

Artificial Intelligence Based COVID-19 Vaccine Design: A Guidebook



Dr. Amit Ray, PhD

Compassionate AI Lab, Fight against COVID-19

Abstract:

This guidebook explains the step-by-step process to develop COVID-19 vaccines using modern artificial intelligence tools and techniques. Here, we explore a very different account of artificial intelligence (AI) based models for [COVID-19 vaccine design frameworks](#). This article explores the workflow of artificial intelligence and systems biology approaches in the vaccine development pipeline. We discussed DeepVaxAI, an Artificial Intelligence based multi-epitope vaccine development system, one of the key project at [compassionate AI Lab to fight against COVID-19 virus](#). The areas include a detailed description of epitope prediction tools for designing epitope-based vaccines and agent-based models for immune system response prediction, along with a focus on their potentiality safety issues.

Researchers across the world proposed several vaccines for COVID-19 utilizing AI based deep learning approaches, however, the safety and ethics of using vaccine is a key issue for combating COVID-19. Any in-silico vaccine design such as the kind discussed here is only the beginning of a process that can take years to go through in vivo testing and final implementation.

Keywords: Artificial intelligence, COVID-19, vaccine design, S protein, vaccine, machine learning, drug discovery.

How to cite this paper: Ray, Amit. (2020). *Artificial Intelligence Based COVID-19 Vaccine Design: A Guidebook*. www.amitray.com, 2020 Aug 5 Available at: <https://amitray.com/artificial-intelligence-based-vaccine-design-covid-19/>, *Compassionate AI Research, COVID Projects*.

Introduction

Effective and early vaccine design is the key challenge in the present COVID crisis. We reviewed, DeepVaxAI, an Artificial Intelligence based multi-epitope vaccine development system, one of our key project at compassionate AI Lab to fight against COVID virus [1]. The objective is to develop most reliable peptide based in-silico vaccines for any given protein sequences with minimum turn-around time. The key objective of our [Compassionate AI lab](#) is to eliminate pain of the humanity. This project is part of that endeavor. Ethics and safety are the two key issues of COVID-19 vaccine design [2, 6].

Presently humanity is going through tremendous loss and suffering due to coronavirus pandemic. Humanity is desperately looking for an urgent solution to come out of this urgent situation. Developing safe and effective vaccine is one of the key solutions for the present crisis. However, long turn-around time for vaccine development is the key obstacle for effective fight against COVID-19 pandemic. In this article, we reviewed DeepVaxAI; the Artificial Intelligence based multi-epitope [vaccine development](#) workflow process automation system. The system architecture of the AI based automatic peptide based vaccine design is shown in the figure below.

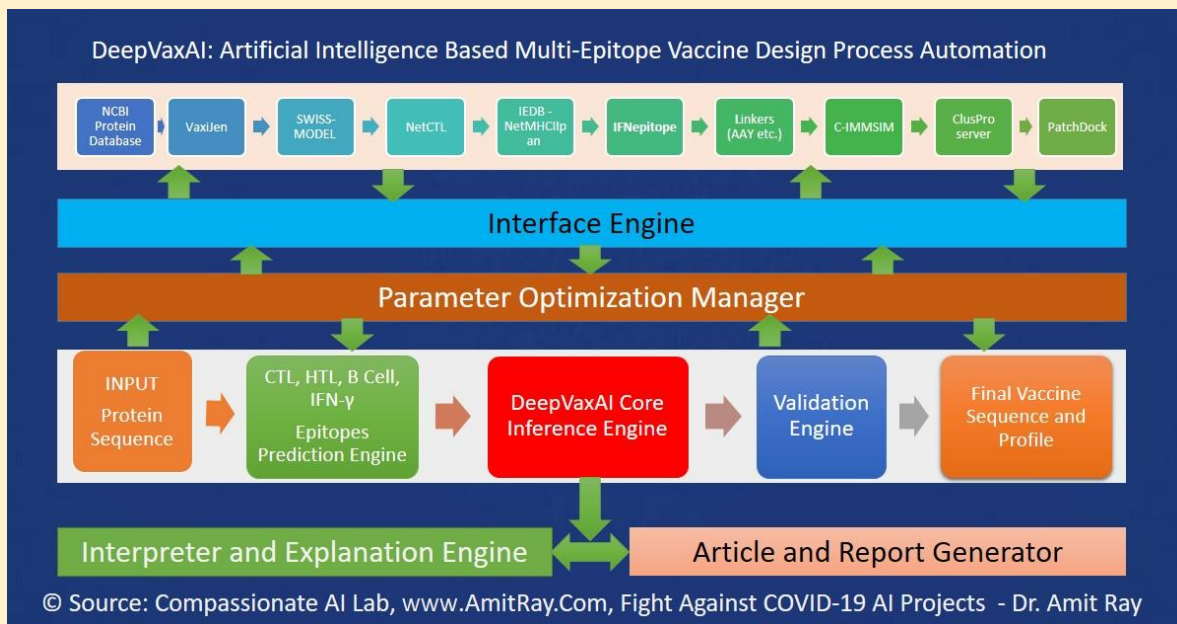


Figure 1: AI Based Vaccine Design System Architecture

The key objective of vaccination is to stimulate the immune system, the natural disease-fighting capabilities of the body [11]. To develop peptic based vaccine epitopes the following properties

are usually preferred: highly antigenic, highly non-allergic, highly non-toxic, significant population coverage, having a strong binding affinity with common human allele.

System Architecture of the Artificial Intelligence Based Vaccine Design System

The interface engine, part of the DeepVaxAI system provides extensive links to major internal and external databases. The interface engine reduces the workload of the research scientists. The key servers linked to the interface engine are NCBI database, IEDB server, NetCTL, VaxiJen, Toxinpred server, ERRAT server, GRAMM-X Simulation web server, LIGPLUS server, AllerTOP, ProtParam, C-ImmSim server, SwissDock, PatchDock, HADDOCK, and YASARA server.

Traditional methods of peptide based vaccine development process is time-consuming, monotonous and very labor-intensive. [Deep Artificial intelligence \(Deep AI\)](#) is a key technology for optimizing the process flow in many application areas [11]. Here, we especially focus for the vaccine development. It can improve process accuracy and remove potential human biases, errors, and repetitions. It will substantially reduce the vaccine development turn-around times. It will enable subject matter experts to focus more on higher value tasks like wet lab experiments. The gap between wet lab experiments and in-silico studies is the key obstacle for vaccine development. With the help of DeepVaxAI, research scientists can now focus more on wet lab experiments and eliminate the gap.

The core inference engine part of the DeepVaxAI system provides facilities for modeling with various AI algorithms. Traditionally, shallow AI includes narrow areas and build models with [deep learning algorithms. Deep learning algorithms like MLP, DNN, CNN, RNN, LSTM are powerful. However, they have many limitations.](#) On the other hand Deep AI includes integration and collaboration of many technologies to provide the highest level of machine intelligence, process automation, automatic interpretation, explanation, report generation, and scientific article generation capabilities. Research scientists will get more time for wet lab experiments, where the true solution exists.

15 Key Steps for Multi Epitope Vaccine Design

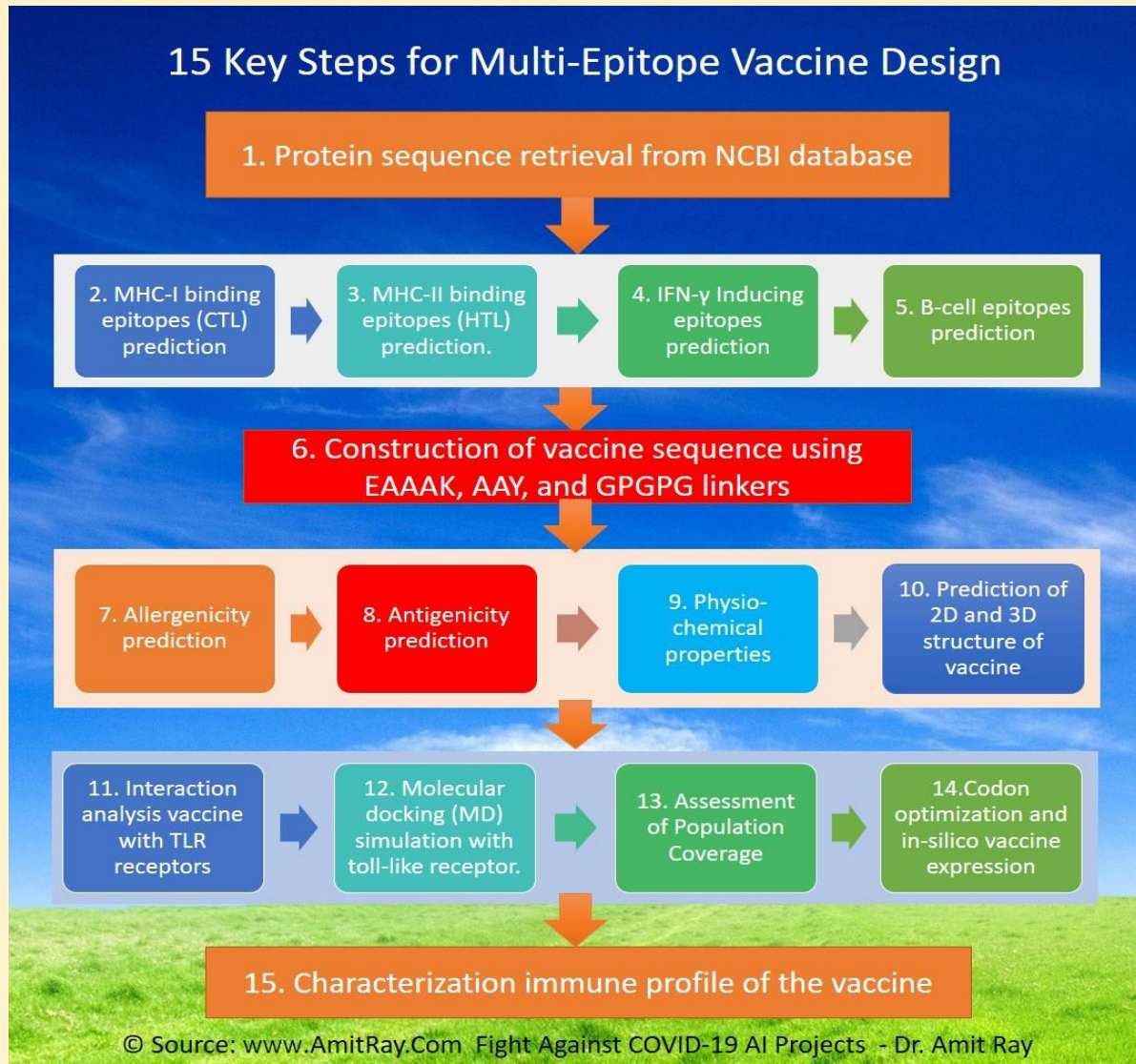


Figure 2: 15 Key Steps for Multi Epitope Vaccine Design

The 15 primary steps for multi-epitope vaccine design are as follows:

1. Retrieval of protein sequence from NCBI database
2. MHC-I binding epitopes (CTL) prediction
3. MHC-II binding epitopes (HTL) prediction.
4. Prediction of IFN- γ Inducing Epitopes
5. B-cell epitopes prediction
6. Construction of vaccine sequence
7. Calculating allergenicity of vaccine sequence
8. Calculating antigenicity of vaccine sequence
9. Reviewing physio-chemical properties of vaccine sequence
10. Prediction of secondary and tertiary structure of vaccine
11. Interaction analysis vaccine with TLR receptors
12. Molecular docking (MD) simulation with toll-like receptor.
13. Assessment of Population Coverage
14. Codon optimization and in-silico vaccine expression
15. Characterization of immune profile of the vaccine.

The 26 Top Vaccine Design Tools and Servers

The 26 Top Vaccine Design Tools and Servers			
Sl. No	Purpose	Server Name	Website Link
1	Protein sequences selection	NCBI database	https://www.ncbi.nlm.nih.gov/
2	Homology check	pBLAST server	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
3	HTL epitopes prediction	IEDB server	https://www.iedb.org
4	CTL epitopes prediction	NetCTL	http://www.cbs.dtu.dk/services/NetCTL/
5	B-Cell epitopes prediction	ABCpred server	http://crdd.osdd.net/raghava/abcpred/
6	B-Cell epitopes prediction	BepiPred server	http://tools.iedb.org/bcell/result/
7	IFN- γ epitopes prediction	IFN- γ epitope server	http://crdd.osdd.net/raghava/ifnepitope/scan.php
8	Physicochemical properties	ProtParam	https://web.expasy.org/protparam/
9	Antigenicity prediction	VaxiJen v2.0	http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
10	Antigenicity prediction	ANTIGENpro server	http://scratch.proteomics.ics.uci.edu
11	Allergenicity prediction	Algpred server	http://crdd.osdd.net/raghava/
12	Allergenicity prediction	AllerTop server	https://www.ddg-pharmfac.net/AllerTOP/
13	Toxicity prediction	ToxinPred tool	http://crdd.osdd.net/raghava/toxinpred/
14	Protein structure assessment	Phyre 2 server	http://www.sbg.bio.ic.ac.uk/phyre2/
15	Ramachandran plot, Protein structure	SWISS-MODEL	https://swissmodel.expasy.org/assess
16	To build tertiary structure	RaptorX server	http://raptorx.uchicago.edu/StructPredV2/predict/
17	Protein structure refinement	GalaxyRefine server	http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE
18	3D protein structure refinement	3Drefine server	http://sysbio.rnet.missouri.edu/3Drefine/
19	3D structure	QMEAN	https://swissmodel.expasy.org/qmean/
20	Ramachandran plot	RAMPAGE server	http://mordred.bioc.cam.ac.uk/~rapper/rampage.php
21	Docking analysis	ClusPro server	https://cluspro.bu.edu/login.php?redir/queue.php
22	Docking analysis	PatchDock server	https://bioinfo3d.cs.tau.ac.il/PatchDock/
23	TLR-3 and vaccine interaction	HADDOCK server	http://milou.science.uu.nl/services/HADDOCK2.2/haddockserver-easy.html
24	Immune dynamics Study	C-ImmSim server	https://www.iac.cnr.it/~filippo/projects/cimmsim-online.html
25	Protein structure validation	ProSA-web	https://prosa.services.came.sbg.ac.at/prosa.php
26	Codon optimization	Java Codon	http://www.jcat.de/ was used to optimize codon

© Source: Compassionate AI Lab, www.AmitRay.Com, Fight Against COVID-19, AI Projects - Dr. Amit Ray

Figure 3: 26 Top COVID Vaccine Design Tools and Servers

Genomic Structure of SARS-CoV-2

The genome of SARS-CoV-2 is a single-stranded positive-sense RNA with the size of 29.8–30 kb encoding about 9860 amino acids. The SARS-CoV-2 protein sequence includes 16 non-structural proteins (nsp1, nsp2, nsp3, ..., nsp16), 4 structural proteins, (E, M, N and S) proteins, and accessory proteins (ORF3a, ORF7a, and ORF8).

The S, N, M, E form the structural proteins that play a vital role in the life cycle of the viral particles. The S protein is shaped like a clove with two subunits S1 and S2 which promotes receptor binding and membrane fusion respectively [7, 8]. The N protein enhances viral entry and performs post-fusion cellular processes necessary for viral survival and growth in the host. The E protein

promotes virion formation and viral pathogenicity while M protein forms ribonucleoproteins and mediates inflammatory responses in hosts. Proteins ORF1a and ORF1ab are papain-like proteases (PL(pro)) involved in viral infection and are potential targets for the development of antiviral drugs.

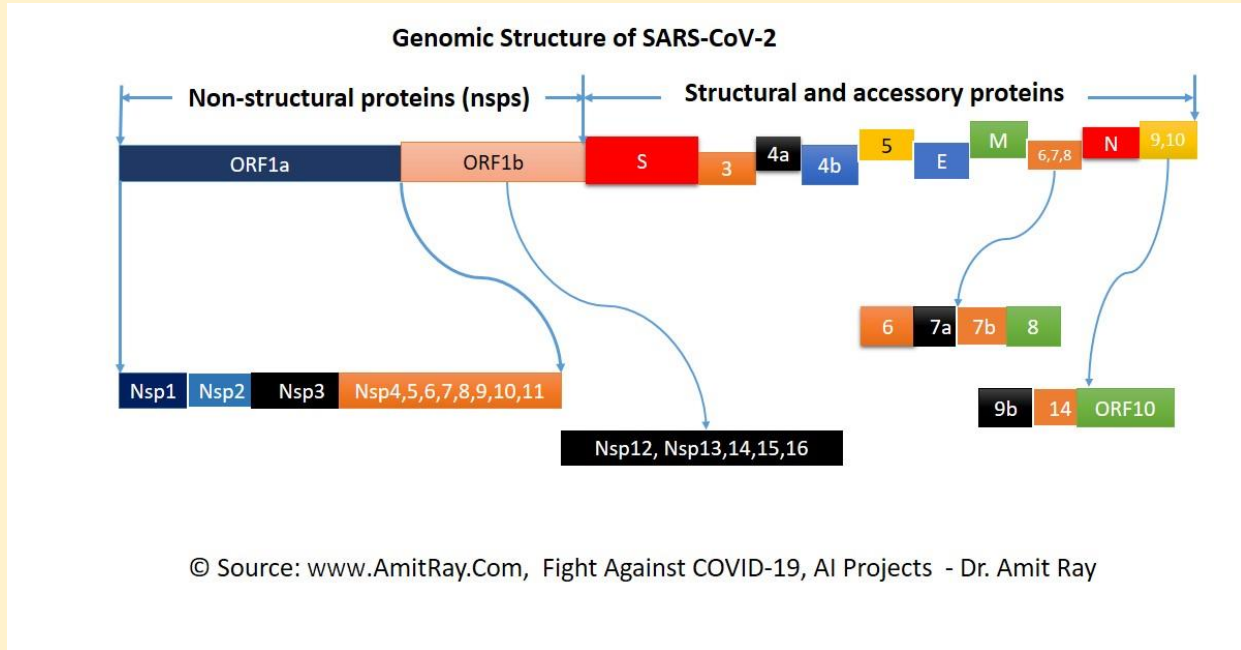


Figure 4: Genomic Structure of SARS-CoV-2

Unusually among coronaviruses, the SARS-CoV-2 S protein is proteolytically cleaved into an S1 subunit (685 amino acids) and an S2 membrane-spanning subunit (588 amino acids), the latter being highly conserved (99%) among CoV families. By contrast, S1 shows only 70% identity to other human CoV strains and the differences are concentrated in the RBD, which facilitates virus entry by binding to angiotensin-converting enzyme 2 (ACE2) on the host cell surface.

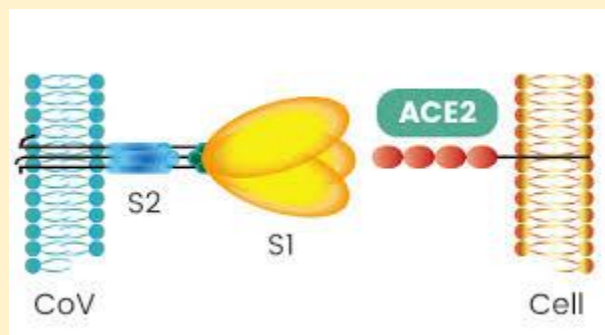


Figure-5: S-Protein and ACE2 Binding

The Candidate Vaccine against SARS-CoV-2

One of our objective is to reduce the turn-around time of the vaccine development process. Hence, we divided the entire study into several phases. First, we have created 15 batches of protein sequences of COVID viruses, randomly selected from the NCBI database.

Our approach is to train the DeepVaxAI system by observing the human behavior of the workflow and parameter optimization, to better automate the end-to-end vaccine design processes. The main workflow includes epitope predictions (HTL, CTL, IFN- γ and B cell epitopes) from the chosen protein sequences; vaccine construction and its quality check. Molecular Docking with immune cell receptor, followed by molecular dynamics simulation (MDS) to check vaccine's stability. Lastly, codon adaptation and immune simulation to understand how the vaccine acquires an immune response.

Vaccine Linkers and Adjuvants

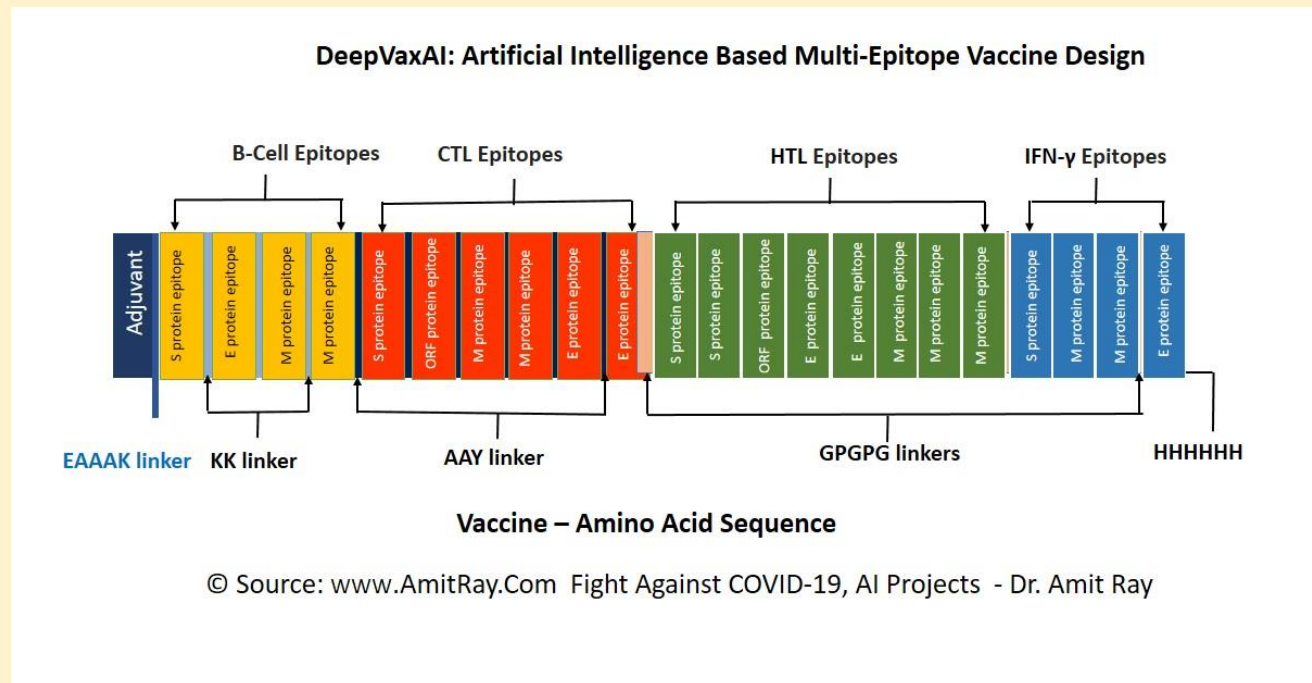


Figure 6: SARS-CoV-2 (COVID) Vaccine Amino Acid Sequence

We analyzed various competitive candidate vaccines to fight against COVID viruses. Finally, we selected the vaccine construct consisted of 563 amino acid residues derived from different peptide

sequences. The immunogenic epitopes were united with the help of linkers; B-cell (KK linkers), CTL (AAY linkers), HTL (GPGPG linkers), and IFN- γ (GPGPG linkers). To enhance vaccine immunogenicity adjuvant was added to the N-terminal of the vaccine with the aid of the EAAAK linker. We analysed human β -defensin-2 (hBD-2), human β -defensin-3 (hBD-3) and Matrix-M1 as adjuvants to enhance the immunogenic response.

Conclusion

We disused the system architecture, tools and techniques of the Deep-AI based vaccine design system. The system constructed a 563 amino acid based vaccine for COVID viruses. But there are still many obstacles to overcome. Manual interventions and checks are required in many places. Special care must be taken for final vaccine selection. For example, spike protein-based SARS vaccine often induce harmful immune responses that cause liver damages. We want the system to take care every possibility to provide the highest level of human safety.

Presently deep learning algorithms has several limitations [17]. However, application of DeepVaxAI in silico methods can be used to design an effective vaccine in lesser time and low cost. Research scientists, process analysts, technical experts, and knowledge workers can drive unprecedented scale of automation while improving performance, accuracy, and data security. Deep-AI based vaccine design system is a powerful end-to-end automaton process harnessing the power of multiple technologies to improve accuracy, effectiveness, and to reduce time and cost for effective vaccine development.

References:

1. Ray, Amit. (2020). Artificial Intelligence Based Multi-Epitope Vaccine Design to Fight against COVID-19. Available at: <https://amitray.com/artificial-intelligence-based-vaccine-design-covid-19/>
2. Ray, Amit. (2020). Gene-Based (RNA-DNA) COVID-19 Vaccines Ethical issues. 10.13140/RG.2.2.17923.73761.
3. Jefferson T. (1998). Vaccination and its adverse effects: real or perceived. Society should think about means of linking exposure to potential long term effect. BMJ (Clinical research ed.), 317(7152), 159–160. doi.org/10.1136/bmj.317.7152.159

4. Andrew F. Read, et al., Imperfect Vaccination Can Enhance the Transmission of Highly Virulent Pathogens, *POLS Biology*.
5. Rauch Susanne, New Vaccine Technologies to Combat Outbreak Situations, *Frontiers in Immunology*.
6. Ray, Amit. (2020). Gene-Based COVID 19 Vaccines Ethical Issues. Retrieved from <https://amitray.com/gene-based-covid-19-vaccines-ethical-issues/>
7. Ray, Amit. (2020). A Comparative Study of Eight Classes of COVID 19 Vaccines. Retrieved from <https://amitray.com/a-comparative-study-of-eight-classes-of-covid-19-vaccines/>
8. Ray, Amit. (2020). Life Cycle Analysis of Coronavirus (COVID-19) – The SARS-CoV-2 Virus. Retrieved from <https://amitray.com/life-cycle-of-coronavirus-covid-19-the-sars-cov-2-virus/>
9. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. (2020) 395:565–74. doi: 10.1016/S0140-6736(20)30251-8.
10. Ray, Amit. (2020). Remdesivir, Favipiravir, hydroxychloroquine, Dexamethasone, Coronil and Swasari A Critical Review. Retrieved from <https://amitray.com/remdesivir-favipiravir-hydroxychloroquine-dexamethasone-coronil/>
11. Ray, Amit. (2020). Fight Against Coronavirus COVID-19: Five Key Ayurveda Techniques. Retrieved from <https://amitray.com/fight-against-coronavirus-covid-19-five-key-ayurveda-techniques/>
12. van Doremalen N, Haddock E, Feldmann F, Meade-White K, Bushmaker T, Fischer R, et al. A single dose of ChAdOx1 MERS provides broad protective immunity against a variety of MERS-CoV strains. *bioRxiv* [Preprint]. (2020) doi: 10.1101/2020.04.13.036293v1.
13. Ray, Amit. (2020). Top Research Updates to Fight against Coronavirus COVID. 10.13140/RG.2.2.33461.09442.
14. John Barthelow Classen, Public should be told that vaccines may have long term adverse effects, *BMJ*

15. Angela S Clem, Fundamentals of Vaccine Immunology, Journal of Global Infectious Diseases.
16. Hudson, Shally (2020), Artificial intelligence, Cognitive Robotics and Human Psychology. DO-10.13140/RG. 2.2. 20153.52323.
17. Ray, Amit. (2020), [Artificial Intelligence to Fight Against COVID-19: Projects](https://amitray.com/artificial-intelligence-to-fight-against-covid-19/)
<https://amitray.com/artificial-intelligence-to-fight-against-covid-19/>
18. Ray, Amit. (2018). 7 Limitations of Deep Learning Algorithms of AI." Retrieved from <https://amitray.com/7-limitations-of-deep-learning-algorithms-of-ai/>