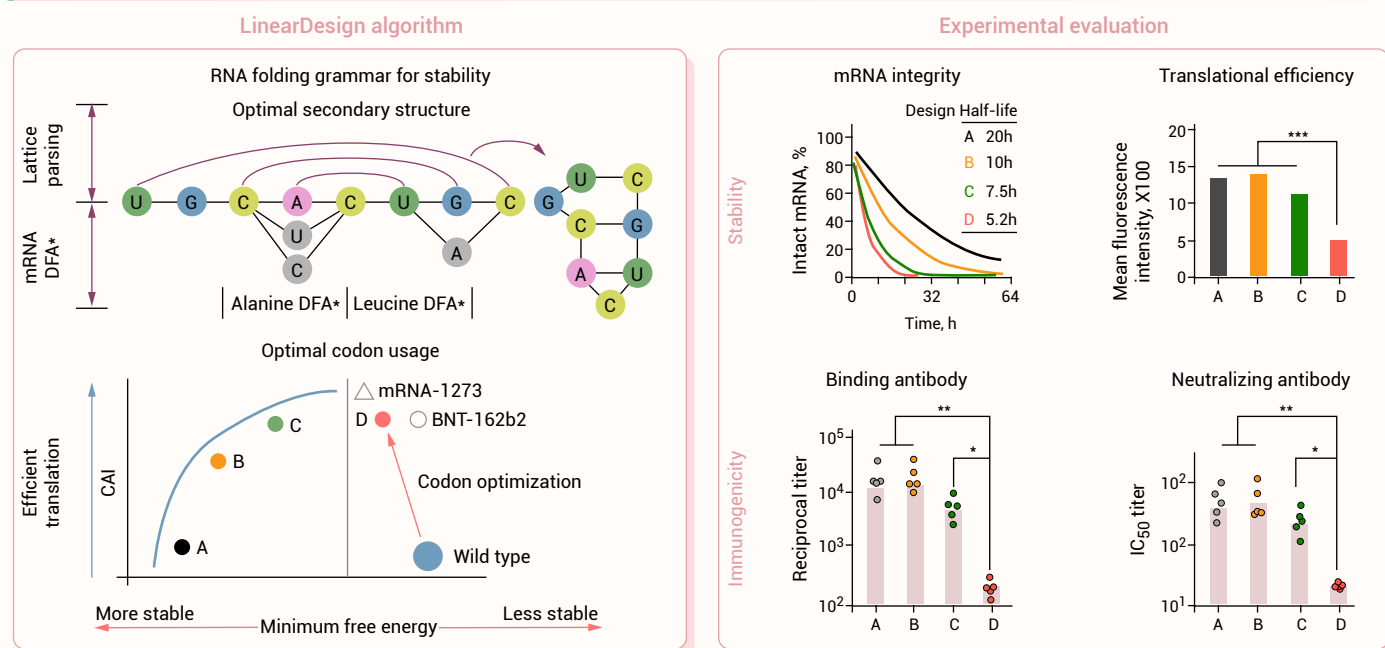
Other aspects

The utilization of AI tools in medical and biological science areas has significantly expedited the vaccine development.¹⁰⁷ On the one hand, AI has revolutionized our way of interpreting nucleotide and amino acid sequences, empowering scientists to harvest vast datasets for computational model training by integrating genomic, proteomic and clinical data. These advanced models not only predict potential B-cell and T-cell epitopes, but also annotate effective vaccine targets, decipher mechanisms of immune responses and optimize the vaccine formulations, and significantly reduce the time taken to screen new vaccine candidates in the lab.^{108,109} On the other hand, these techniques predict viral evolutionary trajectories, further identifying antigen-receptor interactions and are ready for the outbreak of newly emerging pathogens.¹¹⁰

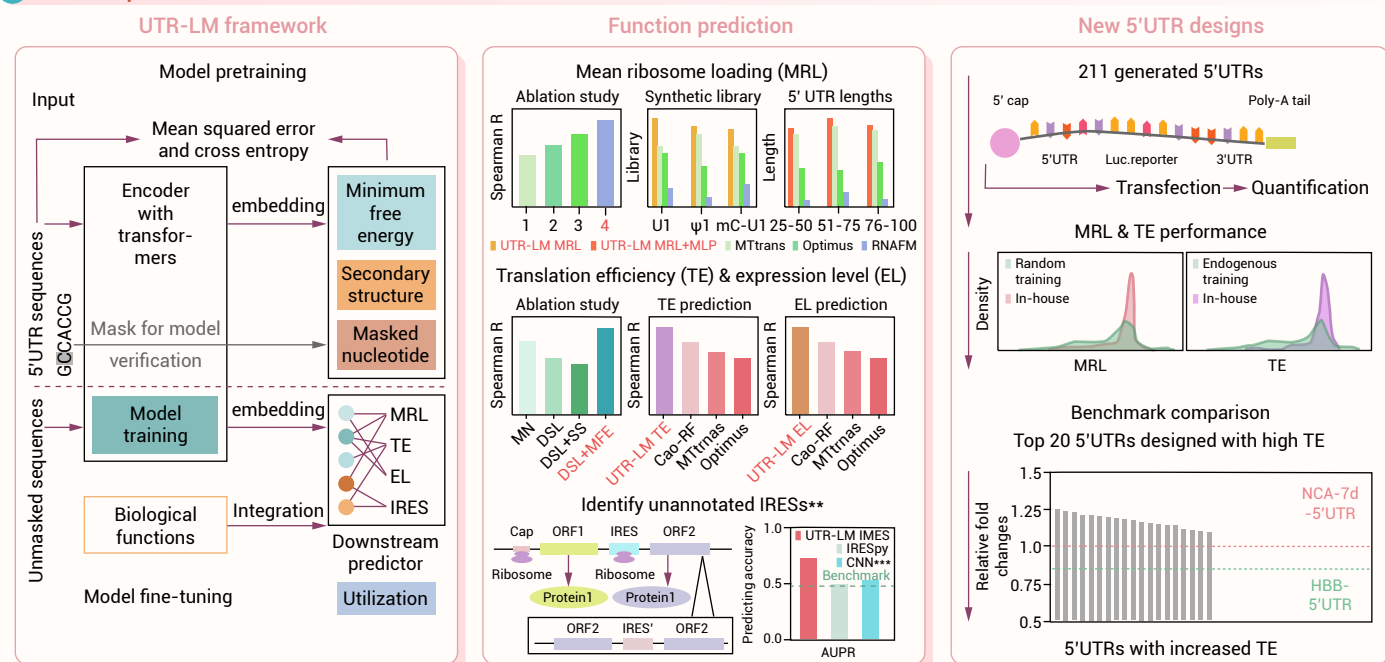
Moreover, machine learning models swiftly engage in vaccine pre-clinical and clinical trials. During the preclinical phase, several models like ImmScope, UniPMT and CapsNet, are developed to identify peptide-major histocompatibility complex class (MHC) pairing relationships, accurately predict whether an antigenic peptides can be recognized, presented to CD4⁺ T cells and subsequently trigger a genuine immune response.¹¹¹⁻¹¹³ Besides, AI models are employed to predict immune response following cancer treatment.^{114,115} We believe that AI will also be capable of participating in predicting the immune protection efficacy after vaccination in the coming future. For clinical trials, they can assist researchers in identifying suitable population samples, optimizing clinical trial designs, predicting trial outcomes, and improving the efficiency and success rate of trials based on a vast amount of clinical, experimental, and bioinformatic data. By integrating an individual's genomic information, the machine learning model can also identify unique neoantigens or autoantigens specific to that individual patient, which aids in the development of personalized vaccines and enhances their immunological efficacy.

As a key facilitator in enhancing vaccine efficacy, the identification of novel adjuvant candidates and the optimization of existing adjuvants are issues worthy of attention in the field of vaccine research and development. Machine learning and molecular simulations can aid in the design and synthesis of a library of small-molecule ligands with broad-spectrum toll-like receptor targeting and activation capabilities.¹¹⁶ Experimental results have revealed

A Codon optimization



B 5'UTR optimization



*DFA, deterministic finite-state automaton **IRESs, internal ribosome entry sites ***CNN, convolutional neural network

1 Random weights frozen, only predictor fine-tuned 2 Random initialization, all parameters fine-tuned 3 Pretrained model frozen, only predictor fine-tuned 4 Pretrained model, all parameters fine-tuned

Figure 4. AI-based mRNA vaccine modification (A) AI-assisted codon optimization. Left panel, overview of LinearDesign algorithm. The algorithm incorporates two factors: RNA folding grammar that increases structural stability and codon usage that facilitates mRNA expression. Scatter plot showing the balanced designs between efficient translation and structural stability. Right panel, experimental validation of stability and immunogenicity of mRNA vaccine after codon optimization. Intact mRNA ratio in the Mg^{2+} buffer was evaluated for mRNA chemical stability. Fluorescence intensity showing protein expression levels of mRNA was determined by flow cytometry 48h after transfection. Serum binding and neutralizing antibody titers indicating immunogenicity of mRNA vaccine behind each design were measured. (B) 5' UTR function predictions using the UTR-LM model. Left panel, 5'UTR sequences were used as training datasets and lateral model fine-tuning process. Biological characteristics of 5'UTR were also incorporated for downstream predictor tasks. Middle panel, the trained model was utilized to predict mean ribosomal loading (MRL), mRNA translation efficiency (TE), mRNA expression level (EL), and unannotated internal ribosome entry sites (IRESs) which is different from cap-dependent ribosome recruitment. Right panel, the model was also fine-tuned for novel 5'UTR generation. The generated sequences were then compared for the translation efficiency by wet-lab experiments. All the predicting functions derived from the UTR-LM model were compared with currently available benchmark methods and showed significant superiority over them.

that these novel gold nanoparticle adjuvants can enhance the antigen-presenting capacity of dendritic cells (DCs) and their ability to transport to lymph nodes. Besides the design of novel TLR-targeting agonist ligands, computational protein design of TLR3 minbinders can precisely target the

hydrophobic sites of TLR3, while ensuring formulation stability and avoiding glycosylation interference, thereby more efficiently driving type I interferon and NF- κ B responses.¹¹⁷ Utilizing computer-aided molecular docking and dynamics simulation methods, Cui et al. screened TLR4-targeting adjuvant

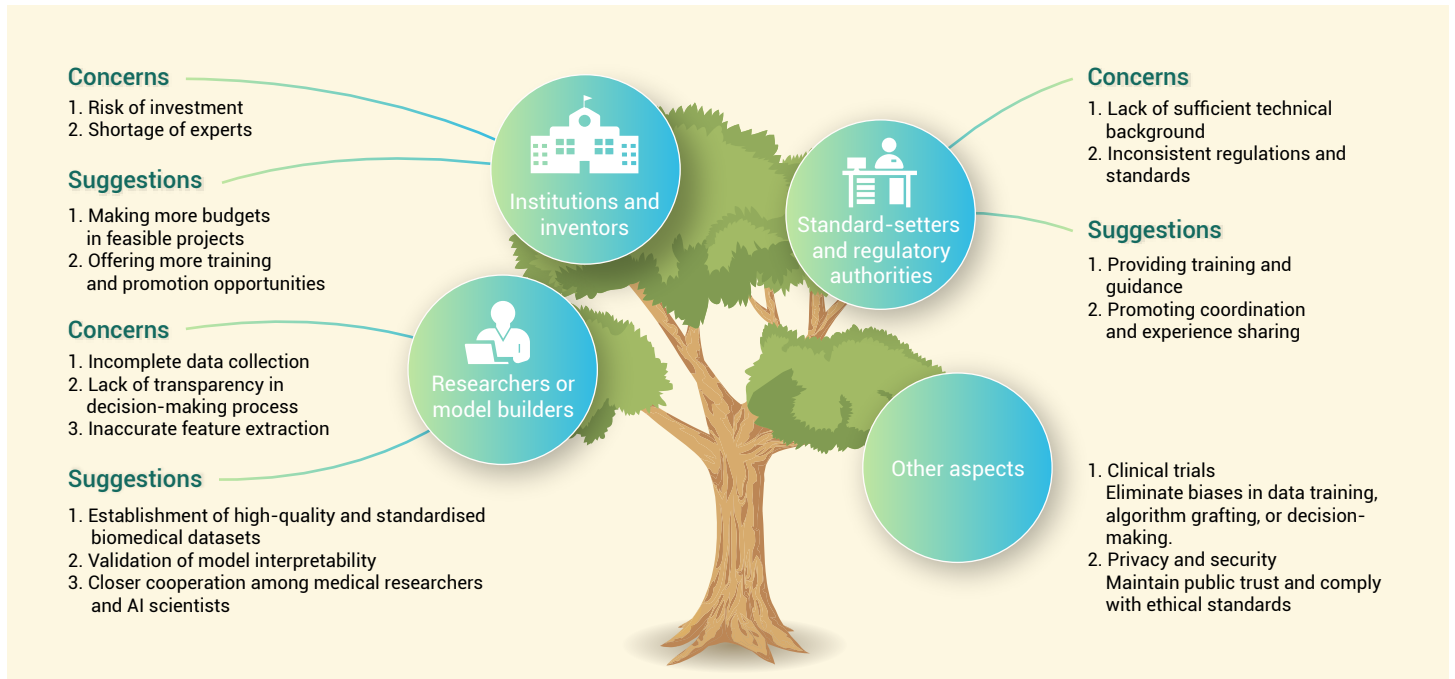


Figure 5. Concerns behind AI-driven vaccine development and proposed suggestions The concerns and recommendations were elaborated from three aspects: researchers and model builders, institutions and inventors, as well as standard-setters and regulatory authorities.

candidates from 34 chrysophanol derivatives.¹¹⁸ Both TLR3 minibinders and the TLR4-targeting adjuvant candidates exhibit high potential as vaccine adjuvants. More recently, a Notch agonist that promotes T cell development and enhances immunity has been *de novo* designed.¹¹⁹

CONCLUSION AND RECOMMENDATIONS

In this review, we listed successful examples of AI-powered vaccine developments, including *de novo* protein subunit vaccine designs, viral vector-based vaccine innovation, and mRNA-based vaccine sequence modifications. Although AI is transforming traditional life science experiments, enabling unprecedented speed and accuracy in diagnoses, personalized therapeutics, as well as vaccine development, there are still several concerns coming from different aspects. We outline some challenges and provide recommendations in response to them.

For institutions and investors, the study, development and application of AI technology require investment of time and money, but its commercial applications may face high risks and uncertainties. Therefore, they must weigh up the benefits and costs. As an emerging technology, we call on the government and investors to continue building on computing infrastructure and allocating more budgets to feasible projects in the AI field (Figure 5). Besides, there is a shortage of experts in the intersection of AI and biomedicine, making it difficult to form a high-level research and experimental development team. AI is a multidisciplinary field that draws on a range of technical and non-technical fields, including computer science, psychology, engineering, mathematics, and philosophy. Institutes and companies that rely highly on AI could offer more training and development opportunities and a better working environment to attract talent (Figure 5).

For researchers or model builders, the biomedical data that is currently accessible often fail to adequately represent the uniform, diversity and complexity of the population as a whole in certain study scenarios, which directly affects the application range of AI models (Figure 5). Given this situation, it is necessary to promote the establishment of high-quality, standardized biomedical datasets as well as data sharing. Besides, many AI models, such as unsupervised neural network models, lack transparency in their decision-making processes, which is particularly crucial in the vaccine development field. Validation of model interpretability through statistical analysis of the datasets or other method will help ensure successful implementation of AI tools into clinical practice (Figure 5). Third, researchers in computer science often struggle with which biological features to extract, which impacts their ability of integrating the latest research findings into language

models, so closer cooperation is required among biomedical researchers and AI scientists to generate models with high accuracy and efficiency during candidate vaccine designs (Figure 5).

For international standard-setters and regulatory authorities, existing regulations and standards may lag behind the development of AI technology (Figure 5). Regulatory authorities may lack sufficient technical background to fully understand and evaluate the feasibility of AI in vaccine design, and may not be able to regulate and guide the application of new technologies in a timely manner. In that vein, providing training and guidance on AI technology to regulators and establishing an expert advisory team to improve their understanding and evaluation of new technologies become a prerequisite. So that they can update and adjust relevant regulations according to the development of AI technology to ensure their adaptability and foresight. Indeed, the U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) has released a guidance for the evaluation of the safety and efficacy of AI-enabled devices used in medical products. Apart from that, vaccine development needs to comply with strict regulations and standards. These standards may differ between countries and regions, which may lead to difficulties in international cooperation and coordination. To avoid adding unnecessary complexity to innovation and acceptance, regulatory authorities could serve as a bridge to promote cooperation and coordination among countries in AI technology and vaccine supervision and share experience and best practices to promote global public health security (Figure 5).

The clinical trial is the most critical and challenging phase in vaccine development. The advent of AI technology holds promise in alleviating difficulties such as time-consuming processes, high costs, and challenges in volunteer recruitment. However, it also faces significant challenges, as data biases—including sampling bias, measurement bias, label bias, etc.—can affect data training and algorithm grafting, ultimately leading to biases in decision-making.¹²⁰ In enhancing AI's involvement in clinical trials, it is crucial to integrate human judgment to precisely define negative and positive outcomes for the model, instead of merely differential analysis of output data. Moreover, a rigorous dissection of racial bias in algorithms is indispensable to uphold fairness in the assessment of medical resources, such as sources of transplanted organs, thereby ensuring equitable healthcare access for diverse populations. Besides, during the nascent phase of AI-assisted clinical trials in particular, strict compliance with privacy and security protocols is important to maintain public trust and ensure compliance with ethical standards, since any violation could erode participant confidence and compromise the trial's integrity.

In conclusion, AI emerges as a highly promising tool in the *de novo* design and optimization of vaccines and thus has broad utilization potential. Larger training datasets and more mature models are required to effectively integrate diverse and personalized vaccine development approaches, ultimately enhancing human health.

REFERENCES

- Watson O. J., Barnsley G., Toor J., et al. (2022). Global impact of the first year of COVID-19 vaccination: A mathematical modelling study. *Lancet Infect. Dis.* **22**:1293–1302. DOI:10.1016/S1473-3099(22)00320-6
- Minor P. D. (2015). Live attenuated vaccines: Historical successes and current challenges. *Virology* **479-480**:379–392. DOI: 10.1016/j.virol.2015.03.032.
- Hill H. A., Elam-Evans L. D., Yankey D., et al. (2018). Vaccination coverage among children aged 19–35 months – United States, 2017. *MMWR. Morbidity and mortality weekly report*. **67**:1123–1128. DOI:10.15585/mmwr.mm6740a4
- Organization W. H. (2024). Global immunization efforts have saved at least 154 million lives over the past 50 years. <https://www.who.int/news/item/24-04-2024-global-immunization-efforts-have-saved-at-least-154-million-lives-over-the-past-50-years>
- Krishnamoorthy Y., Sakthivel M., Eliyas S. K., et al. (2019). Worldwide trend in measles incidence from 1980 to 2016: A pooled analysis of evidence from 194 WHO Member States. *J Postgrad. Med.* **65**:160–163. DOI:10.4103/jpgm.JPGM_508_18
- Jin L., Li Z., Zhang X., et al. (2022). CoronaVac: A review of efficacy, safety, and immunogenicity of the inactivated vaccine against SARS-CoV-2. *Hum. Vaccin. Immunother.* **18**:2096970. DOI:10.1080/21645515.2022.2096970
- Hu L., Sun J., Wang Y., et al. (2023). A review of inactivated COVID-19 vaccine development in China: Focusing on safety and efficacy in special populations. *Vaccines (Basel)* **11**:1045. DOI: 10.3390/vaccines11061045
- Heidary M., Kaviar V. H., Shirani M., et al. (2022). A comprehensive review of the protein subunit vaccines against COVID-19. *Front. Microbiol.* **13**:927306. DOI:10.3389/fmicb.2022.927306
- Richmond P., Hatchuel L., Dong M., et al. (2021). Safety and immunogenicity of S-Trimer (SCB-2019), a protein subunit vaccine candidate for COVID-19 in healthy adults: A phase 1, randomised, double-blind, placebo-controlled trial. *Lancet.* **397**:682–694. DOI:10.1016/S0140-6736(21)00241-5
- Travieso T., Li J., Mahesh S., et al. (2022). The use of viral vectors in vaccine development. *NPJ Vaccines*. **7**:75. DOI:10.1038/s41541-022-00503-y
- Vanaparthi R., Mohan G., Vasireddy D., et al. (2021). Review of COVID-19 viral vector-based vaccines and COVID-19 variants. *Infez. Med.* **29**:328–338. DOI:10.53854/liim-2903-3
- Draper S. J. and Heeney J. L. (2010). Viruses as vaccine vectors for infectious diseases and cancer. *Nat. Rev. Microbiol.* **8**:62–73. DOI:10.1038/nrmicro2240
- Wang S., Liang B., Wang W., et al. (2023). Viral vectored vaccines: Design, development, preventive and therapeutic applications in human diseases. *Signal Transduct. Target. Ther.* **8**:149. DOI:10.1038/s41392-023-01408-5
- Folegatti P. M., Ewer K. J., Aley P. K., et al. (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: A preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet.* **396**:467–478. DOI:10.1016/S0140-6736(20)31604-4
- Sadoff J., Le Gars M., Shukarev G., et al. (2021). Interim results of a phase 1-2a trial of Ad26.COV2.S Covid-19 vaccine. *N. Engl. J. Med.* **384**:1824–1835. DOI: 10.1056/NEJMoa2034201.
- Zhu F. C., Li Y. H., Guan X. H., et al. (2020). Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: A dose-escalation, open-label, non-randomised, first-in-human trial. *Lancet.* **395**:1845–1854. DOI:10.1016/S0140-6736(20)31208-3
- Li J. X., Hou L. H., Gou J. B., et al. (2023). Safety, immunogenicity and protection of heterologous boost with an aerosolised Ad5-nCoV after two-dose inactivated COVID-19 vaccines in adults: A multicentre, open-label phase 3 trial. *Lancet Infect. Dis.* **23**:1143–1152. DOI:10.1016/S1473-3099(23)00350-X
- Wang S. Y., Liu W. Q., Li Y. Q., et al. (2023). A China-developed adenovirus vector-based COVID-19 vaccine: Review of the development and application of Ad5-nCov. *Expert. Rev. Vaccines*. **22**:704–713. DOI:10.1080/14760584.2023.2242528
- Qin F., Xia F., Chen H., et al. (2021). A guide to nucleic acid vaccines in the prevention and treatment of infectious diseases and cancers: From basic principles to current applications. *Front. Cell Dev. Biol.* **9**:633776. DOI:10.3389/fcell.2021.633776
- Pardi N., Hogan M. J., Porter F. W., et al. (2018). mRNA vaccines – A new era in vaccinology. *Nat. Rev. Drug Discov.* **17**:261–279. DOI:10.1038/nrd.2017.243
- Chaudhary N., Weissman D. and Whitehead K. A. (2021). mRNA vaccines for infectious diseases: Principles, delivery and clinical translation. *Nat. Rev. Drug Discov.* **20**:817–838. DOI:10.1038/s41573-021-00283-5
- Ghosh A., Larrondo-Petrie M. M. and Pavlovic M. (2023). Revolutionizing vaccine development for COVID-19: A review of AI-based approaches. *Information* **14**:665. DOI:10.3390/info14120665
- McCarthy J. (2007). What is artificial intelligence. *Stanford University*. <http://www-formal.stanford.edu/jmc/>
- Krenn M., Pollice R., Guo S. Y., et al. (2022). On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**:761–769. DOI:10.1038/s42254-022-00518-3
- Shinde P. P. and Shah S. (2018). A review of machine learning and deep learning applications. 2018 Fourth international conference on computing communication control and automation (ICCUBEA). DOI:10.1109/ICCUBEA.2018.8697857
- Rajoub B. (2020). Supervised and unsupervised learning. Zgallai, W. (ed). In *Biomedical Signal Processing and Artificial Intelligence in Healthcare*. (Academic Press), pp: 51–89. DOI: 10.1016/B978-0-12-818946-7.00003-2
- Bravi B. (2024). Development and use of machine learning algorithms in vaccine target selection. *NPJ Vaccines* **9**:15. DOI:10.1038/s41541-023-00795-8
- Hederman A. P. and Ackerman M. E. (2023). Leveraging deep learning to improve vaccine design. *Trends Immunol.* **44**:333–344. DOI:10.1016/j.it.2023.03.002
- Senior A. W., Evans R., Jumper J., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* **577**:706–710. DOI:10.1038/s41586-019-1923-7
- Du Z., Su H., Wang W., et al. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* **16**:5634–5651. DOI:10.1038/s41596-021-00628-9
- Baek M., DiMaio F., Anishchenko I., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* **373**:871–876. DOI:10.1126/science.abc8754
- He H., He B., Guan L., et al. (2024). De novo generation of SARS-CoV-2 antibody CDRH3 with a pre-trained generative large language model. *Nat. Commun.* **15**:6867. DOI:10.1038/s41467-024-50903-y
- Chakraborty C., Bhattacharya M. and Lee S. S. (2023). Artificial intelligence enabled ChatGPT and large language models in drug target discovery, drug discovery, and development. *Mol. Ther. Nucleic Acids* **33**:866–868. DOI:10.1016/j.omtn.2023.08.009
- Madani A., Krause B., Greene E. R., et al. (2023). Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**:1099–1106. DOI:10.1038/s41587-022-01618-2
- Jablonka K. M., Schwaller P., Ortega-Guerrero A., et al. (2024). Leveraging large language models for predictive chemistry. *Nat. Mach. Intell.* **6**:161–169. DOI:10.1038/s42256-023-00788-1
- Khakzad H., Igashov I., Schneuing A., et al. (2023). A new age in protein design empowered by deep learning. *Cell Syst.* **14**:925–939. DOI:10.1016/j.cels.2023.10.006
- Peng C., Yang X., Chen A., et al. (2023). A study of generative large language model for medical research and healthcare. *NPJ Digit. Med.* **6**:210. DOI:10.1038/s41746-023-00958-w
- Chu A. E., Lu T. and Huang P. S. (2024). Sparks of function by de novo protein design. *Nat. Biotechnol.* **42**:203–215. DOI:10.1038/s41587-024-02133-2
- Awad M. and Khanna R. (2015). Efficient learning machines. Pepper J., Weiss S. and Hauke P. (eds). Theories, concepts, and applications for engineers and system designers (Apress), pp: 1–18. DOI:10.1007/978-1-4302-5990-9.
- Lutz I. D., Wang S., Norn C., et al. (2023). Top-down design of protein architectures with reinforcement learning. *Science* **380**:266–273. DOI:10.1126/science.adf6591
- Correia B. E., Ban Y. E., Holmes M. A., et al. (2010). Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* **18**:1116–1126. DOI:10.1016/j.str.2010.06.010
- Gm H., Gourisaria M. K., Pandey M., et al. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review.* **38**:100285. DOI:10.1016/j.cosrev.2020.100285
- Lyu S., Sowlati-Hashjin S. and Garton M. (2024). Variational autoencoder for design of synthetic viral vector serotypes. *Nat. Mach. Intell.* **6**:147–160. DOI:10.1038/s42256-023-00787-2
- Watson J. L., Juergens D., Bennett N. R., et al. (2023). De novo design of protein structure and function with RFdiffusion. *Nature* **620**:1089–1100. DOI:10.1038/s41586-023-06415-8
- Abramson J., Adler J., Dunger J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**:493–500. DOI:10.1038/s41586-024-07487-w
- Jumper J., Evans R., Pritzel A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. DOI:10.1038/s41586-021-03819-2
- Thirunavukarasu A. J., Ting D. S. J., Elangovan K., et al. (2023). Large language models in medicine. *Nat. Med.* **29**:1930–1940. DOI:10.1038/s41591-023-02448-8
- Ferruz N., Schmidt S. and Höcker B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **13**:4348. DOI:10.1038/s41467-022-32007-7
- Guntuboina C., Das A., Mollaei P., et al. (2023). PeptideBERT: A language model based on transformers for peptide property prediction. *J. Phys. Chem. Lett.* **14**:10427–10434. DOI:10.1021/acs.jpclett.3c02398
- Lin Z., Akin H., Rao R., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**:1123–1130. DOI:10.1126/science.ade2574
- Chu Y., Yu D., Li Y., et al. (2024). A 5' UTR language model for decoding untranslated regions of mRNA and function predictions. *Nat. Mach. Intell.* **6**:449–460. DOI:10.1038/s42256-024-00823-9
- Chen B., Cheng X., Li P., et al. (2025). xTrimoPGLM: Unified 100-billion-parameter

- pretrained transformer for deciphering the language of proteins. *Nat. Methods*. **22**:1028–1039. DOI:10.1038/s41592-025-02636-z
53. Brixi G., Durrant M. G., Ku J., et al. (2025). Genome modeling and design across all domains of life with Evo 2. *bioRxiv*.2025.02.18.638918. DOI: 10.1101/2025.02.18.638918
 54. Nguyen E., Poli M., Durrant M. G., et al. (2024). Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**:ead9336. DOI:10.1126/science.ado9336
 55. Singhal K., Azizi S., Tu T., et al. (2023). Large language models encode clinical knowledge. *Nature* **620**:172–180. DOI:10.1038/s41586-023-06291-2
 56. Zheng S., Li Y., Chen S., et al. (2020). Predicting drug–protein interaction using quasi-visual questionanswering system. *Nat. Mach. Intell.* **2**:134–140. DOI:10.1038/s42256-020-0152-y
 57. Malone B., Simovski B., Moline C., et al. (2020). Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2 leading to universal blueprints for vaccine designs. *Sci. Rep.* **10**:22375. DOI:10.1038/s41598-020-78758-5
 58. Kanampalliar A. M. (2020). Reverse vaccinology and its applications. *Methods Mol. Biol.* **2131**:1–16. DOI:10.1007/978-1-0716-0389-5_1
 59. Santos-Júnior C. D., Torres M. D. T., Duan Y., et al. (2024). Discovery of antimicrobial peptides in the global microbiome with machine learning. *Cell* **187**:3761–3778.e3716. DOI:10.1016/j.cell.2024.05.013
 60. Correia B. E., Bates J. T., Loomis R. J., et al. (2014). Proof of principle for epitope-focused vaccine design. *Nature* **507**:201–206. DOI:10.1038/nature12966
 61. Castro K. M., Watson J. L., Wang J., et al. (2024). Accurate single domain scaffolding of three non-overlapping protein epitopes using deep learning. *bioRxiv*.2024.05.07.592871. DOI: 10.1101/2024.05.07.592871
 62. Fu D., Wang W., Zhang Y., et al. (2024). Self-assembling nanoparticle engineered from the ferritinophagy complex as a rabies virus vaccine candidate. *Nat. Commun.* **15**:8601. DOI:10.1038/s41467-024-52908-z
 63. Song J. Y., Choi W. S., Heo J. Y., et al. (2022). Safety and immunogenicity of a SARS-CoV-2 recombinant protein nanoparticle vaccine (GBP510) adjuvanted with AS03: A randomised, placebo-controlled, observer-blinded phase 1/2 trial. *EClinicalMedicine*. **51**:101569. DOI:10.1016/j.eclinm.2022.101569
 64. Song J. Y., Choi W. S., Heo J. Y., et al. (2023). Immunogenicity and safety of SARS-CoV-2 recombinant protein nanoparticle vaccine GBP510 adjuvanted with AS03: Interim results of a randomised, active-controlled, observer-blinded, phase 3 trial. *EClinicalMedicine* **64**:102140. DOI:10.1016/j.eclinm.2023.102140
 65. Brouwer P. J. M., Antanasijevic A., Berendsen Z., et al. (2019). Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. *Nat. Commun.* **10**:4272. DOI:10.1038/s41467-019-12080-1
 66. Walls A. C., Fiala B., Schafer A., et al. (2020). Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**:1367–1382.e1317. DOI:10.1016/j.cell.2020.10.043
 67. Jacob-Dolan C., Yu J., McMahan K., et al. (2023). Immunogenicity and protective efficacy of GBP510/AS03 vaccine against SARS-CoV-2 delta challenge in rhesus macaques. *NPJ Vaccines* **8**:23. DOI:10.1038/s41541-023-00622-0
 68. Ueda G., Antanasijevic A., Fallas J. A., et al. (2020). Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**:e57659. DOI: 10.7554/eLife.57659
 69. Khattak M. A. A., Ascierto P. A., Queirolo P., et al. (2024). 1084P Phase II study of AI-designed personalized neoantigen cancer vaccine, EVX-01, in combination with pembrolizumab in advanced melanoma. *Ann. Oncol.* **35**:S718–S719. DOI:10.1016/j.annonc.2024.08.1152
 70. Lisanza S. L., Gershon J. M., Tipps S. W. K., et al. (2024). Multistate and functional protein design using RoseTTAFold sequence space diffusion. *Nat. Biotechnol.* DOI: 10.1038/s41587-024-02395-w
 71. Liang Y., Shao S., Li X. Y., et al. (2024). Mutating a flexible region of the RSV F protein can stabilize the prefusion conformation. *Science* **385**:1484–1491. DOI:10.1126/science.adp2362
 72. Che Y., Gribenko A. V., Song X., et al. (2023). Rational design of a highly immunogenic prefusion-stabilized F glycoprotein antigen for a respiratory syncytial virus vaccine. *Sci. Transl. Med.* **15**:ead6422. DOI:10.1126/scitranslmed.ade6422
 73. Bakkers M. J. G., Cox F., Koormeef A., et al. (2024). A foldon-free prefusion F trimer vaccine for respiratory syncytial virus to reduce off-target immune responses. *Nat. Microbiol.* **9**:3254–3267. DOI:10.1038/s41564-024-01860-1
 74. Bakkers M. J. G., Ritschel T., Tiemessen M., et al. (2024). Efficacious human metapneumovirus vaccine based on AI-guided engineering of a closed prefusion trimer. *Nat. Commun.* **15**:6270. DOI:10.1038/s41467-024-50659-5
 75. Williams J. A., Biancucci M., Lessen L., et al. (2023). Structural and computational design of a SARS-CoV-2 spike antigen with improved expression and immunogenicity. *Sci. Adv.* **9**:eadg0330. DOI:10.1126/sciadv.adg0330
 76. Zabaleta N., Dai W., Bhatt U., et al. (2021). An AAV-based, room-temperature-stable, single-dose COVID-19 vaccine provides durable immunogenicity and protection in non-human primates. *Cell Host Microbe* **29**:1437–1453.e1438. DOI: 10.1016/j.chom.2021.08.002
 77. Zhao S., Ke J., Yang B., et al. (2022). A protective AAV vaccine for SARS-CoV-2. *Signal Transduct. Target. Ther.* **7**:310. DOI:10.1038/s41392-022-01158-w
 78. Ploquin A., Szecsi J., Mathieu C., et al. (2013). Protection against henipavirus infection by use of recombinant adeno-associated virus-vector vaccines. *J. Infect. Dis.* **207**:469–478. DOI:10.1093/infdis/jis699
 79. Smalley E. (2017). First AAV gene therapy poised for landmark approval. *Nat. Biotechnol.* **35**:998–999. DOI:10.1038/nbt1117-998
 80. Russell S., Bennett J., Wellman J. A., et al. (2017). Efficacy and safety of voretigene neparovector (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet*. **390**:849–860. DOI:10.1016/S0140-6736(17)31868-8
 81. Grimm D., Lee J. S., Wang L., et al. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J. Virol.* **82**:5887–5911. DOI:10.1128/JVI.00254-08
 82. Dalkara D., Byrne L. C., Klimczak R. R., et al. (2013). In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. *Sci. Transl. Med.* **5**:189ra176. DOI:10.1126/scitranslmed.3005708
 83. Tse L. V., Klinc K. A., Madigan V. J., et al. (2017). Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proc. Natl. Acad. Sci. USA* **114**:E4812–E4821. DOI:10.1073/pnas.1704766114
 84. Bryant D. H., Bashir A., Sinai S., et al. (2021). Deep diversification of an AAV capsid protein by machine learning. *Nat. Biotechnol.* **39**:691–696. DOI:10.1038/s41587-020-00793-4
 85. Mendonca S. A., Lorincz R., Boucher P., et al. (2021). Adenoviral vector vaccine platforms in the SARS-CoV-2 pandemic. *NPJ Vaccines*. **6**:97. DOI:10.1038/s41541-021-00356-x
 86. Fausther-Bovendo H. and Kobinger G. P. (2014). Pre-existing immunity against Ad vectors: humoral, cellular, and innate response, what's important. *Hum. Vaccin. Immunother.* **10**:2875–2884. DOI:10.4161/hv.29594
 87. Rux J. J., Kuser P. R. and Burnett R. M. (2003). Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. *J. Virol.* **77**:9553–9566. DOI:10.1128/jvi.77.17.9553-9566.2003
 88. Flickinger J. C., Jr., Singh J., Carlson R., et al. (2020). Chimeric Ad5.F35 vector evades anti-adenovirus serotype 5 neutralization opposing GUCY2C-targeted antitumor immunity. *J. Immunother. Cancer* **8**. DOI: 10.1136/jitc-2020-001046
 89. Zhu D., Brookes D. H., Busia A., et al. (2024). Optimal trade-off control in machine learning-based library design, with application to adeno-associated virus (AAV) for gene therapy. *Sci. Adv.* **10**:eadj3786. DOI:10.1126/sciadv.adj3786
 90. Marques A. D., Kummer M., Kondratov O., et al. (2021). Applying machine learning to predict viral assembly for adeno-associated virus capsid libraries. *Mol. Ther. Methods Clin. Dev.* **20**:276–286. DOI:10.1016/j.omtm.2020.11.017
 91. Lee J., Woodruff M. C., Kim E. H., et al. (2023). Knife's edge: Balancing immunogenicity and reactogenicity in mRNA vaccines. *Exp. Mol. Med.* **55**:1305–1313. DOI:10.1038/s12276-023-00999-x
 92. Petsch B., Schnee M., Vogel A. B., et al. (2012). Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection. *Nat. Biotechnol.* **30**:1210–1216. DOI:10.1038/nbt.2436
 93. Polack F. P., Thomas S. J., Kitchin N., et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**:2603–2615. DOI:10.1056/NEJMoa2034577
 94. Baden L. R., El Sahly H. M., Essink B., et al. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**:403–416. DOI:10.1056/NEJMoa2035389
 95. Park J. W., Lagniton P. N. P., Liu Y., et al. (2021). mRNA vaccines for COVID-19: What, why and how. *Int. J. Biol. Sci.* **17**:1446–1460. DOI:10.7150/ijbs.59233
 96. Paremskaia A. I., Kogan A. A., Murashkina A., et al. (2024). Codon-optimization in gene therapy: Promises, prospects and challenges. *Front. Bioeng. Biotechnol.* **12**:1371596. DOI:10.3389/fbioe.2024.1371596
 97. Leppik K., Byeon G. W., Kladwang W., et al. (2022). Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat. Commun.* **13**:1536. DOI:10.1038/s41467-022-28776-w
 98. Zhang H., Zhang L., Lin A., et al. (2023). Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* **621**:396–403. DOI:10.1038/s41586-023-06127-z
 99. Zheng W., Fong J. H. C., Wan Y. K., et al. (2023). Discovery of regulatory motifs in 5' untranslated regions using interpretable multi-task learning models. *Cell Syst.* **14**:1103–1112.e1106. DOI: 10.1016/j.cels.2023.10.011
 100. Hinnebusch A. G., Ivanov I. P. and Sonenberg N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**:1413–1416. DOI:10.1126/science.aad9868
 101. Cao J., Novoa E. M., Zhang Z., et al. (2021). High-throughput 5' UTR engineering for enhanced protein production in non-viral gene therapies. *Nat. Commun.* **12**:4138. DOI:10.1038/s41467-021-24436-7
 102. Yuan Y., Wu Y., Cheng J., et al. (2024). Applications of artificial intelligence to lipid nanoparticle delivery. *Particology* **90**:88–97. DOI:10.1016/j.partic.2023.11.014
 103. Alameh M. G., Tombacz I., Bettini E., et al. (2022). Lipid nanoparticles enhance the efficacy of mRNA and protein subunit vaccines by inducing robust T follicular helper cell and humoral responses. *Immunity* **55**:1136–1138. DOI:10.1016/j.immuni.2022.05.007
 104. Xu Y., Ma S., Cui H., et al. (2024). AGILE platform: A deep learning powered approach

- to accelerate LNP development for mRNA delivery. *Nat. Commun.* **15**:6305. DOI:10.1038/s41467-024-50619-z
105. Witten J., Raji I., Manan R. S., et al. (2024). Artificial intelligence-guided design of lipid nanoparticles for pulmonary gene therapy. *Nat. Biotechnol.* DOI: 10.1038/s41587-024-02490-y
 106. Maharjan R., Kim K. H., Lee K., et al. (2024). Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. *Journal of Pharmaceutical Analysis*.100996. DOI: 10.1016/j.jpha.2024.100996
 107. Gao S., Fang A., Huang Y., et al. (2024). Empowering biomedical discovery with AI agents. *Cell* **187**:6125–6151. DOI:10.1016/j.cell.2024.09.022
 108. Wang G., Liu X., Wang K., et al. (2023). Deep-learning-enabled protein-protein interaction analysis for prediction of SARS-CoV-2 infectivity and variant evolution. *Nat. Med.* **29**:2007–2018. DOI:10.1038/s41591-023-02483-5
 109. Liu T., Shi K. and Li W. (2020). Deep learning methods improve linear B-cell epitope prediction. *BioData Min.* **13**:1. DOI:10.1186/s13040-020-00211-0
 110. Han W., Chen N., Xu X., et al. (2023). Predicting the antigenic evolution of SARS-CoV-2 with deep learning. *Nat. Commun.* **14**:3478. DOI:10.1038/s41467-023-39199-6
 111. Shen L.-C., Zhang Y., Wang Z., et al. (2025). Self-iterative multiple-instance learning enables the prediction of CD4⁺T cell immunogenic epitopes. *Nat. Mach. Intell.* DOI: 10.1038/s42256-025-01073-z
 112. Zhao Y., Yu J., Su Y., et al. (2025). A unified deep framework for peptide-major histocompatibility complex-T cell receptor binding prediction. *Nat. Mach. Intell.* **7**:650–660. DOI:10.1038/s42256-025-01002-0
 113. Kalematis M., Darvishi S. and Koohi S. (2023). CapsNet-MHC predicts peptide-MHC class I binding based on capsule neural networks. *Commun. Biology.* **6**:492. DOI:10.1038/s42003-023-04867-2
 114. Chang T. G., Cao Y., Sfreddo H. J., et al. (2024). LORIS robustly predicts patient outcomes with immune checkpoint blockade therapy using common clinical, pathologic and genomic features. *Nat. Cancer* **5**:1158–1175. DOI:10.1038/s43018-024-00772-7
 115. Jiang Y., Zhang Z., Wang W., et al. (2023). Biology-guided deep learning predicts prognosis and cancer immunotherapy response. *Nat. Commun.* **14**:5135. DOI:10.1038/s41467-023-40890-x
 116. Ma J., Wang S., Zhao C., et al. (2023). Computer-aided discovery of potent broad-spectrum vaccine adjuvants. *Angew. Chem. Int. Ed.* **62**:e202301059. DOI:10.1002/anie.202301059
 117. Adams C. S., Kim H., Burtner A. E., et al. (2025). De novo design of protein minibinder agonists of TLR3. *Nat. Commun.* **16**:1234. DOI:10.1038/s41467-025-56369-w
 118. Cui Z., Shi C., An R., et al. (2025). In silico-guided discovery of polysaccharide derivatives as adjuvants in nanoparticle vaccines for cancer immunotherapy. *ACS Nano* **19**:2099–2116. DOI:10.1021/acsnano.4c08898
 119. Mout R., Jing R., Tanaka-Yano M., et al. (2025). Design of soluble Notch agonists that drive T cell development and boost immunity. *Cell* DOI: 10.1016/j.cell.2025.07.009
 120. Cross J. L., Choma M. A. and Onofrey J. A. (2024). Bias in medical AI: Implications for clinical decision-making. *PLOS Digit. Health.* **3**:e0000651. DOI:10.1371/journal.pdig.0000651

FUNDING AND ACKNOWLEDGMENTS

This work is partially founded by the National Natural Science Foundation of China (32571110) and the Young Elite Scientists Sponsorship Program by CAST (2023QNRC001). We thank Prof. Junjie Xu, Prof. Lihua Hou and Dr. Enhao Ma for their helpful suggestions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

F.X.Z. wrote the initial manuscript. X.D.Z. revised the manuscript. Both authors discussed, edited and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

No new data were created or analyzed in this study.