AI-Driven Design of Universal Peptide Adjuvants for Cost-Effective Vaccines

SARA AL-FLITI[1], AMJAD ALQAHTANI[1], SHAHLA ALSHAAR[1], FATIMAH ALHASSAN[1], and YOMNA ALI[1]

[1]KAUST Academy, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Supervisor: Dr. Kerven Durdymyradov.

ABSTRACT Vaccine adjuvants are essential for enhancing immune responses, yet their discovery has traditionally relied on slow, costly, and largely empirical methods that offer limited adaptability to emerging pathogens. Addressing this challenge, we present a novel AI-driven in silico framework, developed from scratch, for the rational design of peptide-based adjuvants (PBAs) targeting Toll-like receptor 4 (TLR4), a central mediator of innate immune activation. A custom protein language model, PeptideGPT (non-hemolytic), was built and fine-tuned on curated datasets of immunogenic, non-toxic peptides. Candidate sequences were generated and systematically filtered through a multi-objective evaluation pipeline incorporating MHC-I binding prediction, antigenicity scoring, allergenicity screening, and toxicity assessment. From this process, seven top-ranked peptides were identified, each predicted to be antigenic, non-toxic, and non-allergenic. Compared to conventional approaches, the proposed computational pipeline demonstrates greater efficiency, enhanced safety profiling, and scalability in design. This work establishes a new paradigm for the accelerated development of safe and adaptable adjuvants through fully digital methodologies.

KEYWORDS Adjuvant, PeptideGPT, Vaccine, TLR4

I. INTRODUCTION

DEVELOPING safe, effective, and affordable vaccines remains a pressing global health priority, particularly in the face of viral threats such as Zika and Hepatitis C (HCV). These pathogens continue to cause considerable illness and mortality, especially in low- and middle-income countries [1]. Traditional vaccine platforms, often based on whole inactivated or attenuated pathogens, are constrained by safety concerns, scalability issues, and challenges in maintaining immunological stability. To address these limitations, synthetic peptide-based vaccines have emerged as an attractive alternative due to their safety, production efficiency, and design flexibility. However, such vaccines often exhibit low immunogenicity and short half-lives, making the use of potent adjuvants indispensable [2] [3]. Adjuvants act as immunostimulatory agents that enhance antigen recognition and promote durable immune responses. Designing peptide-based adjuvants (PBAs) for clinical application remains highly challenging, as they must simultaneously induce robust immune activation while avoiding toxicity, allergenicity, or autoimmune risks [4].

Conventional adjuvant discovery, which relies heavily on trial-and-error screening of natural compounds, recombinant expression, or chemical synthesis, is time-consuming, costly, and offers limited control over molecular structure and overall design flexibility, making it unsuitable for rapid responses to emerging infectious diseases [5]. To overcome these limitations, this project proposes a novel AI-guided approach for designing PBAs that target Toll-like receptor 4 (TLR4), a key immune receptor involved in triggering both innate and adaptive immune responses. TLR4 plays a central role in immunological signaling, making it a promising target for developing broad-spectrum, pathogen-independent adjuvants [4]. Importantly, multiple clinically relevant viruses—including Dengue, Ebola, SARS-CoV-2, Influenza, HCV, HIV, and ZIKV—have been shown to exploit or engage TLR4 pathways, driving pathogenic inflammation and immune dysfunction, as summarized in (Table 1). By focusing on a common immune pathway, the approach allows the development of PBAs that can support a wide range of vaccine platforms, especially for rapidly emerging or mutating viruses [3].

Leveraging recent advances in protein language models and immunoinformatics, this project introduces a generative AI framework for the in silico design of novel immunostimulatory peptides that are both effective and safe [6]. The computational pipeline simulates peptide generation, evaluation,

and optimization entirely in silico, eliminating the need for resource-intensive laboratory screening [7] By integrating predictive tools for immune activation, structural stability, and safety profiling, the approach reduces the time and cost of adjuvant discovery compared to traditional laboratory methods [8]. This represents a major advancement in vaccine development, significantly enhancing scalability, efficiency, and responsiveness to emerging infectious diseases [9] [10]. Although universally applicable, the platform was initially validated using high-priority case studies—Zika virus and Hepatitis C virus (HCV)—due to their significant global health burden, high mutation rates, and immune evasion mechanisms. These examples demonstrated the model's potential to generate broad-spectrum, adaptable adjuvants across diverse pathogens [11]–[13].

A. PROJECT OBJECTIVES

The primary objective of this project is to develop a scalable, cost-effective, and AI-driven in silico framework for the rapid design of peptide-based adjuvants (PBAs) targeting Toll-like receptor 4 (TLR4). By accelerating development timelines and enabling precision vaccine design, this work aims to support the creation of next-generation vaccines that are accessible, adaptable, and capable of addressing both current and future infectious disease threats.

II. RELATED WORK

A. ADJUVANT DISCOVERY

Traditional adjuvants such as aluminum salts and oil-in-water emulsions have been widely employed in licensed vaccines. Aluminum salts are generally regarded as safe and well-tolerated; however, they exhibit limited specificity and tend to bias immune responses toward particular pathways [14]. Oil-in-water emulsions, while capable of enhancing potency, are often associated with increased reactogenicity. These limitations have prompted the exploration of more advanced protein-based adjuvants, such as flagellin, which has been incorporated into fusion constructs like VAX128A/B/C [15]. Despite their promise, empirical strategies for adjuvant development continue to face significant challenges related to scalability, dose-dependent toxicity, and limited design flexibility [16].

B. PROTEIN LANGUAGE MODELS IN IMMUNOLOGY

Recent advances in protein language models (PLMs) such as **ProGen**, **ProtGPT2**, and **RITA** have enabled the generation of biologically plausible protein sequences in an autoregressive manner [17]. These models, typically trained on large-scale datasets such as UniRef, have demonstrated strong capabilities in general protein sequence generation. While ProGen and RITA highlight the potential of PLMs for broad protein design tasks, their applications remain largely generic. In contrast, our work leverages this paradigm by fine-tuning a custom model, PeptideGPT(non-hemolytic), on immune-relevant peptide sequences curated from the Immune Epitope Database (IEDB). This domain-specific adaptation allows for the targeted generation of non-toxic, immunogenic peptide adjuvants optimized for vaccine development.

C. TLR/PRR AGONISM AND SYNTHETIC BIOLOGY

Pattern recognition receptors (PRRs), particularly Toll-like receptors (TLRs) such as TLR4, TLR5, and TLR9, are well-established targets for vaccine adjuvant development [18]. Among these, TLR4 is one of the most clinically validated, with licensed adjuvants such as monophosphoryl lipid A (MPLA) incorporated into vaccines including Cervarix, Shingrix, and RTS,S. TLR4's unique ability to activate both MyD88 and TRIF signaling pathways enables a balanced induction of Th1 and humoral immune responses, positioning it as a universal innate immune receptor applicable across diverse pathogens [19]. Despite decades of research, effective vaccines for the Hepatitis C virus (HCV) and Zika virus (ZIKV) remain elusive or suboptimal. These pathogens pose distinct immunological challenges—chronic infection in the case of HCV, and teratogenic effects for ZIKV [20] [21].

Current vaccine candidates often lack potent, safe adjuvants tailored to these viral properties. Emerging evidence indicates that TLR4 agonists hold significant promise as adjuvants for unlicensed viral vaccines. For ZIKV, although direct trials remain limited, flavivirus studies—particularly with West Nile virus—have shown that formulations combining TLR4 agonists with saponins can elicit sterilizing immunity, broaden antibody repertoires, and induce balanced Th1/Th2 responses, suggesting translational potential for Zika [22] [23]. In HCV, preclinical studies using MPL, a clinically approved TLR4 agonist, have demonstrated marked improvements in antibody titers and Th1-skewed cellular immunity [24] [25]. Collectively, these findings position ZIKV and HCV as strong candidates for TLR4-targeted peptide adjuvant development, offering a pathway to boost protective immunity in the absence of licensed vaccines. Advances in synthetic biology and computational immunology further strengthen this approach. Tools such as TLRpred, motif scanning, and in silico docking enable early prediction of PRR–ligand interactions, reducing experimental costs and timelines [26]. Our pipeline builds on this foundation by integrating TLR4-focused PRR activation filtering directly into the post-generation scoring stage—a critical validation layer that is absent in most PLM-driven peptide studies.

D. MULTI-OBJECTIVE OPTIMIZATION AND EVALUATION

Most computational approaches to adjuvant discovery have traditionally focused on screening existing sequences with predefined scoring tools. Few have integrated end-to-end pipelines capable of balancing critical parameters such as MHC binding, TCR recognition, innate immune activation, structural stability, and toxicity. To address this, our framework builds upon established tools including NetMHCpan, VaxiJen, DeepImmuno, and other immunogenicity scoring models, enabling candidate ranking and refinement through a multi-objective optimization strategy [27]–[30].

Recent reviews highlight this gap, consistently noting that systematic, multi-objective computational approaches for adjuvant design remain underdeveloped and urgently needed to achieve parity with the sophistication of epitope-based vaccine design [31]–[33]. To the best of our knowledge, this work represents the first application of PLM-driven peptide generation to adjuvant discovery, incorporating a fully integrated multi-objective evaluation pipeline. By combining generative AI with rigorous immunoinformatics filtering, the framework offers a path toward designing universal peptide adjuvants tailored for emerging viral threats and contributing to the development of therapeutic vaccines.

III. METHODOLOGY

This project established a baseline end-to-end computational workflow, as shown in (Figure 1). for the in silico design of novel peptide adjuvants targeting Toll-like receptor 4 (TLR4). All experiments were conducted in a GPU-accelerated Jupyter environment, with reproducibility ensured through the use of fixed random seeds, version-controlled dependencies, and structured logging of all intermediate results.

A. DATA ACQUISITION AND FILTERING

A total of 35,000 peptide sequences were collected from the **Immune Epitope Database (IEDB)** and supplemented with 102,000 background sequences obtained from **UniRef90**. To maximize biological relevance, strict filtering criteria were applied to the IEDB dataset. Only **linear peptides** were retained, as these are both compatible with transformer-based sequence modeling and practical for chemical synthesis. Furthermore, only **T-cell assays with positive outcomes** were selected to ensure the inclusion of experimentally validated immunogenic peptides.

The dataset was further restricted to **Homo sapiens** epitopes and limited to **MHC Class I-restricted peptides**, given the central role of CD8 cytotoxic T-cell responses in antiviral immunity and adjuvant function. A minimum peptide length of **eight amino acids** was imposed, as shorter sequences lack stable MHC-I binding. To provide negative contrast during training, background "non-epitope" peptides of matched lengths were sampled from UniRef90.

This multi-step filtering process ensured that the final corpus was **biologically validated, human-specific, and directly relevant to immune activation**, providing a robust foundation for downstream model training and evaluation.

B. MODEL SELECTION AND FINE-TUNING

The generative backbone employed was **PeptideGPT non-hemolytic**, a transformer-based protein language model pretrained on safe, non-hemolytic peptide sequences. This design choice minimized the likelihood of generating hemolytic (toxic) candidates while maintaining broad coverage of peptide sequence diversity. To further guide model learning, two special tokens—**<TLR4>** and **<BG>**—were introduced to label immunogenic and background examples, thereby conditioning the model to distinguish target-relevant peptides.

The curated dataset was partitioned into **training (90%)** and **validation (10%)** subsets. Fine-tuning was conducted using the **Hugging Face Transformers** library in **PyTorch** with GPU acceleration. **Mixed precision training** (fp16/bf16) was applied to improve computational efficiency. Hyperparameters such as learning rate, batch size, and learning-rate scheduler were optimized iteratively. Model performance was evaluated using **validation** loss and **perplexity**, which served as primary convergence indicators.

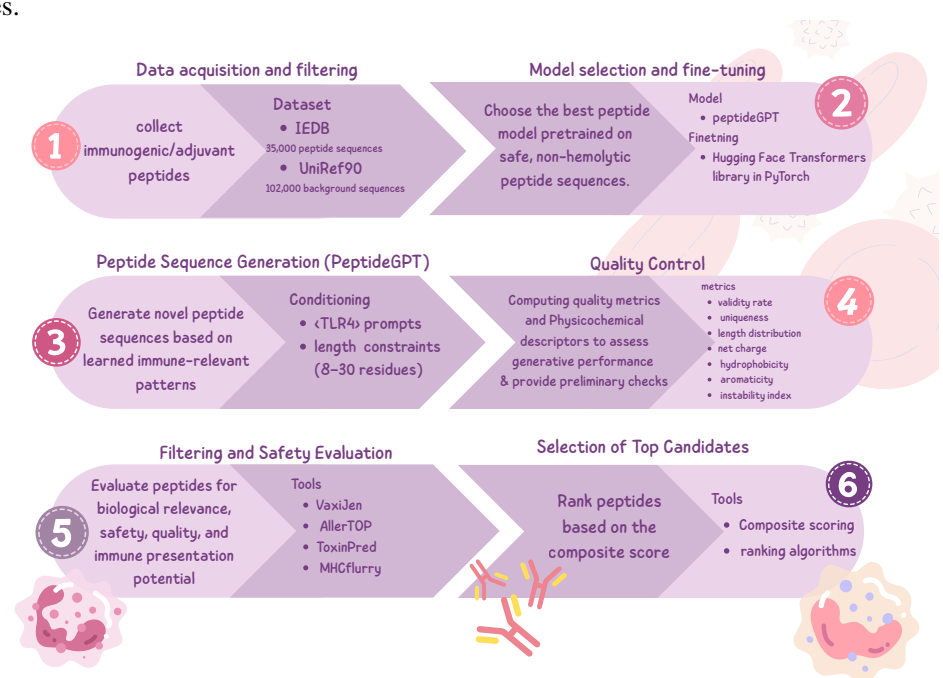


Figure 1. Illustration of the AI-driven pipeline for peptide adjuvant design.

Table 1: Examples of viral diseases engaging TLR4

Virus / Disease	Viral or Host Factor Engaging TLR4	Cell Types Involved	Reported Consequence	Vaccine Cost (per dose)
Dengue virus (DENV, Dengue fever)	Secreted NS1 protein, oxidized phospholipids	Endothelial cells, macrophages	Vascular leak, cytokine storm, endothelial dysfunction	$\approx$ \$66 (Dengvaxia, US CDC private sector [34])
Ebola virus (EBOV, Ebola virus disease)	Shed glycoprotein (GP)	Monocytes, macrophages, endothelial cells	Hyperinflammation, vascular permeability, shock	\$98.60 (ERVEBO, UNICEF [35])
SARS-CoV-2 (COVID-19)	Spike protein (S1), DAMPs (HMGB1, oxidized lipids)	Monocytes, alveolar macrophages, lung epithelial cells	Cytokine storm, lung inflammation	\$136.75−141.80 (Pfizer/Moderna, US CDC [36])
Influenza A virus (IAV, flu)	Viral proteins (M1, NP), oxidized phospholipids	Alveolar macrophages, epithelial cells	Acute lung injury, cytokine release, ARDS	\$19.48−73.36 (seasonal flu vaccines, CDC [37])
Respiratory Syncytial Virus (RSV, bronchiolitis)	Fusion protein (F glycoprotein)	Airway epithelial cells, macrophages	Inflammatory cytokine induction, bronchiolitis severity	\$290−306.80 (Arexvy, Abrysvo, mRESVIA, CDC [38])
Human cytomegalovirus (HCMV)	Viral glycoproteins; induces TLR4 expression	Dendritic cells, monocytes	Proinflammatory cytokine production, immune activation	N/A — no licensed vaccine [39]
Japanese Encephalitis Virus (JEV)	NS1/NS1 proteins, viral RNA byproducts	Microglia, neurons, astrocytes	Neuroinflammation, NF-κB activation, neuronal damage	\$290−366 (IXIARO, US retail [40])
West Nile virus (WNV)	Viral envelope proteins, DAMPs	Microglia, neurons	Neuroinflammation, BBB disruption	N/A — no licensed human vaccine [41]
Hepatitis C virus (HCV)	Core protein, NS5A interactions	Monocytes, hepatocytes, dendritic cells	Enhanced TLR4 signaling, liver inflammation, fibrosis	N/A — no licensed vaccine [42]
HIV-1 (AIDS)	gp120, Tat, viral RNA	Monocytes, dendritic cells, astrocytes	Chronic immune activation, neuroinflammation	N/A — no licensed vaccine [43]
Herpes simplex virus (HSV-1/2)	Envelope glycoproteins, DAMP release	Keratinocytes, dendritic cells, neurons	TLR4-mediated NF-κB activation, inflammation	N/A — no licensed vaccine [44]
Zika virus (ZIKV)	NS1 protein (like DENV)	Endothelial cells, placental macrophages	Vascular leakage, placental inflammation	N/A — no licensed vaccine [45]

C. PEPTIDE GENERATION AND QUALITY CONTROL

The fine-tuned model was conditioned with **<TLR4> prompts** to generate novel candidate peptides. Post-processing steps included removal of non-canonical amino acids, enforcement of length constraints **(8–30 residues)**, and elimination of duplicate sequences. To evaluate generative performance, quality metrics such as **validity rate**, **uniqueness**, and **length distribution** were computed.

In addition, physicochemical descriptors—including **net charge**, **hydrophobicity**, **aromaticity**, and the **instability index**—were calculated to provide preliminary checks of biological plausibility prior to downstream immunological evaluation.

D. POST-GENERATION FILTERING AND EVALUATION

Following generation, candidate peptides were filtered to exclude sequences with extreme physicochemical properties—such as excessive net charge, highly hydrophobic composition, or long homopolymer runs. This step reduced the risk of aggregation and removed sequences unlikely to fold or function correctly. The filtered set was subsequently evaluated using three independent **safety** and **immunogenicity** screening tools:

- **ToxinPred** — predicted peptide toxicity, enabling elimination of potentially harmful sequences.
- **VaxiJen** — computed antigenicity scores to identify peptides with a higher probability of stimulating immune responses.
- **AllerTOP** — assessed allergenic potential, removing candidates predicted to cause allergenic reactions.

Only peptides simultaneously classified as non-toxic, antigenic, and non-allergenic were advanced to the final evaluation stage.

E. MHC-I BINDING PREDICTION AND 9-MER REFINEMENT

To assess compatibility with human immune presentation, peptides were evaluated using **MHCflurry**. This tool predicts binding affinity of peptide fragments to a panel of common **MHC-I alleles**. Since MHC-I molecules typically present **9-mer cores**, MHCflurry automatically scanned longer parent peptides and **identified their most favorable 9-mer subsequences**. These predicted 9-mers were scored for binding strength (reported in nanomolar affinity, with lower values indicating stronger binding) and for presentation likelihood. This refinement step not only improved biological realism but also broadened **population coverage**, as candidate peptides could be matched to globally prevalent HLA alleles.

F. RANKING AND REPORTING

The final candidate pool was scored across multiple dimensions: **antigenicity** (VaxiJen), **safety** (ToxinPred and AllerTOP), and **MHC-I binding/presentation** (MHCflurry). Results were normalized and aggregated into **composite ranking system**, prioritizing peptides that combined strong antigenicity, safety, and favorable HLA presentation profiles.

The top-ranked peptides, both as full parent sequences and refined 9-mer cores, were exported as final outputs.

G. GOVERNANCE AND SAFETY

Safety was embedded throughout the pipeline. A **non-hemolytic generative model** reduced inherent toxicity risks, while filtering with **ToxinPred** and **AllerTOP** ensured exclusion of unsafe candidates. Restriction to human, experimentally validated, positive T-cell epitopes further improved relevance and minimized noise. All computational results remain **in silico predictions**; therefore, shortlisted peptides require **wet-lab validation** through binding, cytokine release, and functional immune assays before consideration for use as adjuvants.

To further optimize **MHC-I compatibility**, broaden coverage across human populations, and reduce synthesis costs, the parent peptides were trimmed into **9-mer windows**. MHCflurry predictions highlighted several high-affinity binders, particularly to HLA-A*02:01 and HLA-B*27:05.

IV. RESULTS

Using the fine-tuned **PeptideGPT_non_hemolytic** model conditioned on <TLR4>, a total of **50 novel peptide sequences** were generated. Initial quality control revealed that approximately **90% were predicted to be non-toxic** by ToxinPred, with the majority also scoring as antigenic (VaxiJen) and non-allergenic (AllerTOP). This demonstrated that the model successfully biased generation toward safe and immunogenic candidates.

From this pool, peptides underwent sequential filtering for physicochemical plausibility (charge, hydrophobicity, length constraints), redundancy removal, and evaluation through immunoinformatics tools. The filtered peptides were then ranked by composite criteria incorporating antigenicity, safety, and predicted MHC-I presentation.

Seven top candidates (Seq1–Seq7) emerged from this ranking. These represented diverse profiles — including balanced all-round peptides (Seq1), highly cationic immune stimulators (Seq2, Seq5), hydrophobic sequences with potential membrane interaction (Seq6), and short stable anchor peptides (Seq3, Seq7). All top candidates were non-toxic, antigenic, and non-allergenic, meeting the essential requirements for further development as adjuvants. The sequences, along with their physicochemical descriptors and safety evaluations, are summarized in (Table 2,3).

The full results and the complete set of generated potential peptide adjuvants are provided in the supplementary files available at the linked GitHub repository.

V. DISCUSSION

The pipeline successfully generated a manageable set of **50 candidate peptides**, of which **90 percent passed initial toxicity and allergenicity filters**, validating the effectiveness of the PeptideGPT non-hemolytic model in biasing generation toward safe adjuvants. From this, filtering and ranking produced **seven top candidates** with complementary biochemical profiles.

The diversity of these peptides is notable. Highly cationic sequences (Seq2, Seq5) are predicted to strongly stimulate immune responses through electrostatic interactions with receptor complexes, while hydrophobic candidates (Seq6) may mimic natural pathogen-associated peptides by embedding into membranes and facilitating TLR4 signaling. Shorter peptides (Seq3, Seq7) demonstrate minimalistic yet stable scaffolds, which are advantageous for synthesis and modular vaccine design. Seq1, meanwhile, emerged as the most balanced candidate across all properties, making it an excellent all-rounder.

The trimming of peptides to **9-mers** further enhanced biological plausibility by aligning with MHC-I presentation requirements. Several 9-mers demonstrated strong predicted binding to globally prevalent alleles such as HLA-A02:01, with nanomolar affinities (e.g., FIFFAIFYI, VLSFIFFAI), suggesting potential for broad immunogenic coverage. LRMRLMFGK and KWYILRMRL added allele diversity with predicted binding to HLA-B27:05 and HLA-A*24:02 respectively. Together, these results indicate that the combination of generative modeling, toxicity/allergenicity filtering, antigenicity scoring, and MHC-I binding predictions can yield biologically plausible adjuvant candidates. The final shortlist balances **safety, immunogenicity, and presentation feasibility**, making these peptides strong contenders for next-generation adjuvant design.

VI. CHALLENGES AND OBSTACLES

A. MODEL SIZE AND GPU REQUIREMENTS

The chosen model, **PeptideGPT (non-hemolytic)**, is based on a large Transformer architecture with millions of parameters. Fine-tuning such a model required access to high-end GPUs, specifically NVIDIA A100s, as smaller GPUs (e.g., Tesla V100) often ran out of memory during training. Even with gradient checkpointing and mixed-precision training enabled, fitting long peptide sequences into the model proved challenging. Managing memory allocation and preventing CUDA out-of-memory errors became a recurring obstacle, especially during training runs with larger batch sizes.

Table 2: Final Top Candidates (parent peptides)

ID	Sequence	Length (aa)	Charge	Hydrophobicity	Mol. Wt. (Da)	ToxinPred	VaxiJen	AllerTOP	Key Notes
seq1	FTLRAVLSFIFFAIF YICLSTTFR	24	+2	0.17 (balanced)	2877.8	Non-Toxin	Antigen	Non-Allergen	Strongest all-round candidate; balanced, antigenic, safe
seq2	VTLRMTMRMAM RTAMRTARRTAIR IAMRIA	30	+8	-0.29 (polar)	3538.9	Non-Toxin	Antigen	Non-Allergen	Highly cationic, strong immune stimulation potential
seq3	MTLRMNFP	8	+1	-0.12 (polar)	1009.4	Non-Toxin	Antigen	Non-Allergen	Very short, minimalistic, stable anchor peptide
seq4	ITLRARLRRLGEP DPV	15	+2	-0.34 (polar)	1862.4	Non-Toxin	Antigen	Non-Allergen	Arginine/Lysine-rich; likely TLR4 interaction
seq5	ITLRIKFIKKLSLLN KMIILK	21	+6	-0.06 (neutral)	2527.7	Non-Toxin	Antigen	Non-Allergen	Strongly cationic & antigenic, safe profile
seq6	ITLRALGALFVVV GLIFIWSPVF	23	+1	0.31 (hydrophobic)	2532.5	Non-Toxin	Antigen	Non-Allergen	Hydrophobic; may insert into membranes for TLR4
seq7	VTLRARLTE	9	+1	-0.29 (polar)	1058.4	Non-Toxin	Antigen	Non-Allergen	Small, flexible, antigenic, safe

Table 3: Top-Ranked 9-mers from Parent Candidates

Rank	Peptide (9mer)	Antigenicity	Best Allele	Affinity (nM, lower = stronger)	Presentation Score
1	LRMRLMFGK	0.95	HLA-B*27:05	84.9	0.51
2	KWYILRMRL	0.95	HLA-A*24:02	200.3	0.35
3	FIFFAIFYI	0.85	HLA-A*02:01	12.8	0.87
4	VLSFIFFAI	0.85	HLA-A*02:01	14.8	0.85
5	SLLNKMIIL	0.73	HLA-A*02:01	52.6	0.91
6	LNKMIILK	0.73	HLA-A*02:01	97.7	0.87
7	TLRMNFP	0.65	HLA-B*07:02	188.5	0.41

B. TRAINING AND GENERATION TIME

Fine-tuning was computationally expensive and time-consuming. Each training run required several hours of continuous GPU use, with validation checks further extending runtime. Candidate generation was also non-trivial: producing thousands of peptides with sampling and subsequent filtering took additional hours, slowing the overall workflow. This technical bottleneck limited the ability to perform multiple training iterations, experiment with alternative hyperparameters, or test larger datasets. As a result, optimizing the pipeline for both efficiency and reliability was a key challenge throughout the project.

VII. CONCLUSION

This work presents a **novel AI-driven framework, developed entirely from scratch**, for the rational design of peptide-based adjuvants targeting TLR4. By fine-tuning the custom protein language model **PeptideGPT (non-hemolytic)** on curated immune-relevant datasets and integrating a multi-objective evaluation pipeline, the study demonstrated that generative modeling can be systematically combined with immunoinformatics to produce biologically plausible and safe peptide candidates.

From 50 generated sequences, the pipeline distilled a **shortlist of seven top-ranked adjuvants**, each predicted to be non-toxic, non-allergenic, and antigenic, with strong potential for broad HLA presentation. These candidates exhibited complementary biochemical features—ranging from cationic to hydrophobic to balanced profiles—offering multiple modes of interaction with TLR4 and mimicking natural immune activation pathways. Importantly, refinement into 9-mer cores enhanced biological realism by aligning with MHC-I presentation and expanded population coverage across globally prevalent HLA alleles.

Beyond the immediate candidates identified, this study establishes a generalizable, scalable, and cost-effective digital workflow that addresses long-standing limitations of empirical adjuvant discovery. By embedding safety, immunogenicity, and receptor-specific presentation into a unified computational framework, this work sets a foundation for next-generation vaccine design strategies that are adaptable to rapidly emerging viral threats.

While experimental validation remains essential, the results underscore the transformative potential of combining protein language models with multi-objective optimization in vaccine research. Ultimately, this framework offers a new paradigm for accelerating adjuvant discovery, paving the way toward safer, faster, and more adaptable vaccines with global impact.

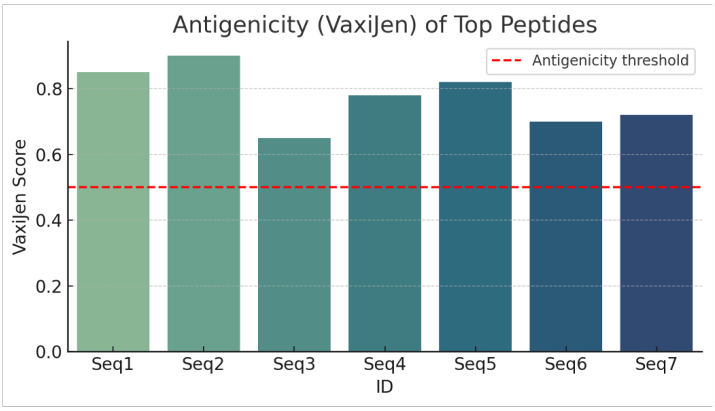


Figure 2

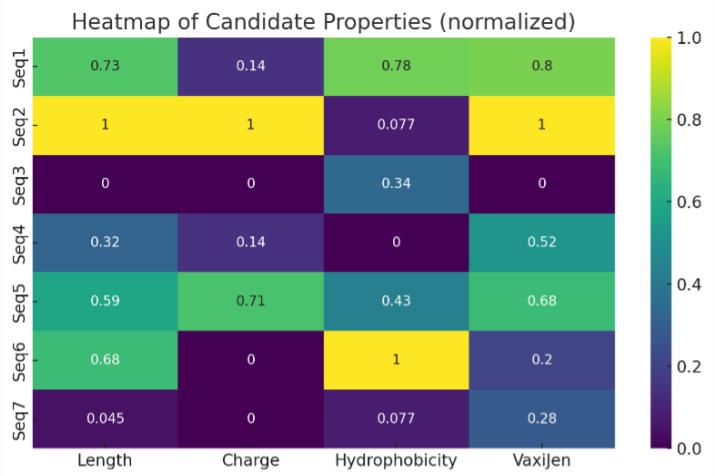


Figure 3

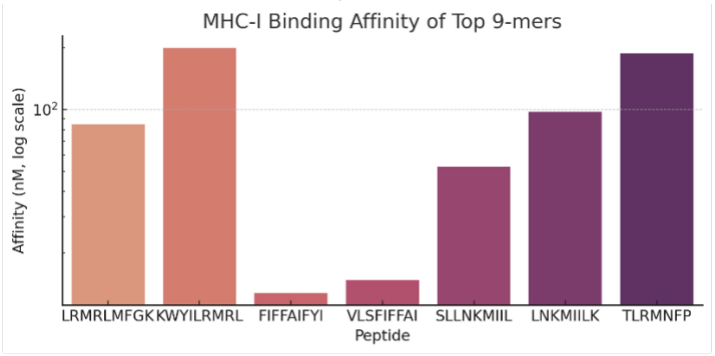


Figure 4

VIII. FUTURE WORK

Future efforts will focus on **structural and experimental validation** of the predicted peptide adjuvants. Planned computational analyses include **docking simulations** with the **TLR4/MD-2 complex** and **molecular dynamics simulations** to assess structural stability and binding persistence. These in *silico* studies will be complemented by **wet-lab assays**, including *in vitro* immune activation experiments and *in vivo* preclinical studies, to confirm biological efficacy. Collectively, these steps will be critical for advancing the shortlisted candidates from computational predictions toward **clinical applicability**.

IX. CODE SOURCE AND SUPPLEMENTARY FILES

https://github.com/sarafliti/project-kaustAI.git

References

[1] A. Farlow, E. Torreele, G. Gray, K. Ruxrungtham, H. Rees, S. Prasad et al., "The future of epidemic and pandemic vaccines to serve global public health needs," Vaccines, vol. 11, no. 3, p. 690, 2023. [Online]. Available: https://www.mdpi.com/2076-393X/11/3/690/htm

[2] F. Polack, S. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart et al., "Safety and efficacy of the bnt162b2 mrna covid-19 vaccine," New England Journal of Medicine, vol. 383, no. 27, pp. 2603–2615, 2020. [Online]. Available: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2034577

[3] M. Skwarczynski and I. Toth, "Peptide-based synthetic vaccines," Chemical Science, vol. 7, no. 2, p. 842, 2015. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5529997/

[4] G. Del Giudice, R. Rappuoli, and A. Didierlaurent, "Correlates of adjuvanticity: A review on adjuvants in licensed vaccines," Seminars in Immunology, vol. 39, pp. 14–21, 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29801750/

[5] S. Lee and M. Nguyen, "Recent advances of vaccine adjuvants for infectious diseases," Immune Network, vol. 15, no. 2, pp. 51–57, 2015.

[6] T. Chen, M. Dumas, R. Watson, S. Vincoff, C. Peng, L. Zhao et al., "Pepmlm: Target sequence-conditioned generation of therapeutic peptide binders via span masked language modeling," arXiv preprint arXiv:2310.03842, 2023, [cited 2025 Aug 18]. [Online]. Available: https://arxiv.org/pdf/2310.03842

[7] A. Shah, C. Guntuboina, and A. B. Farimani, "Peptide-gpt: Generative design of peptides using generative pre-trained transformers and bioinformatic supervision," arXiv preprint arXiv:2410.19222, 2024, [cited 2025 Aug 18]. [Online]. Available: https://arxiv.org/pdf/2410.19222

[8] D. B. Olawade, J. Teke, O. Fapohunda, K. Weerasinghe, S. O. Usman, A. O. Ige et al., "Leveraging artificial intelligence in vaccine development: A narrative review," Journal of Microbiological Methods, vol. 224, p. 106998, 2024, [cited 2025 Aug 18]. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167701224001106

[9] W. Y. Zhang, X. L. Zheng, P. S. Coghi, J. H. Chen, B. J. Dong, and X. X. Fan, "Revolutionizing adjuvant development: harnessing ai for next-generation cancer vaccines," Frontiers in Immunology, vol. 15, p. 1438030, 2024, [cited 2025 Aug 18]. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11349682/

[10] S. S. Rawat, A. K. Keshri, R. Kaur, and A. Prasad, "Immunoinformatics approaches for vaccine design: A fast and secure strategy for successful vaccine development," Vaccines, vol. 11, no. 2, p. 221, 2023, [cited 2025 Aug 18]. [Online]. Available: https://www.mdpi.com/2076-393X/11/2/221/htm

[11] M. Alharbi, A. Alshammari, J. F. Alsabhan, S. I. Alzarea, T. Alshammari, F. Alasmari et al., "A novel vaccine construct against zika virus fever: insights from epitope-based vaccine discovery through molecular modeling and immunoinformatics approaches," Frontiers in Immunology, vol. 15, p. 1426496, Jul 2024, [cited 2025 Aug 18].

[12] W. Ezzemani, M. P. Windisch, H. Altawalah, F. Guessous, R. Saile, S. Benjelloun et al., "Design of a multi-epitope zika virus vaccine candidate–an in-silico study," Journal of Biomolecular Structure and Dynamics, vol. 41, no. 9, pp. 3762–3771, 2023, [cited 2025 Aug 18]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/35318896/

[13] S. Zubair, F. Parvaiz, T. Abualait, K. Al-Regaiey, T. Anwar, M. Zafar et al., "Computational design of multi-epitope vaccine against hepatitis c virus infection using immunoinformatics techniques," PLoS One, vol. 20, no. 1, 2025, [cited 2025 Aug 18]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/39854342/

[14] S. Reed, M. Orr, and C. Fox, "Key roles of adjuvants in modern vaccines," Nature Medicine, vol. 19, no. 12, pp. 1597–1608, 2013.

[15] N. Petrovsky, "Comparative safety of vaccine adjuvants: A summary of current evidence and future needs," Drug Safety, vol. 38, no. 11, pp. 1059–1074, 2015.

[16] R. Desikan, M. Germani, P. van der Graaf, and M. Magee, "A quantitative clinical pharmacology-based framework for model-informed vaccine development," Journal of Pharmaceutical Sciences, vol. 113, no. 1, pp. 22–32, 2024.

[17] A. Madani, B. Krause, E. Greene, S. Subramanian, B. Mohr, J. Holton et al., "Large language models generate functional protein sequences

across diverse families," Nature Biotechnology, vol. 41, no. 8, pp. 1099–1106, 2023.
[18] N. Garçon, M. Van Mechelen, and M. Wettendorff, "Development and evaluation of as04, a novel and improved adjuvant system containing mpl and aluminum salt," in Immunopotentiators in Modern Vaccines, 2005, pp. 161–177.
[19] C. Chen, C. Zhang, R. Li, Z. Wang, Y. Yuan, H. Li et al., "Monophosphoryl-lipid a (mpla) is an efficacious adjuvant for inactivated rabies vaccines," Viruses, vol. 11, no. 12, p. 1118, 2019.
[20] D. O'Hagan, R. Lodaya, and G. Lofano, "The continued advance of vaccine adjuvants – 'we can work it out.'," Seminars in Immunology, vol. 50, p. 101430, 2020.
[21] D. Johnson, "Tlr4 agonists as vaccine adjuvants: A chemist's perspective," Expert Review of Vaccines, vol. 12, no. 7, pp. 711–713, 2013.
[22] Y. Wang, L. Ling, Z. Zhang, and A. Marin-Lopez, "Current advances in zika vaccine development," Vaccines, vol. 10, no. 11, p. 1816, 2022.
[23] A. Andrianov and T. Fuerst, "Immunopotentiating and delivery systems for hcv vaccines," Viruses, vol. 13, no. 6, p. 981, 2021.
[24] M. Kayesh, M. Kohara, and K. Tsukiyama-Kohara, "Tlr agonists as vaccine adjuvants in the prevention of viral infections: an overview," Frontiers in Microbiology, vol. 14, p. 1249718, 2023.
[25] T. Weber and H. Kim, "The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production," Synthetic and Systems Biotechnology, vol. 1, no. 2, pp. 69–79, 2016.
[26] G. Nagpal, K. Chaudhary, P. Agrawal, and G. Raghava, "Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants," Journal of Translational Medicine, vol. 16, no. 1, p. 181, 2018.
[27] Q. ul Ain, M. Batool, and S. Choi, "Tlr4-targeting therapeutics: Structural basis and computer-aided drug discovery approaches," Molecules, vol. 25, no. 3, p. 627, 2020.
[28] A. Romerio and F. Peri, "Increasing the chemical variety of small-molecule-based tlr4 modulators: An overview," Frontiers in Immunology, vol. 11, p. 1210, 2020.
[29] Sunita, A. Sajid, Y. Singh, and P. Shukla, "Computational tools for modern vaccine development," Human Vaccines and Immunotherapeutics, vol. 16, no. 3, pp. 723–735, 2019.
[30] K. Singleton, A. Joffe, and W. Leitner, "Current trends, challenges, and success stories in adjuvant research," Frontiers in Immunology, vol. 14, p. 1105655, 2023.
[31] R. Rappuoli, G. Alter, and B. Pulendran, "Transforming vaccinology," Cell, vol. 187, no. 19, pp. 5171–5194, 2024.
[32] B. Pulendran, P. Arunachalam, and D. O'Hagan, "Emerging concepts in the science of vaccine adjuvants," Nature Reviews Drug Discovery, vol. 20, no. 6, pp. 454–475, 2021.
[33] J. Kim, M. Rosenberger, N. Rutledge, and A. Esser-Kahn, "Next-generation adjuvants: Applying engineering methods to create and evaluate novel immunological responses," Pharmaceutics, vol. 15, no. 6, p. 1687, 2023.
[34] CDC, "Cdc vaccine price list: Dengue (dengvaxia)," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html
[35] UNICEF Supply Division, "Ebola vaccine ervebo pricing data," 2024, [cited 2025 Aug 27]. [Online]. Available: https://www.unicef.org/supply
[36] CDC, "Cdc vaccine price list: Covid-19 (comirnaty, spikevax)," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html
[37] ——, "Cdc vaccine price list: Influenza vaccines," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html
[38] ——, "Cdc vaccine price list: Rsv vaccines (arexvy, abrysvo, mresvia)," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html
[39] NIAID, "Cytomegalovirus (cmv) vaccine research — no licensed vaccine yet," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.niaid.nih.gov/diseases-conditions/cmv
[40] GoodRx, "Ixiaro (japanese encephalitis vaccine) prices and coupons," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.goodrx.com/ixiaro
[41] CDC, "West nile virus — no licensed human vaccine," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.cdc.gov/westnile/prevention/index.html
[42] WHO, "Hepatitis c — no licensed vaccine," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c
[43] HIV.gov, "No licensed hiv vaccine," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.hiv.gov/hiv-basics/overview/history/hiv-vaccine-research
[44] NIAID, "Herpes simplex virus (hsv) vaccine research — no licensed vaccine," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.niaid.nih.gov/diseases-conditions/stds-herpes
[45] CDC, "Zika virus — no licensed vaccine," 2025, [cited 2025 Aug 27]. [Online]. Available: https://www.cdc.gov/zika/prevention