Computational Immunoinformatics and Generative Modelling: AI-Driven Acceleration of Vaccine Antigen Design and Development

Dr. Agata Grabowska, Associate Professor of Computer Science, Wrocław University of Science and Technology, Poland

1. Introduction

Vaccines are the most important tool for maintaining public health. As vaccines significantly reduce morbidity, hospitalization rates, spending on medical help, and death rates, they are crucial for safeguarding public health. In the present landscape, each year, more than 100 vaccines go through several development phases and hit the market to extend human life. While the current vaccine development process is time-consuming, new and reinfecting diseases indicate that the urgency for vaccines is increasing. Novel vaccine developments have the potential to eradicate diseases from the world. At the same time, existing vaccine development processes can be criticized because of their prolonged time frame.

Because of its capacity to manage, process, analyze, and make conclusions more precisely and efficiently than human intelligence, artificial intelligence (AI) has the ability to transform research initiatives and medical services. The adoption of AI in the field of healthcare and life sciences poses a significant challenge for many medical industries, including healthcare organizations, pharmaceutical, and biotechnology companies. Currently, AI solutions and vaccine designs are in their infancy. A vaccine development timeline has the potential to be shortened to just a few days by AI-based solutions. In emergency situations such as pandemic outbreaks, epidemic diseases, or bioterrorism, disease prevention is an important but legitimate demand. AI systems can stimulate knowledge creation, broad and deep exploration beyond human cognitive capacity, as well as generate and evaluate beneficial vaccines for vaccination.

1.1. Background and Significance

The development of a vaccine is the most effective way to address an infectious disease outbreak. Vaccines prevent deadly and debilitating infections and save millions of lives worldwide. The century of the vaccine has witnessed the successful development of vaccines against many pathogens, including meningococcal, rotavirus, and human papillomavirus. While immense progress has been made, there is still extensive room for improvement. The most recent severe pandemic was caused by a coronavirus agent. Many developed countries have worked tirelessly to develop vaccines to fight the disease, and several of them have made it to market within approximately a year.

However, there are still many public health issues that must be resolved in order to bring the outbreak under control. The current approach to vaccine development takes a long time, requires a significant financial investment, and frequently results in a low rate of success. It follows the traditional path of stepwise target identification, vaccine design strategies, preclinical testing, and multiple phase clinical trials to demonstrate safety and efficacy before mass vaccination. Leveraging big data, machine learning, and AI technologies can greatly optimize the vaccine development process. These techniques have become increasingly popular due to their ability to rapidly construct accurate predictive models. Integrating AI methods into vaccine design and validation would not only speed up the process but also increase the likelihood of developing vaccines that will ultimately be accepted for human use. These approaches could reduce the risk of vaccine study delays and ineffective investments, as well as stimulate the investment of funds and time in promising work.

1.2. Scope and Objectives

Scope and Objectives

Vaccines have fundamentally changed the fabric of modern societies, reducing the impact of infectious diseases that have ravaged human populations in recorded history. New pathways for questing and formulating vaccine candidates in laboratories are being enabled by recent advances in biotechnologies and huge volumes of relevant data, including clinical trial results, binding properties of antigens, immune cell-MHC interactions, and more. There is also a convergence of data and tools from traditional vaccine development wet labs with medical imaging data on health outcomes, gene interaction networks, and cellular phenotypic data. A critical foundation of vaccines,

though, still remains the mimicry control of human immunological memory, established in preclinical testing regimens in animal models. In recent times, vaccine researchers are integrating modalities targeted therapies, including T cells and recombinant proteins.

Several techniques have been reported to have utilized AI in data analytics and prediction tasks in the process of vaccine development. Two systematic reviews have recently cataloged the vaccines developed using methodologies and identified gaps and opportunities to accelerate vaccine development. In this work, we thoroughly identify overlaps with the key themes identified in a systematic review of ML in vaccine development and build on their advancements using both an empirical and theoretical framework. The future promises of AI in the accelerating part of vaccine science are discussed in terms of clinical and biorepository trials and the creation of broader databases representing vaccine-mediated human disease processes with a view to modeling pathophysiological pathways and applying knowledge metrics. Such a vision is presented as it may best depict the potential for AI in permutation with wet lab techniques in vaccine signature-enhanced personalized therapy. In short, the discussion progresses from AI as a computational adjunct to wet labs to AI-aided in silico full surrogate vaccine development theory. As the objective of this paper is to discuss implications and evidence base with the view to provoke scientific debate, there is a structure to this paper, and a roadmap is presented in terms of various vaccine science components where AI has been integrated.

2. Foundations of Vaccine Development

Vaccines mimic the presence of the virus or bacteria they protect against, so that the body can learn to fight the infection. They generate an immune response and thereby enable long-term immunity, which can reduce the severity of the illness. There are various types of vaccines, such as live attenuated vaccines, inactivated vaccines, killed vaccines, subunit vaccines, recombinant vaccines, polysaccharide vaccines, conjugate vaccines, DNA and RNA vaccines, and viral vector vaccines. Not all vaccines are perfect, and there are challenges in the development of these vaccines. For example, pathogens have diverse molecular characteristics, vaccines may inadequately stimulate the immune responses, and there is a lack of tools to forecast immunogenic properties. Thus, artificial intelligence provides numerous opportunities to address the modern problems of vaccine development.

Vaccine development is a complex process involving research and development of the vaccine, hosting preclinical data, and submitting an investigational new drug application for Phases I, II, and III clinical trials. This process requires a significant amount of time, resources, and money. Moreover, hundreds of potential vaccine targets need to be screened, which is computationally intensive. For several decades, vaccines have been developed using the best immunological and virological techniques. However, traditional approaches for the development of vaccines are constrained by enormous costs and time requirements, while global health problems like the epidemic call for faster and more creative solutions. Public-private partnerships between the government, biotech, and academia, among others, are designing ways to reform and expedite this lengthy and costly process, including leveraging AI and data analytics. In all cases, big data have already been vital to accelerating the vaccine production route, whether via more effective target selection, the transition to readouts, or the efficient definition of settings for the conduct of clinical trials.

2.1. Basic Concepts and Processes

Each antibody-producing B cell presents one type of antigen. These surface-expressed antigens, which cells or microorganisms possess, are specifically scanned by macrophages, dendritic cells, and others. Even if the virus has not infected many cells in the human body, if B cells of many kinds are produced by inoculating the antigens themselves, they circulate in the body. Triggered antigens have undergone differentiation into plasma cells that secrete large amounts of the corresponding antibody. To adapt the immune system, it should be prepared in advance to prevent the disease from developing before the virus infects a large number of cells. Before humans are harmed, their antigens are presented, their blood tools are present, and they are tested. The clinical significance of antibody production due to vaccination is very important.

The production of a vaccine includes basic research, initial discovery of antigens, preclinical research, and clinical trials. The vaccine, alone or in combination with other drugs, is initially tested in vitro or in vivo. Finally, it is tested in volunteers in clinics. In each step of discovery, testing will only be carried out if there is scientific evidence certifying safety, quality, efficacy, and so on. Principle 4 also follows the principle of protective patent production. In order to rapidly produce a vaccine to combat an

emerging epidemic, we will discuss the principles of vaccine production. The application of antigenic peptides, receptors of mRNA vaccines, CD16 of testing, gene editing, and others is to determine the antibodies. A vaccine is a drug that is safe in almost all cases for application to the general population, so a lot of care is taken in each development test. Every year, each national management body performs the same tests, and when proven in clinical tests, a vaccine goes on the market. This major investment will develop a drug.

Concurrently, over the years, various vaccine technologies have been developed, such as inactivated pathogens, live and dead vaccines, adjuvant vaccines, subunit vaccines, mRNA vaccines, DNA vaccines, and vector vaccines. These vaccines are made using a variety of manufacturing technologies. Among them, the mRNA vaccine attracted attention during the pandemic. Most companies rely mainly on large trials to show the protective effect in an emergency. Administer the vaccine to one person and conduct a simple examination to mark it as the antibody.

2.2. Challenges in Traditional Approaches

Traditionally, developing a vaccine from target identification to clinical approval requires a considerable period of following preclinical testing and clinical trials and can take 10 to 15 years. The process is expensive, and work, time, and resources are allocated inefficiently. Identifying effective antigens requires testing thousands of candidates, followed by establishing purity, identifying conformation, and optimizing synthesis. In most cases, advisability cannot be determined for advanced efficacy testing. This is posed by a lack of suitably precise and integrative preclinical systems. The determination is hard, expensive, and time-consuming. Approaches from animals may not predict human immune responses and do not promise an appropriate comparison of the trial. Additionally, desirable immune responses may be rare or can decline naturally over time, which may enforce extended Phase 3 trials with increased costs. If developed, a Phase 3 vaccine is not guaranteed to be effective due to stochastic or cost-based decisions in earlier stages with an imperfect fit between testing models, hampered diversity, impurity, limited replication opportunities, capacity constraints, and a lack of adaptability. Incessant benefits and determination in Phase 2 to diminish feasibility, as recommended by biomarker response, may be impossible to obtain if a vaccine requires revaccination once, twice, or more.

Challenges translate preclinical success into human clinical approval through the understanding of the immune response pathway. Traditional research practices and statistical evaluations in a well-defined clinical model are not suitable for the identification of innate biomarkers and the discovery of human immune pathways. Under traditional approaches, the innate immune system pathway cannot be predicted in the absence of a standardized signal demonstration in the first-in-human study of vaccines. Some proposals are shared as adaptive therapeutics; however, they have not been tested for vaccines in human subjects. In summary, quality development, a significant need for drugs, and strategic food and drug administration are vaccinological problems with some special strengths and weaknesses. There is an urgent need to provide vaccines at an adequate pace and positive information for licensure.

3. Machine Learning in Vaccine Development

In recent years, interest in utilizing data-driven approaches, including machine learning, to streamline the process of vaccine development has grown. Vaccine development poses a unique set of challenges, but machine learning techniques, including classification, regression, and clustering, can play a key role in overcoming some of them. These techniques are employed to build predictive models, analyze complex datasets, and gain insights from the data. This can help researchers make data-driven decisions and yield more accurate, efficient, and reliable results than traditional methods.

Machine learning models use predetermined qualities of proteins to predict their function. This information can be used to design vaccine antigens. They also use sequence or structural information to predict the immune response that will be elicited by these antigens, the conditions required to produce the biggest and most long-lasting response, and which hosts a vaccine will work in. Machine learning can mine clinical trial datasets to identify key timepoints and correlates of protection to measure in studies. Antigen design, by identifying the best parts of an infectious agent to help the immune system identify it later, is frequently undertaken by experimental studies and by computational methods. Conducting a large number of vaccine studies to create a predictive mathematical model could accelerate vaccine development and provide valuable insight into the B cell selection process.

3.1. Overview of Machine Learning

Machine learning is an area of artificial intelligence that can lead to the development of algorithms, computational models, and statistical models that can be effectively used by computers to act or react in response to various situations given to them. ML makes use of data, various methods, and produces automatic decisions and predictions without the active participation of humans. The widely used methods in machine learning include unsupervised and supervised learning or the convergence of both as a hybrid approach. Supervised learning is where the ML algorithm is trained on entering the data and its expected output. Neural networks are the evolving and most powerful outcome of supervised learning. Some applications of supervised learning are animal sound detectors in the wild, robots playing soccer, and implanting artificial neural networks into humanoid robots. Unsupervised learning is raw and has direct potential for understanding data; one classic example is its use in data clustering.

The design and development of vaccines hinge on a critical mass of biological, chemical, immunogenic, and pathophysiological data. All processes of the development of vaccines generate a large volume of complex data. Often, it is not possible to identify relationships or trends in data through traditional data analysis. Machine learning converts the data into computational models based on complex mathematics. The machine learning model learns and makes predictions. Uniquely in the development of vaccines, through using big data, machine learning will predict the immunogenicity, reactogenicity, and efficacy of vaccines before getting exposed to humans. This will avoid the failure of vaccines due to safety and efficacy, which is happening in the trial phase of vaccine development.

3.2. Applications in Antigen Design

As the fundamental step in vaccine development, the prediction of optimal antigens or immunological target proteins can accelerate the progress of vaccine development. However, the scientific hypotheses and design strategies that we rely on are based on the successful or failed data of previous vaccines and studies. Can machines or computers also make predictions? Based on the data provided, in addition to the relationship between antigens and host protective responses, a series of cooperative calculations based on previous algorithms and software can be used to screen or design immunogenic antigens. In addition to assessing the effectiveness of antigens, some

studies use models for B cell epitope prediction and T cell epitope selection. Compared with traditional 3D structural design strategies, these operations can help us save time and cost at the screening stage.

Furthermore, the machine algorithms combine antigens that can induce T cell responses and B cell responses, so that the vaccine can elicit a more efficient long-term cellular immunity. In summary, these methods use machine learning to screen out and design high-efficiency antigens to improve the success rate of the downstream of the vaccine. For instance, novel CTL epitopes of FAP were identified through epitope prediction algorithms to facilitate the progression of personalized peptide vaccinations. A machine learning algorithm was built to determine the most unusual FAP peptides as new epitopes and demonstrated that these peptides were capable of inducing proven T-cell immunogenicity and clinical activity. In turn, isolated T cells were then expanded into T-cell products and used to personalize T-cell immunotherapy. Databases can also predict antigens or T cell epitopes based on published research.

3.3. Applications in Immune Response Prediction

The next area showcasing the potential of machine learning in predicting vaccine outcomes is in determining how individuals are likely to respond to a vaccine with greater specificity. Predicting vaccine-induced immune responses leverages advanced computational approaches that analyze complex, often multi-site data mined directly from the immune response. This generation of improved immune predictions is then used to identify correlates of protection and immune markers that can inform the rational design of new vaccines that target specific immune activities.

The promise of machine-learning immune predictions is evidenced by strong performance when predictions have been compared to actual clinical outcomes. Studies have demonstrated the ability of computational approaches to predict influenza vaccine immune responses in young and aged human populations, to predict mRNA-based COVID-19 vaccine immune responses in cancer patients, and to predict malaria vaccine immune responses in healthy volunteers and children in Kenyan and Malian populations. The prediction accuracies, in all tests, often correlate with disease risk and outcomes. Predictive algorithms, therefore, have the potential to be used to create preventive interventions and therapeutics in response to either emerging diseases or customized vaccines for specific low-immune responder populations that are at high

risk of disease acquisition. Facilitating the rational selection of vaccine candidates with higher efficacy based on immune responses will also reduce the risk of investing in vaccines that trial as 'reactogenic' in the target population or display unexpected immune enhancement leading to failed trial outcomes late in the development process. These advances in predictive immunology critically hinge on the greater ability for AI to predict which immune responses are protective as a function of pathogen, tissue microenvironments, and host variability.

3.4. Applications in Clinical Trial Efficiency

A key bottleneck in bringing vaccines to market is the time and energy of clinical trials. Machine learning (ML) can improve the process at all stages of development. Beyond identifying basic inclusion/exclusion criteria for trial participants, an ML model can suggest which patients are likely to succeed based on past performance. Drug makers increase their likelihood of success the broader the participant pool. For individuals wanting to participate in a trial, ML models can be insightful. It is more likely to include a participant who will succeed based on the characteristics of the model.

The design of clinical trials is another significant area that may be enhanced by AI. During extensive, high-volume data analyses, AI can notice patterns that might suggest a need to alter the design of a study in real time. In order to create more flexible designs that may generate advantageous interim findings, many algorithmic designs are being developed. The FDA has expanded the use of such strategies in recent years, causing approvals that some fear are derived from less restrictive analyses. Also, the FDA recognizes that such techniques may be especially useful in emergency situations, such as the drug development response to the pandemic. Because of the significantly increased statistical power achieved by many such AI-backed adaptive trials, several biotechs and regulators have begun placing more trust in their interim findings. These are attractive to all stakeholders because they are able to offer the results of an ongoing study more quickly, often resulting in accelerated drug approvals.

4. Case Studies and Examples

The application of AI to vaccine development has long been discussed. Here, we provide an overview of several case studies, which exemplify the successful integration of AI with specific vaccine-related immunological memory data, techniques, and tools. Applied efforts like these are critical in establishing the real-world applicability of using

AI within this context. Robust examples, such as those described in this section, provide evidence that innovation in vaccine strategies can be achieved by integrating AI and, critically, with demonstrable results. Discussions on incentives and changes impacting the broader application of AI-based technologies in biology and precision medicine are further detailed.

4.1 Case study 1: Prediction of best formulations for a Leishmania vaccine

The development of a human vaccine against parasitological disease leishmaniasis has been a target for many decades, but as yet no effective vaccine is available. Recently, there has been a drive to reverse this, and significant consideration has been given to the late-stage clinical trial of a veterinary canine vaccine for the same disease. AI has been used to resolve the impasse of identifying the best Leishmania canis antigens to include within a prospective protective human vaccine. Exploratory results have since been validated experimentally in a preclinical assay with success and are currently informing downstream antigens to be taken to clinical trial.

4.1. Successful Implementation of AI in Vaccine Development

In this subsection, we look at cases where AI was successfully applied to assist in vaccine development. The first examples illustrate instances in which important aspects of AI were used successfully, like vaccine design and animal study modeling, but the final animal models created using AI were not used in the development of the final vaccines. The second group of cases concerns studies in which substantial AI-based work was instrumental in human studies that led to successful vaccine development. In almost every case, the successful AI system is commercially available for use in research. Further, a vaccine was developed using new lipid nanoparticles and AI help from its development. It took researchers just 42 days to launch human trials, a process that usually spans six months. They identified a sequence for an mRNA vaccine with the potential for clinical success.

Lots of lessons you could learn here – mainly, better and continual validation and upgrading of AI systems. The program is considered the most valuable and successful AI-assisted vaccine development effort to date. AI was used to build animal models for various diseases; they added humans to the models and were able to predict that a specific vaccine increased disease in humans to a level that matched animal data-based

predictions. While the vaccine was never officially licensed, it was used through a compassionate use research protocol, and many individuals were vaccinated. Because of the computation they used, they utilized funding to reformulate AI tools and modeled the outbreak in multiple countries; those with the vaccine and those without were evaluated. Vaccines were randomized at the end of May 2019, a week after the first results came out. The vaccine was the only one that showed efficacy, and a formal protocol from May 2019 used it only. Another vaccine was also ready to enter the trial at that time but was scrapped. AI was used to predict vaccine efficacy. It was used to analyze real-time sequencing data to ensure the vaccine was the right strain, including enhancement results. The quality system they designed also helped.

5. Future Directions and Challenges

Emerging trends that are likely to shape the future of AI in vaccine research include the ongoing advancements in unsupervised representation learning, deep learning, reinforcement learning, graph neural networks, advanced data integration techniques, and different federated learning, transfer learning, and domain adaptation methodologies. These research directions will allow for the development of more efficient and interpretable methods for integrating heterogeneous data sources. Especially, the continuous collection, curation, and mining of data—a process often referred to as big data analytics—will be crucial in the design of accurate machine learning models. There are numerous obstacles that stand in the way of a successful translation of these innovative AI strategies into the clinic. Primary considerations are the cost and the required infrastructural capability to implement such a system in low-income areas or remote populations where healthcare resources are limited. It is crucial that the successful implementation of these AI vaccines has a low threshold for infrastructural use in low- to middle-income countries.

With the rapid development of more complex data and ML approaches, it will be essential to ensure continuous training, fine-tuning, and expansion of the dataset by integrating novel information. Furthermore, disciplinary interaction between computational and experimental scientists needs to be encouraged to foster innovations in this field. Ethical aspects of data collection, especially regarding patient data privacy, as well as issues related to explainability and ethical use of ML-generated decision-support systems in healthcare and vaccine development need to be considered. Despite

these challenges, big data ML remains a hopeful tool in outlining computational protocols to design and develop new vaccines and optimize vaccine processes as well. Future advancements in big data ML should hopefully ease the development of computational methods for the easy design and production of vaccines. Ethical issues regarding data storage and dissemination in vaccine trials can be overcome using blockchain technology. Hopefully, further developments in the technology and industry's experience will assist in overcoming these challenges.

5.1. Emerging Trends in AI-Based Vaccine Development

The field of AI is fast evolving, with newer algorithms including generative models and reinforcement learning slowly pushing the envelope. Likewise, several methodologies and practices in the AI space are proving to be promising for accelerating the discovery and delivery of newer vaccines. The integration of AI with genomics and biotechnology to decipher gene expression signatures in immune cells and individual B and T cell receptor sequences for the design of novel vaccines is gaining traction. The use of AI-based deep learning models for direct prediction of significant changes in viral genomes and the computational design of antigens that will elicit broadly neutralizing antibodies pose potential avenues for facilitating the development of universal vaccine modules that account for potentially significant vaccine escape mutants. Moreover, the application of AI-based systems biology approaches to integrating multi-omic data to identify combinatorial correlates of durable immunity is a novel approach to accelerate vaccine development. Predictive analytics integrating big data and structural biology aims at the computational prediction of broadly neutralizing antibody epitopes and the creation of antigens with conserved and immunogenic viral regions. Artificial intelligence-based development of personalized vaccines using genomic and immune profiling data is currently emerging.

The steady growth in computational power and the accessibility of big data have fueled the growth of data-driven AI, making it a convenient tool for analytical workflows. Research has thus far shown greater capacity for discovering subtle similarities while designing novel vaccines through AI. In the near future, several trends in this space are anticipated, including trends around harnessing the integration of AI with genomics and biotechnology, advancements in big data-driven AI for better data pools, demonstrating robust and premature AI vaccines, and trends in AI for predicting real-time genomic

signatures. The third trend that holds great potential refers to the nuanced computational design of vaccines against rapidly shifting strains of rapidly evolving viruses, by harnessing AI tools such as through the computational prediction of broadly neutralizing antibody epitopes and the merging of antigens that allows the integration of human protein features with viral characteristics in novel conserved and immunogenic vaccine antigens.

5.2. Ethical and Regulatory Considerations

Machine learning and AI-driven algorithms can process and analyze large volumes of heterogeneous data in an unbiased manner, helping to develop new insights and accelerate experimental validation. In the context of vaccine development, these tools can be used to increase the coverage of applicable experimental data, including in silico, preclinical, and clinical data, in model building and predictions. Predictive data-driven methodologies will be advantageous for providing real-time risk assessment, such as the safety and efficacy of licensed drugs and the prioritization of repurposed drugs for clinical trials from preclinical to Phase III trials in humans. Additionally, such data-driven analyses can be used to interrogate the efficacy of vaccine candidates using a systems vaccinology approach. AI tools can analyze vaccine sequence information, such as the homology of potential antigens with human proteins, the ability to perform in silico host-pathogen interactions, and fortified epitope selection, including neoepitope protection that will lengthen vaccine coverage. These approaches are critical to exploit and validate, where applicable, extensive available data relevant to COVID-19. Preclinical and experimental data-driven analysis can also be used to analyze immune profiles of various vaccine candidates. The ethical and regulatory considerations to be taken into account for the application of these methodologies are discussed further under that heading.

Widespread use of AI in the vaccine development process will also have implications for data privacy. Given that AI relies on data input and its decision-making is data-driven, the proportion of data directly accessed by AI must be optimally manipulated to ensure appropriate data privacy. Given that AI will have an increasing role in fields where ultimate decisions made by AI often have a significant ethical and value dimension, transparency of the AI methodology is an increasing requirement upon society for the ethical acceptance of AI decisions. This is especially important, given that AI

methodologies will change over time, with numerous variations of AI and methodologies evolving individually. It is essential for stakeholders and the community to understand not just the output, but the method of the output calculation and the protocol or methodology behind that calculation. It is also essential that compliance is met for the use of human data in AI applications and that data protection laws are adhered to. Furthermore, there must be reporting in relation to the use, auditing, and management of personally identifiable data. It is recognized that AI-driven methodologies may encounter bias, as the AI program will produce results based on the data it is fed and on previously learned results. For many AI algorithms, bias can be minimized by using techniques such as reverse proxy bias methods and more abstract higher level feature spaces in the design of the application. Because the use of AI relies heavily on data sampling and has the potential to result in a range of outcomes, ethical guidance will need to be updated and developed in relation to this advancing technology. The use of such technologies in vaccine development should adhere to ethical standards. Finally, it is important to highlight that the increasingly global nature of vaccine and therapeutic development in response to pandemics has repercussions for international regulatory and ethical frameworks. To facilitate ongoing research and international collaboration, it will be essential that the development of ethical guidelines is proactive, clear, and transparent, and that guidelines are developed in concert with regulatory and research entities to facilitate vaccine nomination and development while upholding the highest ethical, regulatory, and scientific standards.

6. Future Direction

We are in a period of transition in the application of machine learning and data analytic research to expedite the development of vaccine products. In the near future, rather than using data analytic models to resolve problems associated with traditional approaches to vaccine R&D that utilize a hypothesis-driven approach, the next generation tools will be used to resolve issues germane to AI-supported vaccine research and development programs. For example, novel AI-based models are evolving to facilitate analysis of "-omic" data and expedite the research stage of system-based approaches to vaccine development. Further, as real-world data analytic approaches become more refined and gain in accuracy, there is the potential to use such data to provide an early indication of efficacy or safety issues associated with newly developed vaccine candidates.

The ability to harness the very large amount of data that are acquirable from numerous data sources and apply state-of-the-art AI approaches to make associations that spur development and answer pertinent questions has great potential. The need to improve analytic tools to conduct a broad set of activities, such as investigational product characterization, and to track the increasingly expansive complexity of vaccine development is imperative. The models we develop must be dynamic and iterative as vaccine development can be thought of as involving a "web of evidence" with the "center" essentially being unknown until the new vaccine is used globally. To accomplish this, research must focus on developing new machine learning tools to analyze certain types of natural language processing, identify critical content in public documents, as well as tools to mine social media for lay opinions about vaccines. New data analytic tools are then required to mesh these disparate data streams and integrate them with other structured data streams. We believe that these methods could add significant new value to the vaccine R&D process; however, the development of these methods is highly complex and will require cross-discipline collaboration. Furthermore, improved efforts to integrate across the public and private sector, as well as with global stakeholders, will be critical to ensure that our USV is built on a solid, adaptable foundation. Since these 21st century science methods will continue to evolve and be refined over time, we need to build an infrastructure that enables our methods and models to evolve in real time as new best practices are identified.

In the near future, AI-based solutions for vaccine R&D will likely evolve to encompass the analytics of electronic health records to conduct drug surveillance activities as part of pre- or post-market activities. We anticipate that just as systems biology and related concepts evolved through a number of iterations, this research on 21st century bioinformatics will evolve, as well as evolve into type 3 AI innovations. Fortunately, the proposed research that we contemplate will likely be able to evolve to keep pace with these new changes.

7. Conclusion

Our analysis highlights how AI and related methodologies can assist researchers in developing efficacious and safe vaccines in an efficient manner. Traditional vaccine discovery methods take several years in high-cost settings, and given the paradigm employed, are not guaranteed to scale from discovery to manufacturing and on to

clinical use. For identifying a suitable immune target, a precise and detailed understanding of the biology of the pathogen and immune system is required. Thus, vaccinology has always been "data-driven" and, as such, has much in common with big-data-driven machine learning approaches. Trials can often, therefore, take a "one dose fits all" approach and are often underpowered in their ability to account for subtleties in immune phenotype that might predict protection. These are problems that AI approaches excel at addressing: mapping high-dimensional non-linear probability distributions and extracting signal from noisy data. While AI is not a panacea, the case studies shown here of recent work to accelerate vaccines illustrate the potential improvements that AI can offer, moving away from a cost-inefficient and slow "one target, one trial" paradigm to a data-informed vaccine pipeline. However, AI is not neutral, and it brings with it a number of ethical considerations. There are also limitations in current approaches, mainly around availability and use of health data, highlighted particularly in the pediatric domain, and there is relatively little investment currently in R&D outside the field of immuno-oncology. The work described here is by no means exhaustive, but we hope it will show how the field can develop in the future to ensure that worldwide public health is better prepared for pathogens of epidemic and pandemic potential.