AI & Data Science for Vaccine Deployment

Leveraging AI and data science applications for improved vaccine coverage

DataKind®

Gavi
The Vaccine Alliance

Funded by Wellcome

datakind.org

Conflict of Interest Disclaimer: Mention of any specific example, company, organization, or solution included in this report does not serve as an endorsement by Gavi or DataKind®. All case studies are presented solely for the purpose of information sharing and to provide an appropriate acknowledgement of third party contributions. Gavi and DataKind recognize that some companies may have commercial interests in presenting a favorable public opinion of artificial intelligence solutions and therefore certain studies or examples may have some inherent bias associated. Gavi and DataKind have not made independent determinations about the accuracy of the underlying facts of case studies included in the report and disclaim all liability for any inaccuracies or incompleteness in the information in the report contributed by third parties.

Landscaping artificial intelligence and data science applications for the vaccine deployment life cycle

Abbreviations 3
Glossary 5
Abstract 7
I. Introduction 8
The challenge 8
Landscape report 8
Scope and strategic focus 9
Methodology 9
II. AI and data science for the vaccine deployment life cycle 11
Overview of thematic areas 11
Overview of AI and data science potential in each thematic area 12
Area: Vaccine supply chain (VSC) management 13
Area: Cold chain equipment (CCE) management 15
Area: Vaccine forecasting and waste management 16
Area: Vaccine demand generation and vaccine acceptance 19
Area: Health workforce management 22
Area: Vaccine campaigns, information platforms, and data ecosystems 23

Area: Immunization service delivery 26
Area: Data availability to support vaccination activities 27
Area: Data quality of vaccine-related information capture 29
III. Challenges and limitations of AI and data science in the vaccine life cycle 31
Introduction 31
Ethical considerations 31
Technical considerations 33
Adoption of AI and data science solutions 37
IV. Recommendations and conclusion 42
References 46
Appendix 57
Project partners 57
Stakeholder table 57
Stakeholder survey 58
Evaluation of the stakeholder survey 60

Abbreviations

Abbreviation	Definition
AI	Artificial intelligence
API	Application programming interface
CCE	Cold chain equipment
CDC	United States Centers for Disease Control and Prevention
DOT	Data Observation Toolkit
DRC	Democratic Republic of the Congo
EBS	Event-based surveillance
EIR	Electronic Immunization Registry
eLMIS	Electronic Logistics Management Information System
EPI	Expanded Programme on Immunization
FHIR	Fast Healthcare Interoperability Resources
G7	The Group of Seven (G7)
G20	The Group of Twenty (G20)
GHSC	The Global Health Security Consortium
GDPR	General Data Protection Regulation
INFUSE	Innovation for Uptake, Scale and Equity in Immunization
IoT	Internet of Things
LLM	Large language model
LMIC	Low- and middle-income country
ML	Machine learning
NLP	Natural language processing
PII	Personally identifiable information
VSC	Vaccine supply chain

VVM	Vaccine vial monitor
WHO	World Health Organization

Glossary

Term	**Definition**
Artificial intelligence	A field of computer science focused on creating systems capable of performing tasks that typically require human intelligence. These tasks include problem-solving, decision-making, understanding natural language, and recognizing patterns and images. Artificial intelligence systems can range from simple, rule-based algorithms to complex neural networks that mimic the way the human brain operates.
Data science	An interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. Data science combines aspects of statistics, computer science, and information science to analyze and interpret complex data. It encompasses data analysis, machine learning, and big data analytics, aiming to provide a basis for decision-making and strategic planning.
G7	The Group of Seven (G7) is an intergovernmental political and economic forum that has the largest and most advanced economies in the world, consisting of Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. The European Union is a "non-enumerated member".
G20	The Group of Twenty (G20) is a group of the world's largest economies that meets regularly to coordinate global policy on trade, health, climate, and other issues, including Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, the United Kingdom, the United States, the European Union, and the African Union.
Machine learning	A subset of artificial intelligence that involves the development of algorithms that can learn from and make predictions or decisions based on data. Unlike traditional programming, where instructions are explicitly provided to perform a task, machine learning algorithms adjust their parameters based on the patterns found in data, improving their performance over time without being explicitly programmed to do so.
Vaccine deployment	The term "vaccine deployment" is used in this report to refer to the comprehensive process of procuring vaccines from manufacturers or distribution centers, delivering them to health facilities and communities, and ensuring their safe and effective individual administration. It encompasses two primary components: 1. Vaccine delivery: This involves the logistical aspects of transporting vaccines, maintaining their cold chain, and distributing them to healthcare facilities, clinics, and vaccination sites. It includes storage, transportation, and supply chain management to ensure the vaccines are readily available.

	2. Vaccine administration: This step involves the actual provision of vaccines to individuals. Healthcare professionals and trained personnel administer the vaccines according to established schedules, dosages, and routes. This includes vaccination campaigns, routine immunization programs, and the monitoring of adverse events.
Zero-dose children	Children who have not received any routine vaccines. Gavi's operational definition of zero-dose children are those who lack the first dose of diphtheria-tetanus-pertussis containing vaccine (DTP1).

Abstract

Immunization is foundational to enabling good health and well-being for individuals worldwide. Globally, vaccinations can prevent dozens of life-threatening diseases and have saved millions of lives. Yet, despite advancements in vaccine technology and distribution, challenges to reaching global immunization targets persist.

These challenges occur in managing vaccine quality and supply across manufacturing and distribution, workforce training and development, communications, population targeting, and service delivery. Challenges in delivering vaccines and achieving vaccination coverage are exacerbated by various factors, including inadequate data coverage and quality, and megatrends, such as climate change. In the face of these challenges and the widening vaccination gaps, health teams around the globe are seeking innovative solutions to increase vaccine coverage.

Digital innovations are essential tools to close this widening gap, and global investment in digital technologies has highlighted the value of these tools. However, with a rapidly changing digital landscape, health actors need to be aware of the potential of data-driven solutions to be a part of broader national eHealth strategies. This report seeks to provide a primer on the value of artificial intelligence (AI), machine learning (ML), and data science as part of the digital innovation landscape and how health actors can apply these technologies to aid vaccine deployment.

Within this report, readers will encounter critical considerations for applying AI, ML, and data science across the vaccine life cycle. Further, this report will expose readers to opportunities to utilize this suite of technologies to improve vaccine delivery coverage, equity, and efficiency. Readers will explore the limitations of these technologies, some of the challenges and adoption barriers, and how to overcome them.

Finally, this report concludes with recommendations for implementing and sustaining AI, ML, and data science solutions in healthcare systems to support achieving universal health coverage and vaccination targets.

I. Introduction

Immunization is a cornerstone of universal health coverage and global health security.[1] Vaccines prevent millions of deaths and offer protection against over 20 life threatening diseases.[2] The measles vaccine alone prevented over 23 million deaths worldwide from 2000 to 2018.[3] Gavi-supported health systems have averted the potential future deaths of more than 17.3[4] million children in low- and middle-income countries (LMICs).

Over the last five years, many of these gains have been eroded. The COVID-19 pandemic resulted in the largest decline in childhood vaccinations in three decades,[5] leaving over 20.5 million children under-vaccinated,[6] including 10.2 million zero-dose children (children who have never received any vaccines). Inadequate healthcare resources, a shortage of trained workers, gender-related barriers, and global issues like conflict and climate change compound existing challenges in vaccine access and hinder equitable vaccine deployment.

To overcome these challenges, G7 and G20 countries prioritize harnessing digital innovation in immunization.[7] Using AI, ML, and data science is seen as a critical pathway. These technologies have already contributed to global healthcare - supporting healthcare workers,[8] strengthening health systems,[9] and improving patient diagnosis, treatment,[10] and disease monitoring.[11] In vaccine deployment, AI, ML and data science-powered solutions have the potential to contribute to enhanced vaccine coverage and positively affect human health and health equity. AI-driven algorithms are already being used to optimize distribution and forecasting,[12] while monitoring systems and sensors ensure real-time oversight of the vaccine cold chain.

The challenge

Successful implementation and scaling AI, ML, and data science interventions require a degree of shared knowledge and understanding of these technologies to foster collaboration among a range of stakeholders such that they can be tailored to individual country contexts, adhere to the regulatory landscapes, and be supported by sustainable infrastructure. However, key stakeholders face knowledge gaps and insufficient appreciation of the potential applications of AI and data science applications for vaccine deployment, including the optimal application, risks, and benefits. This hampers decision-making on prioritizing these investments, hindering the rapid and effective adoption of relevant technologies.

Landscape report

This report aims to address these knowledge gaps by raising awareness about innovative AI, ML, and data science tools for vaccine deployment - offering a discussion about their applications, limitations, and ethical considerations. It serves as a comprehensive resource for a global audience aiming to understand the current digital landscape and future potential of AI and data science in the field of vaccine deployment. This report also provides concrete use cases and highlights opportunities for technologies to enhance coverage, equity, and efficiency in vaccine deployment.

The primary audience includes stakeholders across geographic contexts directly engaged in vaccine deployment (e.g., frontline health organizations, governments, and intergovernmental bodies) and those that influence the field, such as policymakers, the private sector, technology companies, and funders (currently or aiming to be) involved in various aspects of vaccine deployment.

Scope and strategic focus

The report is focused on the deployment phase of the vaccine life cycle, encompassing vaccine delivery and administration, and landscapes key thematic and cross cutting areas, such as data quality and information platforms. The scope and thematic areas, summarized in the next section, were identified as priorities in stakeholder discussions with Gavi and its partners across the Gavi Alliance and Innovation for Uptake, Scale and Equity in Immunization (INFUSE) Pacesetters. AI tools are also accelerating vaccine research and development and vaccine production; however these areas are beyond the immediate scope of this report.[13,14]

Definition of AI, ML and data science
In this report, AI is defined as encompassing advanced computational and modeling techniques, involving the development of algorithms and models enabling machines to simulate human-like intelligence. AI systems can perform complex tasks, analyze and learn from vast and complex data sources, and generate insights, often mimicking or exceeding human capabilities.[15] Generative AI is defined as a category of AI models and algorithms that are designed to generate new content, such as text, images, and audio often based on patterns and examples they were trained on. ML, a subfield of AI, is defined as developing algorithms and statistical models facilitating computer learning and decision-making based on data. Data science is characterized as a multidisciplinary approach that combines mathematical and statistical methodologies, specialized programming, advanced analytics, AI, and ML with subject matter expertise, with the goal of extracting meaningful insights from data to inform decision-making.

Methodology

This report uses two core research approaches to landscape the use and potential applicability of existing AI, ML, and data science tools.

1. Desk research. A literature review was performed based on resources identified in a keyword search in Google Search and Google Scholar. To support interpretation and analysis, key findings were synthesized and tagged according to thematic area and AI, ML, and data science techniques. The list of potential resources, scientific studies, and publications was limited to publicly available and free of charge resources. Accordingly, potentially relevant sources requiring payment are not included in this report.
2. Stakeholder survey. A survey was developed and administered to Gavi teams and partners to gain perspectives from a broad range of stakeholders operating in a variety of geographies across the outlined priority areas. Groups currently working on identifying opportunities to use AI and data science-powered technologies across the vaccine

deployment life cycle were also engaged. These included Gavi's Health Systems Immunization Strengthening (HSIS) team, Gavi Alliance partners, including UNICEF and the World Health Organization (WHO), and stakeholders from the INFUSE Pacesetters. The topics covered in the survey included challenges in applying AI and data science technologies, such as data availability and use, opportunities for applying AI-powered solutions in vaccine deployment, and key ethical and technical considerations. Since stakeholders in the survey already collaborate with Gavi to varying degrees, perspectives on similar challenges from outside the immunization domain (e.g. cold chain challenges around the food supply chain) were not captured in this approach.

The findings were analyzed and synthesized to provide insights on the following:

1. Overview of the thematic area, emphasizing goals, key processes, and associated challenges.
2. Examination of current practices applying AI and data science technologies, opportunities, and challenges.
3. Benefits and positive impacts of AI and data science technologies, showcasing enhancements in efficiency and accuracy.
4. Risks and potential challenges associated with the application of AI and data science technologies.

Employing these methods facilitated a comprehensive landscaping of these areas, encompassing assessments of the application, maturity, and potential value and adoption barriers of AI and data science techniques in vaccine deployment to fully harness their potential.

II. AI and data science for the vaccine deployment life cycle

Overview of thematic areas

The thematic areas in this report are based on the vaccine deployment life cycle, illustrated in Figure 1, which comprises nine interconnected components.

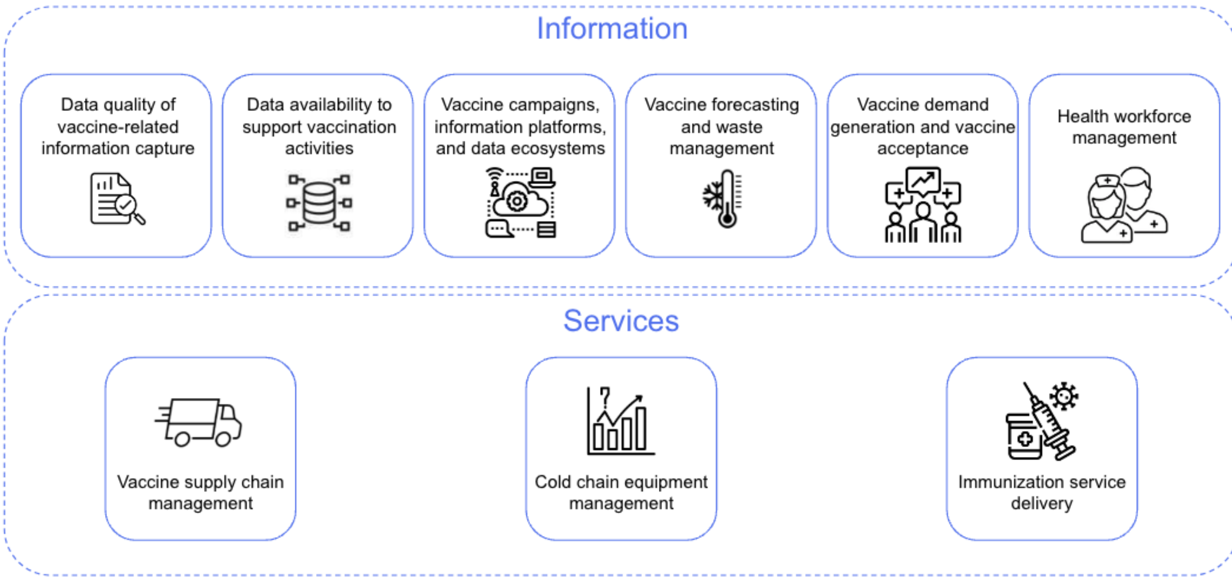

Figure 1: Vaccine deployment life cycle

For the purpose of this report, the areas are defined as:

Vaccine supply chain management	Activities, tools, resources, and planning necessary to ensure that vaccines stay safe and effective and reach all those who need them.[16]
Cold chain equipment management	Storing and transporting vaccines at recommended temperatures from the point of manufacture to the point of use.[17]
Vaccine forecasting and waste management	Ensuring efficient utilization of vaccines and preventing unnecessary loss by analyzing disease prevalence, demographics, and immunization schedules to forecast vaccine need, and minimizing discarding of unused or expired doses.
Vaccine demand generation and vaccine acceptance	Strategic initiatives and communication efforts to increase awareness, acceptance, and demand for vaccines within a target population.
Healthworkforce management	Efforts to train and retain skilled health workers through recruitment activities, workplace safety, gender equity, job satisfaction, performance management, and career development opportunities.[18]
Vaccine campaigns, information platforms and ecosystems	Coordinated national or subnational public health efforts to plan, implement, and disseminate information and reach vaccine target groups through various channels and technologies to ensure widespread coverage and awareness.[19]

Immunization service delivery	Systematic and organized administration of vaccines to individuals or communities, ensuring effective coverage and protection against preventable diseases.
Data availability	Timely and reliable access to and use of data by users whenever and wherever required.[20]
Data quality	Measurement of how well a dataset meets criteria for accuracy, precision, relevance, completeness, and timeliness.[21]

Table 1: Overview of thematic areas

Overview of AI and data science potential in each thematic area

From forecasting supply and demand to health immunization service delivery and health workforce management, AI and data science can play a pivotal role by addressing distinct challenges in each area. In supply chain management, ML algorithms, data pipelines, and dashboards can be utilized to monitor and optimize logistical processes, ensuring a robust supply chain, and leading to efficient and timely delivery of vaccines. Data science solutions can automate specific tasks, or support stakeholders in managing complex networks involving material, information, and resources, with AI specifically aiding in vehicle routing for distribution efficiency.

In Cold Chain Equipment (CCE), Internet of Things (IoT) devices, such as electronic time-temperature loggers, can be deployed to continuously monitor temperatures, enabling ML algorithms to dynamically adjust the cooling efforts, optimize the energy consumption of the system, and keep vaccines viable. Real-time data enables the automated flagging of potential temperature breaches, helping stakeholders to act quickly and prevent or reduce damages.

Knowledge of supply and demand at any given time is essential for determining order volumes and ensuring sufficient supply, while reducing wastage. In the area of vaccine forecasting and waste management, predictive modeling based on historical data, demographic information, vaccination rates, and public perception allows managers and health professionals to make informed decisions regarding supply distribution. Optimizing resource allocation is especially critical when demand outstrips supply, as exemplified by the shortage of cholera vaccines in 2022. This scarcity resulted in the temporary abandonment of the two-dose vaccine regimen.[22]

Similarly, predictive analytics can be utilized to optimize health workforce management processes. Reliable approximations, created by forecasting algorithms based on historical data, appointment rates, or observed trends in similar environments, allow managers to optimize the allocation and scheduling of staff to efficiently meet healthcare demands. Analysis of historical data regarding no-shows allows health institutions to strategically overbook appointments and ensure optimal utilization of staff and immunization equipment while minimizing wastage of these limited resources.

AI and data science can play a pivotal role in understanding and addressing misinformation and vaccine hesitancy in the area of vaccine demand generation. Through real-time content monitoring of social media posts and news outlets with Natural Language Processing (NLP) techniques, areas of concern can be detected and analyzed to inform counterstrategies. Additionally, tracking changes in public perception, as well as vaccination rates over time creates transparency about the effectiveness of information and promotional campaigns.

Access to high-quality data is key for any of the techniques presented in this report, from real-time monitoring to the training and application of ML algorithms. This transparency and sharing of data between the various stakeholders along the vaccine life cycle can be achieved through information platforms and data ecosystems. These integrate various data streams, aiding in efficient tracking of vaccine administration and verification, with examples like the United States Centers for Disease Control and Prevention's (CDC) North Star Architecture and decentralized open data platforms. The effectiveness of such an information platform and data ecosystem greatly depends on the number of stakeholders and the volume and quality of data they share on the platform.

Immunization service delivery, such as vaccination appointments, are one of the most important points of data collection. Data on vaccination status, demographic information, and potential adverse events are best collected by health institutions and vaccination sites, either during the appointment or through targeted post-appointment outreach initiatives. As mentioned above, high-quality data is essential for any ML algorithm, dashboard, or other monitoring system. Smartphone and tablet applications can support these processes to decrease the workload of healthcare professionals. Such applications can also improve data quality by guiding the user through the data capturing process and employing guardrails that, for example, only allow certain entries for specific fields. In scenarios where data quality cannot be ensured during the capturing process, automated data quality monitoring can mitigate the downstream effects of low-quality data, such as biased decision-making or wrong results of ML algorithms. Tools, such as the Data Observation Toolkit (DOT), can be deployed to highlight data quality issues and enable users to improve the reliability of the data over time.

Area: Vaccine supply chain (VSC) management

Introduction to the area

Supply chain management focuses on the design and coordination of relationships and flows of material, information, and resources across a network of organizations involved in logistical processes.[23] Supply chain management ensures stable and efficient processes to deliver the right product in the right quantity and condition to the right place for the right customer at the right time and price.[24]

The impact of supply chain problems in vaccine deployment can leave individuals and communities vulnerable to outbreaks. Delays and other challenges in vaccine delivery can expose people, especially the most vulnerable, to the threat of potentially deadly diseases and place a strain on the mental health and well-being of caregivers, such as a mother who is concerned about protecting her child's health. Beyond the individual, repeated supply chain disruptions can erode public confidence in healthcare systems and government agencies responsible for ensuring the availability of vital medications like vaccines. This loss of trust can have long-term consequences for public health initiatives, global health security, and overall health outcomes.

Potential AI, ML, and data science interventions

AI-informed route optimization for the VSC

A crucial part of transportation in general, and for last-mile delivery in particular, is the vehicle routing problem, namely the design and optimization of routes of individual vehicles or fleets to ensure operational efficiency in vaccine distribution. Simulation models digitally recreate various characteristics of the supply chain, such as fleet size and composition (e.g., vehicle type), and can include real-time information about stock levels or appointment schedules at different locations to inform operational and tactical decision-making. During the COVID-19 pandemic, a study in Norway modeled the national VSC, from arrival of the immunization materials in the country and initial storage to their allocation and distribution to all municipalities in the country. Through real-time tracking of transports, stock levels, and infection rates in different areas, as well as economic and ecological factors such as the cost and CO_2 emissions per transport, the model assisted in defining an allocation strategy of vaccination materials.[25]

While the study concluded that digital models of the VSC can improve operations and overall resource efficiency, this is dependent on linking a variety of supply chain data, such as shipping volumes, delivery timestamps, and transport routes to inform AI models. This application of AI is promising and a potential area for practical implementation, however, it requires accurate input for the models. In many low-resource countries, significant supply chain processes are not digitally and accurately monitored, and interoperability between IT systems makes data sharing complex and error prone. This could lead to model inaccuracies if foundational data availability and quality issues are not addressed first.

Risk management along the VSC

AI can significantly enhance the capacity for risk management in supply chain management and has been deployed with great success in cases, such as sentiment analysis of online news articles to understand regional risk patterns affecting the supply chain, as well as to evaluate the reliability of suppliers and forecasting demands.[26]

With known or common risks, such as flight delays due to weather conditions, AI models can be deployed for predicting specific supply chain disruptions.[27] ML algorithms can consume historical data and current trends to assign risk scores and predict potential failures in the supply network. This evaluative capacity allows for the creation of robust contingency plans, thereby ensuring the continuity of vaccine production and distribution even in the face of supply turbulence.

The predictive nature of AI-driven risk management also has the potential to identify previously unknown areas within the supply chain that may develop into risk factors and require additional investment in infrastructure or process improvement, such as creating stable and diversified supplies of raw materials and components. For example, American Airlines used ML algorithms on historic supplier data to predict the likelihood of shipments not arriving on time to be loaded onto a projected onward flight. This allowed the airline to plan cargo freights more efficiently, improve cargo space utilization, and decrease overall fuel consumption.[28]

Area: Cold chain equipment (CCE) management

Introduction to the area

Vaccines typically need storage at stable low temperatures, which pose challenges both in reliably ensuring these conditions and monitoring any breaches along the entire supply chain (i.e., transporting and storing the vaccines). If vaccines are exposed to improper conditions, such as too low or too high temperatures[1] or light, their potency is reduced or, in the case of below freezing temperature, can be destroyed entirely. Usage of a compromised vaccination might require revaccination which increases cost and might erode patients' confidence.[29] Accordingly, affected vials should no longer be used. This lack of usable vaccines and access leaves individuals and communities exposed to negative health effects that could have been mitigated with vaccination. The downstream effects of negative health impacts often result in lost economic productivity. The direct impacts of spoiled vaccines also result in increased cost for the replacement of spoiled vaccines and can cause broader supply chain disruptions by necessitating replacement of the spoiled material.

To mitigate the impacts of potential loss, the current state of the art CCEs include electronic time-temperature loggers (e.g., LogTags), freeze tags that alert users if vials exposed to sub-zero temperatures for over one hour, heat exposure monitoring strips, and vaccine vial monitors (VVMs) that measure the temperatures of individual vaccines and change colors in case of heat accumulation.[30] This monitoring allows for sustaining and maintaining vaccines and provides data that can be used in data-driven analytics approaches.

At the time of writing, the WHO provides planners and decision makers with a list of 75 pre-approved temperature monitoring devices, varying in cost and level of sophistication.[31] General procurement guidelines of CCE recommend that countries not only consider the initial cost of any devices, but also factor in any additional costs for installation, operation, maintenance, and potential replacements. The UNICEF Supply Division provides support to countries in the CCE ordering process through long-term arrangements with suppliers to ensure availability and price stability for a typical period of 24 months.[32]

Multiple case studies tested and evaluated the use of temperature monitoring devices in the VSC in LMICs. In Indonesia, a study focused on the transport and storage of hepatitis B vaccines concluded that the use of VVMs can reduce the risk of heat damage.[33] Another study in Thailand assessing the cold chain for measles and hepatitis B vaccines using logging devices found that excessively cold temperatures pose a more significant risk than heat exposure.[34]

---

[1] For example, the suitable temperature for the JYNNEOS Smallpox and Monkeypox Vaccine is 2°C to 8°C.

Potential AI, ML, and data science interventions

Deployment of real-time temperature loggers

The CCE and dry store temperature mapping tool, developed and maintained by the WHO, deploys electronic time-temperature loggers to assess the suitability of vaccine storage and transportation units, including warehouses, refrigerators, and transport boxes. Over a period of up to 72 hours, temperature loggers are placed at various locations within the unit being examined and the data is subsequently evaluated to determine how and where equipment can best be stored.[35] While this approach focuses on the general suitability of a specific cooling unit, it does not consider any anomalies (e.g., damage to a transport box during a transport) or changes of equipment over time. As such, the process supports strategic planning but lacks tactical and operational monitoring and analysis. This limitation is likely due to the capabilities of WHO-approved temperature loggers that only support wired connectivity. Accordingly, the loggers have to be manually collected and connected to a computer to extract the readings. Wireless sensors and a digital platform to monitor temperatures could address these limitations and allow for a continuous monitoring process that enables operational adjustments which ultimately reduces damages to, and accordingly wastage of, vaccine equipment.[36,37]

Through the use of real-time loggers, up-to-date time series of temperature information can be captured and made available for continuous analysis and monitoring. Based on this data, ML algorithms can be applied to predict CCE failures (predictive maintenance)[38] and give technicians and engineers sufficient lead time for replacement to avoid down times or damages to the vaccines.

Dynamic cooling to increase efficiency and lower cost

Typically, refrigeration control systems aim at maintaining stable temperatures and manage the cooling process solely based on this indicator. While this approach ensures that temperatures stay within the necessary ranges, it does not consider energy efficiency and the resulting economic factors, especially relevant in low-resource environments. Provided with additional information about the environment, such as the ground temperatures at a warehouse location, the materials of various components (e.g., heat capacity of refrigerators or transport boxes), and historic or projected cost of power consumption, AI models can combine data about cooling cycles and external conditions, such as temperature and energy prices, to assist in designing operating models and cooling strategies that ensure adequate temperatures at a lower cost. This could be achieved, for example, by over-chilling a warehouse or transport unit in times of low energy cost and warming it up (within given thresholds) in peak consumption times with higher prices.[39]

Area: Vaccine forecasting and waste management

Introduction to the area

Accurately forecasting vaccine demand and consequently reducing vaccine wastage are critical components of effective immunization programs, particularly when resources are limited and there is high global demand for vaccines. Vaccine waste due to damage, such as cold chain disruption, is covered in the previous section, and here - the effects of wrongly anticipated supply and demand are discussed.

Oversupply of vaccination material leads to wastage and increased cost due to inefficient usage of CCE capacities, while undersupply can cause stock-outs and missed vaccination opportunities.[40] Accurate demand forecasting and reducing vaccine wastage aligns with the focus areas of "vaccine forecasting, procurement, and supply" and "sufficient, predictable resources" outlined in the Immunization Agenda 2030.[41] As vaccine wastage cannot be entirely eliminated, Gavi recommends a target of no less than 25% wastage in the first year of a vaccine and 15% for subsequent years on a national level. For single- or two-dose vials, this number is substantially lower at only 5%.[42]

These are aggressive targets, as waste has been estimated at as much as 33%,[43] and as such, novel approaches, like utilizing predictive ML algorithms to forecast wastage and analyzing data to understand the root causes, should be considered.

Potential AI, ML, and data science interventions

Reducing vaccine wastage through accurate demand forecasting
One potential solution to lower vaccine wastage is to accurately predict the number of vaccines required at various distribution points to ensure optimal availability without significant excess leading to wastage. Reliable forecasts of vaccine demands can inform stakeholders, such as vaccine manufacturers and health facilities, in the design and operational planning of sourcing, such as ordering and delivery planning, and manufacturing processes. Furthermore, comparisons of vaccine demand and supply forecasts can highlight bottlenecks along the supply chain that cause disruptions and inefficiencies. Key processes in vaccine forecasting involve analyzing various data points, such as historical vaccination data, demographic developments, current disease incidence, and the effect of immunization campaigns.

Due to limited access to reliable forecasting models, health clinics, especially in low-resource areas, often rely solely on historical data, such as past vaccination rates, to inform order volumes, which is problematic. However, this approach is problematic as it assumes future demand will mirror past numbers, failing to consider sudden changes in vaccine demand or supply chain disruptions during events like pandemics. Implementing buffer stocks can mitigate this issue, but resource shortages may prevent its feasibility.

AI can enhance demand forecasting by incorporating real-time data, such as disease surveillance and weather patterns, which impact disease spread and, consequently, vaccine demand. The forecasting accuracy of these algorithms and models could potentially be further improved by incorporating expected incident numbers as modeled by the Vaccine Impact Modeling Consortium.[44] However, the effectiveness of these tools is contingent upon the quality and granularity of available data.

To improve forecast quality, the Vaccine Modeling Initiative developed spreadsheet-based simulation models, considering different timeframes of historical data, arrival rates, vial sizes, and buffer stock levels.[40] Although this approach likely benefits from the broad adoption and usage of tools, such as Excel, the tools often rely on human input of estimated demands.[45]

ML models are being piloted to forecast vaccine uptake rates and adjust supply chain dynamics accordingly. In a case study in Niger, Mueller et al.[46] utilized the Python-based software platform HERMES to implement a detailed simulation model as a virtual representation of the Expanded Programme on Immunization (EPI) VSC, including various storage locations and health institutions. The study reports increased vaccine availability from 69% to 100% and a cost reduction of up to 34%.

Challenges in using AI and data science to forecast vaccine demands include data availability, variability in disease outbreaks, logistical constraints, and uncertainties in vaccine uptake. While individual stakeholders along the supply chain, such as vaccine manufacturers, might implement various forecasts on a company level, this information is typically not shared between different entities, hindering transparency and collaboration. Therefore, the Global Health Security Consortium (GHSC) proposes a neutral broker of information to increase forecast quality on a global level. Global cooperation of governments, health organizations, and the private sector, based on centralized demand forecasting, such as forecasting on a national or subnational level, can enable efficient resource allocation, policymaking, and the creation of incentives to ultimately increase vaccine availability and coverage.[47]

Event-based monitoring/surveillance

Event-based surveillance (EBS) involves the use of non-traditional data sources, such as social media, news articles, and other online information, to detect and monitor disease-related events. AI tools, especially NLP and ML, are instrumental in this process, as they can efficiently scan and analyze large volumes of data for potential indicators of disease outbreaks.

One notable application of AI in EBS was during the early stages of the COVID-19 pandemic. For instance, an EBS engine developed by BlueDot, which incorporated AI capabilities, was able to identify and flag a local media report about the emerging disease in Wuhan, China, on December 31, 2019. This allowed BlueDot to notify their clients about the potential outbreak five days before the WHO made a formal announcement.[48,49]

Such AI-powered surveillance tools are critical for early detection, offering an advantage in responding swiftly to emerging public health threats. They complement traditional surveillance methods, providing early warning signs and enriching the data pool with real-time, diverse information sources. However, it is crucial to balance AI insights with human expertise for validation and contextual understanding, ensuring a comprehensive and accurate approach to disease surveillance.

Utilizing ML to forecast demand and understand driving factors

The integration of AI and data science in vaccine demand forecasting and wastage management offers several benefits: AI models can process complex, real-time (if available) datasets to provide more accurate forecasts, leading to better vaccine allocation and reduced wastage. Through continuous re-evaluation and improvements, ML algorithms can improve over time, adapting to changing patterns in vaccine uptake and disease incidence. By reducing vaccine wastage and optimizing supply chains, AI-driven models can significantly lower operational costs and increase vaccine availability.

ML models can not only be utilized to forecast supply and demand metrics but also to gain an understanding of the importance of the various contributing factors as showcased in a case study in Tanzania. In that case, one model determined the three-month rolling average of vaccine utilization, and factors related to the health facilities to have the highest impact on the model. However, it is important to note that feature importance has to be understood as a characteristic of a given model, not the object of study itself, as a different model in the same study identified the vaccine type to be the most important feature. By utilizing daily vaccination information from 710 health facilities and a regression model, the study aimed at bi-weekly forecasts on a health facility level.[50]

While reliable forecasts can support resource allocation on a national and subnational level, societal factors, such as gender and age, can still impact vaccination rates. For example, a study in The Gambia found that despite a sufficient supply of immunization material, vaccines remained unavailable to some women and children. This underscores the significance of addressing equity and inequities to ensure the effectiveness of vaccination initiatives.[40]

Area: Vaccine demand generation and vaccine acceptance

Introduction to the area

Vaccine demand is the desire or willingness of individuals in a community to receive vaccination. Vaccine demand on an individual human or household level is highly connected to a decision-maker's vaccine hesitancy (i.e., reasons why individuals choose not to get vaccinated). Instead of seeing vaccine acceptance as simply being for or against vaccination, it should be understood as a range of opinions. On this spectrum, individuals' positions may shift in response to external influences, such as new information or information presented in a different way. Through data analysis, the movements of individuals and communities along the continuum and their dependencies on external factors can be understood, which allows the formulation, implementation, and evaluation of corrective actions.

The Health Section of the UNICEF Programme division defines the spectrum of vaccine acceptance as depicted in Figure 2 below.

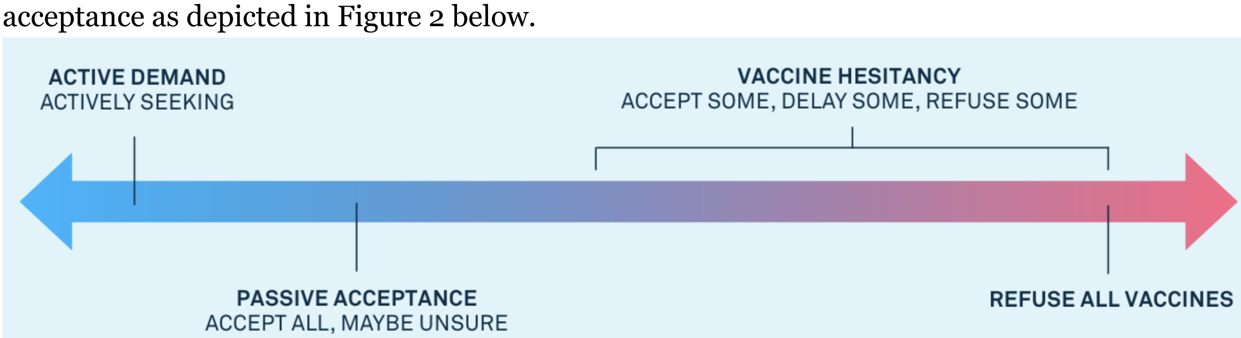

Figure 2: Continuum of vaccine acceptance sourced from UNICEF's Vaccine Messaging Guide[51]

In recent years, the spread of disinformation and misinformation[2] on social media and other forms of digital communication, referred to as an infodemic, has increased vaccine hesitancy, which was ranked as a threat to public health by the WHO in 2019 and continues to persist as a threat.

The Health Section of the UNICEF Programme division suggests the following iterative, four-step process to manage vaccine misinformation:

1) Prepare: Review of the information ecosystem (both online and offline) lays the foundation for any of the following phases by providing an understanding of the relevant information channels, and narratives, as well as metrics to quantify them.
2) Listen: This phase focuses on the design and implementation of a social listening system that enables constant and efficient monitoring of the information ecosystem to detect early signals and emerging patterns around vaccine misinformation.
3) Understand: Analysis of the signals and specific content collected in the previous phase, leads to a comprehensive understanding of the actors that spread misinformation, their motivation, the target audience, and the resulting risks for the public perception of a given topic. This situational understanding allows the formulation of strategies and actionable recommendations to address the issue.
4) Engage: The recommendations are implemented to prevent or debunk misinformation and strengthen the system by (re)building public trust. The implementation and monitoring of quantifiable metrics regarding the impact of a particular engagement are key to success. Accordingly, phases two to four should form a continuous process that dynamically addresses new challenges as they arise.[52]

Data-driven approaches can help identify positive demand responses to awareness campaigns and policy, as well as identify potential sources of vaccine hesitancy; in both cases, this information can inform vaccination campaign decisions.

Potential AI, ML, and data science interventions

Real-time content monitoring

AI and data science tools can automate the gathering and pre-processing of information (e.g., the 'Listen' step in the above UNICEF process), reducing complexity and manual workload. Especially during the COVID-19 pandemic, various studies to scan social media content for immunization-related misinformation and vaccine hesitancy were conducted. For example, an observational study conducted in the United Kingdom and the United States analyzed information from over 300,000 social media posts using AI algorithms and concluded that this approach, coupled with surveys and other conventional methods of assessing public attitude, could help address the concerns around vaccines to maximize uptake.[53]

---

[2] While both terms describe the spread of false information, disinformation constitutes deliberate misinformation to achieve a shift in public perception.

Compared to one-time studies that require a static dataset, a continuous process requires enhancements of the data collection by integrating it into a data pipeline that periodically collects and saves new content as it becomes available. The resulting data source would allow not only for the analysis of close to real-time information but also inform the evaluation of any initiatives to address vaccine misinformation that have been implemented.

Once the data has been collected, ML algorithms can be applied to distill relevant information and support the "Understand" step. Classification models can, for example, be used to detect misinformation and gauge public sentiment across multiple topics.[54]

The selection of the technique highly depends on the use case: while a general understanding of the areas of concern could, for example, be achieved through topic modeling, tracking changes in the public perception of immunization programs might require different approaches, such as sentiment analysis.

Through the mining and analysis of news and social media information with NLP techniques and large language models (LLMs), the causes of vaccine hesitancy can be better understood and addressed. The knowledge gained from these analyses, paired with generative AI, allows targeted information campaigns to address vaccine hesitancy and increasing demand for vaccines.

Analysis of medical data, such as vaccination appointments

LLMs can assist professionals in the analysis of medical data and information about vaccination appointments. Their ability to process and organize large volumes of data, including doctors' notes and patient records, can enhance data interoperability in healthcare settings. LLMs can develop specialized knowledge in various medical disciplines by analyzing medical data. This includes the capacity to be fine-tuned on domain-specific medical literature, ensuring they remain current and relevant; the models can also be adapted to different languages and contexts.[55] Utilizing this approach, stakeholders can develop a better understanding of public concerns regarding specific vaccines, which in turn allows them to appropriately formulate targeted campaigns and measure these campaigns' effectiveness.

Providing personalized information about immunization

Beyond responding to common concerns, misconceptions, or misinformation, AI can be used to create positive and proactive personalized information. Many of today's LLMs, such as Anthropic's ClaudeAI or OpenAI's ChatGTP, provide a simple chat interface to allow non-technical users to communicate with the model.

In the era of digital health communication, chatbots have emerged as an innovative tool, particularly highlighted during the COVID-19 pandemic. They have shown potential in delivering personalized and timely information, crucial in managing public health crises. A noteworthy example of this application is seen in a randomized, controlled trial conducted across Thailand, Hong Kong, and Singapore.[56] This study evaluated the impact of COVID-19 vaccine chatbots on the confidence and acceptance of vaccines among adult guardians of children and seniors, particularly targeting those hesitant or delaying vaccination.

The findings from the study presented a nuanced picture. In Thailand, individuals interacting with the chatbot exhibited a smaller decrease in confidence regarding the vaccine's effectiveness compared to those who did not use the chatbot. However, the situation was different in Hong Kong and Singapore, where some chatbot users reported a decrease in vaccine acceptance and confidence in vaccine safety. These varied outcomes underscore the complexity and the need for tailored approaches in using chatbots for health communication, as their effectiveness can significantly differ based on cultural and societal contexts.

A critical insight from the study was the influence of sociodemographic factors on the effectiveness of chatbots. Notably, underrepresented groups and individuals with lower education levels showed more pronounced improvements in vaccine confidence and acceptance when using chatbots. This suggests that chatbots could be an effective medium for reaching and positively influencing underrepresented or less-educated populations, who are often at the periphery of traditional health communication strategies.

Additionally, LLMs' ability to analyze patient data, medical records, and physician inputs can lead to the automatic generation of comprehensive and structured reports.[3] This not only saves time but also reduces the risk of errors, a critical factor in medical data management.[57]

As in all data analysis applications, there are potential challenges and risks with LLMs, and those are outlined in section III.

Area: Health workforce management

Introduction to the area

In the context of this report, workforce management can be divided into two areas: training of staff to ensure high-quality service delivery and resource allocation, for example the assignment of personnel to specific shifts based on the expected workload and availability.

Workforce management tools are utilized to digitize paper-based, manual processes, such as workload forecasting and staff scheduling, to ensure adequate staffing and fair treatment of personnel. Where available, forecasts around vaccine uptake and availability (see section, Area: Vaccine forecasting and waste management), as well as real-time monitoring of workloads and staffing levels, can increase the system's accuracy and allow decision-makers on a health facility level to optimize their operations.[58,59]

Potential AI, ML and data science solutions

Effective health workforce management in immunization can be greatly facilitated by leveraging AI and data science tools across various phases and key functions of the vaccine life cycle. Establishing a robust data pipeline is essential for real-time analysis, involving the importation of pertinent data through application programming interfaces (APIs) or connected tools. This serves

[3] Of particular concern are privacy issues with such sensitive data. See Section III: Challenges and limitations of AI and data science in the vaccine life cycle.

as a foundational step for subsequent AI-based or manual analyses. The integration of dashboards, automated reports, and continuous metric monitoring, such as vaccination rates or coverage within specific communities, enhances reporting capabilities. Moreover, forecasting models, powered by ML algorithms like regression models, enable the prediction of future events. For instance, by analyzing a rolling average of disease incidents and other input variables, these models can forecast an increase in vaccination appointments. In the planning and scheduling phases, the outputs from previous steps, such as vaccination appointment forecasts, empower managers and health professionals to adjust staffing based on anticipated workloads. This strategic approach not only helps in optimizing healthcare worker efficiency but also ensures adequate patient care and minimizes wait times.[55]

In addition to the above, generative AI can support the training of medical staff and assist in daily operations, for example by automatically generating knowledge tests and streamlining information retrieval from training resources.

Through the use of generative AI, such as enabling natural language queries and interactive data analysis, health institutions in the United States expect to streamline their daily workflows, increase productivity, and focus health professionals on their core duties instead of manual data entry, maintenance, and analysis.[55]

Area: Vaccine campaigns, information platforms, and data ecosystems

Introduction to the area

Vaccine campaigns are a delivery strategy used to quickly reach large numbers of individuals with one or more vaccines and can span the entire vaccination deployment life cycle, from identifying target populations and training the health workforce to monitoring and reporting the progress of vaccination efforts. To establish a functional and scalable vaccination infrastructure, both at local and global levels, it is important to allow seamless information sharing among the multiple stakeholders. This encompasses various processes and IT systems, necessitating the implementation of efficient, robust, and secure data flows. The integration of AI within this framework can significantly enhance data processing capabilities, informing timely and accurate decision-making. Such an approach is crucial for optimizing vaccine distribution and administration, thereby improving overall public health outcomes.

Potential AI, ML, and data science interventions

Vaccine campaigns

The goal of vaccination campaigns is to reduce outbreaks and strengthen population immunity by prioritizing access to follow-up vaccinations, particularly for under-immunized children in humanitarian or hard-to-reach contexts. This can be achieved through routine immunizations, or supplementary immunization activities.

Although both approaches are well established, the spatial variation of their effectiveness and degree of coverage is often unknown, especially in rural areas of LMICs. Combining vaccination rates and demographic information and applying geospatial analysis can help highlight particular areas of low immunization coverage, informing targeted outreach programs and campaigns to ensure full immunization (i.e., administration of subsequent doses of a vaccine where necessary).

A study of various vaccine types (e.g., against polio and measles) in Nigeria, Ethiopia, the Democratic Republic of the Congo (DRC), Cambodia, and Mozambique predicted the spatial distribution of vaccine coverage as depicted in Figure 3. Furthermore, the researchers applied covariate analyses to determine the variables that negatively impact coverage. While many of the contributing factors varied between the different countries, travel duration and distance to the nearest main settlement negatively impacted immunization rates ubiquitously, underlining the vulnerability of remote communities.

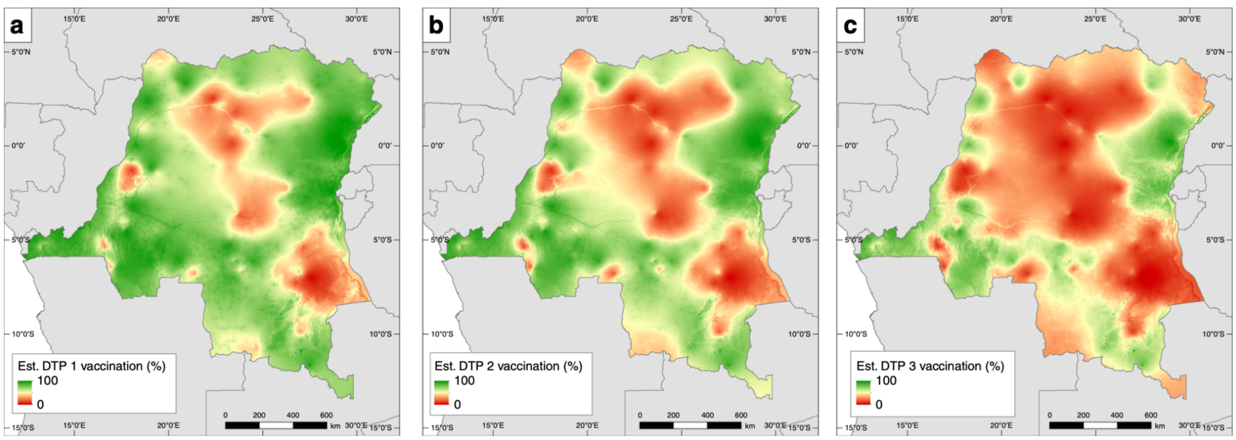

Figure 3: Estimated diphtheria-tetanus-pertussis vaccine dose 1, 2, and 3 vaccination coverage in children under five years old at 1 × 1 km resolution for DRC in 2013–2014[60]

Information platforms and data ecosystems

In both the literature review and stakeholder survey conducted for this research, data availability and transparency along the entire vaccine deployment life cycle, including vaccination rates within communities, demand, and stock levels of immunization materials in different health facilities or administrative regions, and delivery schedules of logistic service providers, were repeatedly flagged as key challenges regarding the application of AI and data science tools. Since decisions regarding vaccination strategies are often made on a national level but are implemented on a local level, this transparency is crucial to monitor the impact of policies on various levels and across time. The lack of standardized and interconnected systems across the United States, for example poses a challenge to efficiently tracking vaccine administration and verification.[61] Similarly, an analysis of the data saved in the VaxTrac electronic immunization registry in Sierra Leone concluded that data sharing restrictions and inconsistent workflows between different health facilities are major implementation challenges and negatively impact data quality even within the same software system.[62] Acknowledging the challenges and downsides of multiple district vaccine data management tools in Tanzania, such as low-quality, consistency, accessibility, and availability, the different systems were combined into a single vaccine information management system.[63]

Although emerging technologies, such as LLMs, show potential around in context learning (i.e., not relying on dedicated training data for a given task), data availability and quality is generally a potentially limiting factor in data science applications.

Recognizing the challenge above, the CDC has designed the "North Star Architecture" as a common framework to provide and access data and enable informed decision-making. By integrating data from various data streams across different stakeholders, space, and time, the framework allows data sharing between the different parties, reducing manual workload and increasing response times.[64]

Together with the Virginia Department of Health in the United States, the CDC piloted a project to implement the North Star Architecture in the form of a modular, cloud-based, configurable data pipeline that combines different data streams on lab reporting, case reporting, and vaccines. After pre-processing, quality control (e.g., through geocoding information, deleting of duplicate entries, and joining of datasets), and standardization, the prototype was able to join the formerly dispersed information to allow for analysis and actionable insights. The research team concluded that the prototype saves time and manual effort, increases processing time, and creates a single source of truth for any analysis.[65]

Alternatively, to the North Star Architecture approach, LLMs show potential in data engineering tasks, such as combining disparate datasets or cleaning and processing unstructured data. Through training of foundation models, such as GPT-4, LLMs can be utilized in the extraction, transformation, and loading of data. Normalized datasets that adhere to a common structure can easily be merged and stored on the same platform for analysis, reporting, or further processing. Furthermore, LLMs can assist in the creation and optimization of complex structured query language queries. Lastly, LLMs can be utilized to create transparency regarding data lineage, for example by creating lineage diagrams that illustrate the flow of data, making it easier for humans to analyze dependencies and relationships between datasets.[66]

Regarding data analysis, LLMs can empower non-technical users by answering their questions (formulated in natural language) through a conversational interface. Open-source frameworks, such as LangChain, allow organizations to connect their own data sources to be analyzed by the LLM.[67]

During the COVID-19 pandemic, a team of researchers piloted the COVID-19-Vaccine Platform, a decentralized, open data platform on vaccinations in Germany to facilitate transparency and data sharing between regional and local actors while adhering to data security protocols and the European General Data Protection Regulation (GDPR). The research team recognized that beyond a technical solution, an ecosystem of roles, responsibilities, and processes is needed to maximize the systems' overall efficiency.[68]

Development of information products

Information and communication are crucial components of various stages of the vaccine life cycle, from vaccine campaigns to educating communities about vaccines to direct communication between health workers and patients. This important but labor-intensive task can cause additional workload for health professionals and can require skill sets from other domains, such as public relations, marketing, and psychology. Information products can include bulletins, situation reports, slide decks, dashboards, and infographics.

Through the generation of text or semi-automated analysis, generative AI tools could support policymakers, managers, and healthcare professionals both in productivity and access to information.

Area: Immunization service delivery

Introduction to the area

Immunization service delivery focuses on the administration of a vaccine to a patient, spanning from appointment scheduling through capturing of vaccine data to monitoring of adverse events following the immunization. The application of AI and data science in the realm of vaccine service delivery represents a significant advancement in healthcare management. This approach not only enhances the efficiency of vaccine administration but also plays a crucial role in data collection, ensuring the optimization of resources and the minimization of risks.

Potential AI, ML, and data science interventions

Optimization of appointment scheduling

One of the key challenges in vaccine administration is managing appointment schedules to minimize wastage and maximize vaccine utilization. This already-difficult task is further complicated by no-shows that lead to a risk of the budgeted vials remaining unused and resulting in wastage. This is especially crucial if multi-dose vials that can serve numerous individuals are opened by a health facility expecting multiple patients, as some may miss their appointment. With manufacturers typically recommending all doses in a vial be used within 24-48 hours of opening, this leaves little time for the vaccination site to fully utilize the vaccines and can cause ripple effects for scheduling. Publishing of available doses daily by vaccination sites, paired with targeted outreach programs to increase demand, can reduce this type of wastage. A robust data pipeline is needed to gather, aggregate, and publish this information in close to real-time.

Alternatively, AI algorithms can use historical vaccination and appointment data to predict the likelihood of no-shows and allow efficient overbooking of appointments. For example, through strategic overbooking, the Canadian Province of Alberta reported to have lowered vaccine wastage rates for a COVID-19 vaccine to 3%.[42] In Tanzania, the “Connected Health AI Network vaccine forecasting tool” efficiently matched supply and demand on a health facility level, reducing vaccine wastage by 96%.[69]

Post-vaccination health monitoring

Post-vaccination monitoring is another critical aspect where AI and data science play a pivotal role. AI algorithms can sift through vast amounts of patient data to identify patterns and flag potential adverse events following vaccination. A research team at MIT's Computer Science and Artificial Intelligence Lab applied ML predictions to clinical data to investigate the efficacy of a specific vaccination on different populations.[70] This proactive approach to monitoring not only ensures patient safety but also contributes to the overall efficacy and trust in vaccination programs.[71]

Designing a network of vaccination facilities
In regions or situations where a new vaccination infrastructure must be set up, such as establishing a new network of vaccination sites, based on (expected) demands, prior analysis of key factors, including existing health infrastructure and population density, is essential to ensuring equitable and efficient design. For example, many high-income countries adapted the following approach during the COVID-19 pandemic:

+ Existing health facilities, including hospitals and primary care physician offices, were used during the initial ramp up of vaccination efforts.
+ Subsequently, additional vaccination sites were added in densely populated areas to relieve pressure on the existing facilities and reach as many people as possible.
+ In the last wave, mobile solutions, such as buses or smaller sites, were used to reach specific or underserved populations.[72]

Area: Data availability to support vaccination activities

Introduction to the area

The literature review and stakeholder interviews identified the scarcity of high-quality data as one of the primary challenges in the deployment of data science solutions for vaccine management. As high-quality data lays the foundation for data-based decision-making, as well as the training of any ML model, this is one of the primary bottlenecks and challenges and should therefore be prioritized in any data project or initiative.

To systematically address issues around data availability, the three following methods of digital data capturing should be considered:

+ Manual data collection - staff such as community health workers manually collect and save information, such as vaccination appointments, and surveys around vaccination status.
+ Automated data collection - implemented through devices, such as trackers and data loggers, as described in the CCE Management section. For example, location trackers in the fleet of a transportation network allow dynamic route optimization for individual vehicles.
+ Data shared by partners or other organizations within the same industry - the implementation of a shared platform as described in the Vaccine Campaigns, Information Platforms, and Data Ecosystems sections could further increase data availability. In the absence of such a platform, partnering organizations can be connected via APIs to allow

data transfer. Alternatively, open datasets can be sourced from platforms, such as the WHO Global Health Observatory,[73] UNICEF,[6] or the Global Health Data Exchange.[74] Due to the dynamic nature of the vaccination landscape and the differing focus of the individual platforms and datasets, data from different sources can be joined to allow multifaceted analyses.

In many regions of the world, a significant number of individuals without birth certificates or "unaccounted" individuals pose significant challenges to consistent and reliable data capture, as noted by an assessment study of the VaxTrack immunization in Sierra Leone.[62] The Better Immunization Data (BID) initiative underscores the importance of proactively addressing this by involving a diverse range of stakeholders in the data capture process, including community health workers, facility staff, and district information officers to support closing vaccine data gaps.[75]

Potential AI, ML, and data science interventions

Manual data collection can be streamlined and improved by addressing the data literacy of the people tasked with the collection, and improving the data capture tools. Understanding the value of data capture and how this data is used (i.e., the necessary information and quality standards the data must adhere to) can help increase motivation, the stringency of the capturing process, and the overall quality of the data. For example, the use of tablets decreases friction in the process and the workload for the people involved. The absence of such tools requires multi-step processes, involving multiple parties and ultimately resulting in a higher probability of miscaptions. The application of optical character recognition can support staff and increase efficiency in transferring handwritten information into digital data.[76] However, especially in high-risk medical settings, this process should still feature continuous human interaction for quality control.

Automated data collection through IoT devices reduces human involvement in the initial setup of the data infrastructure and continuous monitoring. This has implications for the required skill sets, which might be a limiting factor in low-resource environments. While data can be captured by following defined processes in a standard operating procedure or following instructions presented by the capturing device, setting up the data infrastructure requires technical expertise, as well as domain-specific knowledge. Furthermore, biometric systems, such as VaxTrack, can support ensuring vaccine adherence, but acceptance of such technology is crucial, and ethical challenges arise from their application.

Technical constraints such as low storage capacity, and restricted internet access or bandwidth in low-resource environments pose challenges to data storage and ultimately to data availability, which lays the foundation for any AI and data science application. For example, if data is collected through mobile devices, such as tablets, transferring this data to a centralized data storing device, such as a data warehouse, where it is consolidated and used for decision-making or process automating might not always be feasible. IT architectures, such as edge computing, can address this challenge by shifting parts of the computation and storage capacity of the overall network to the end clients. With this approach, only consolidated metrics need to be transferred to the central data center, thus reducing the bandwidth requirements for the transfer.[77]

Sharing data not only increases transparency but also fosters the development of synergies. It leverages collective knowledge, allowing different entities to benefit from shared insights and innovations. This collaborative approach can significantly strengthen the overall system or network, leading to more effective vaccine distribution, monitoring, and administration strategies. However, competitive dynamics often pose a significant challenge to this ideal. Companies and organizations, driven by market competition and proprietary interests, tend to be protective of their data. This reluctance to share data can stem from concerns over losing competitive advantage, intellectual property issues, or data privacy regulations. In such an environment, the adoption of a collaborative data-sharing model is hindered. One potential solution to this challenge is the implementation of an objective third party that can act as a mediator and custodian of shared data. The role of this third party would be to ensure that data sharing adheres to agreed-upon standards, maintaining data privacy and security while facilitating access to necessary information.

Area: Data quality of vaccine-related information capture

Introduction to the area

In the stakeholder survey, six out of 10 organizations from various locations and industries reported data availability and quality to be the primary challenges when applying AI and data science technologies across the immunization life cycle.

Low-quality data can directly impact the accuracy and reliability of ML models, as well as reports and monitoring systems pertaining to real-world events, processes, or situations. Such discrepancies can significantly erode trust in data-driven applications, as stakeholders may question the integrity and utility of the data underpinning these systems. This erosion of trust can potentially lead to a broader skepticism towards the value of data-driven insights. Consequently, users may be less inclined to engage with data-enabled tools in an effective and informed manner, undermining the potential benefits these technologies offer.

There are two dimensions to the challenge of low-quality data:

+ If the data is used to inform decision-making, low-quality data can misrepresent the situation, increasing the risk of wrong decisions.
+ If data is used to train an AI model and/or automate processes, low-quality data inevitably leads to biased model output and incorrect results.

Data quality is a multifaceted issue, and understanding its different scenarios is crucial for effective data management. Generally, data quality concerns can be categorized into intrinsic, and contextual/relational scenarios. Intrinsic quality issues refer to errors in individual data points, such as missing values, wrong data types, entries outside of a predefined value range, or plain miscalculations, such as incorrectly capturing patient names in a vaccination appointments dataset. Contextual quality issues, while individually accurate, violate certain contextual conditions or expectations. For instance, the height information of a patient of 1.8 meters might generally be plausible but becomes questionable if the age information of this patient is recorded as six months. A good understanding of the data that is being captured, as well as the underlying processes, is key to defining the relevant data quality scenarios and monitoring strategies.

Potential AI, ML, and data science interventions

Data Observation Toolkit

Acknowledging the importance of systematically improving data quality, DataKind, in cooperation with globally distributed frontline health partners, Ministries of Health, frontline health workers, and funders, developed DOT.[78]

The DOT is an open-source, community-informed toolkit capable of automated monitoring and detection of inconsistent or problematic data in a relational database. DOT is designed such that it can sit as close as possible to the point at which health-gathered data by frontline health workers using digital tools syncs with the server and enters the database, at which point a series of tests can be applied.

At its core, DOT uses two powerful data integrity and validation libraries—DBT and Great Expectations.[79] Many out-of-the-box tests from both libraries are provided for classic data quality scenarios and for common scenarios related to the community health domain, such as specific protocols for the community case management of childhood diseases (malaria, pneumonia, and diarrhea) and maternal, newborn, and child health.

DOT also provides a simplified user interface as a management layer where tests can be easily configured and results are saved to a DOT database so that data integrity over time can be tracked.

III. Challenges and limitations of AI and data science in the vaccine life cycle

Introduction

Integrating AI and data science techniques into vaccine deployment brings ethical and technical considerations that are crucial for the successful deployment of these technologies. While not the only considerations—cultural and policy-related issues should also be considered—these pillars are essential to ensuring the responsible and transparent application of AI and data science tools in vaccine deployment. Understanding the analysis and implementation context is necessary for the successful and responsible implementation of data-driven solutions.

An overarching challenge in applying AI, ML, or data science techniques in a field is the contextually appropriate handling and interpretation of data.[80] While technologists and data scientists may lack domain expertise in the health or immunization sector, it is essential that these technologists collaborate with domain experts to mitigate potential challenges and limitations arising from data-driven applications. Addressing the complexity of ethical considerations necessitates a multidisciplinary approach. The breadth of talents required includes expertise in privacy, security, data science, health, legal, and community engagement, at the very least, in the review process, if not throughout the solution development process.

Ethical considerations

The use of AI and data science techniques in vaccine deployment activities involves processing sensitive health data.[80] This necessitates a robust consideration of ethical concerns to ensure responsible data use. This section will focus on three ethical considerations: data privacy, bias mitigation, and addressing gender inequities. Acknowledging the broader spectrum of important considerations, such as political, regulatory, community, and environmental impact concerns, this report recommends a multidisciplinary approach to tackle the complexity of ethical challenges inherent in the use of AI and data science technologies. This approach should encompass relevant expertise and thorough review processes throughout the development, deployment, and use of such technologies.

Data privacy

Data privacy challenges include the potential for inadequate or insufficient privacy and security protections of personal health data, particularly when dealing with large-scale datasets, such as electronic health records, to inform the development and training of AI and data science tools. A related set of potential challenges or uncertainties to be addressed prior to any analysis project are those of data ownership and consent.[80,81] Ensuring consistent and informed consent is a critical issue, as disparities or lack of uniformity in obtaining permission to use health data can impact the ethical foundation of AI and data science applications, leading to potential breaches of individual privacy and trust. This calls for a sustained community engagement and trust building strategy.

To navigate these challenges successfully, a comprehensive understanding of legal and ethical frameworks is indispensable when handling health-related data. This emphasizes the need for adherence to ethical guidelines to ensure the responsible and secure application of AI in vaccine deployment[81] as well as regular and transparent coordination between technologists and health and immunization experts when designing, developing, and deploying these solutions.

Mitigating biases

Measures to address and mitigate biases are equally essential in the ethical development of AI and data science solutions for vaccine deployment.[81] Risk of bias is a highly likely occurrence and can occur throughout the development stages of these solutions, including the definition of target variables, class labels, and the selection of features and proxies within AI algorithms.[80] Given that the underlying data used to train models may lack bias mitigation considerations, particularly in historical data, it is crucial to address both historical biases and biases in current and future data collection, processing, and utilization.

Automation bias, where human decision-makers unquestioningly follow AI recommendations, similarly risks potentially contributing to biased outcomes in vaccine allocation decisions.[80] The use of LLMs, such as ChatGPT, in healthcare settings has also raised concerns regarding bias and transparency.[80,82] Furthermore, the risks posed by these techniques of generating potentially inaccurate content in healthcare settings can lead to severe consequences, elevating ethical concerns.[82] To counteract bias, developers and providers of AI models should engage in ongoing activities such as error reporting and auditing of deployed models.[81] As with ensuring adherence to data protection standards, the necessity for a collaborative approach between technologists, health experts, and communities affected to reduce biases is crucial for error identification, documentation, and the establishment of ethical safeguards.[81]

Gender equity considerations

The use of AI and data science techniques in vaccine deployment has the potential to integrate perspectives from diverse stakeholders into the development process, enhancing transparency and inclusiveness.[80] However, human decisions throughout the vaccine deployment process and the development of AI technologies to support them can lead to the embedding of biases and inequities in AI and data science systems. While numerous equity issues exist, this report focuses specifically on gender equity.

The gender-related barriers in immunization are multifaceted. Behavioral factors influenced by gender contribute to variations in disease severity and vaccine uptake, leading to higher mortality rates among specific demographic groups. Challenges in collecting and analyzing sex- and gender-disaggregated data on populations further compound these challenges due to a myriad of issues. These include incomplete and non-standardized health records, difficulties in collecting data in remote or humanitarian crises contexts, and limited resources and training on gender-responsive data collection practices.

Equally, AI and data science techniques can inadvertently contribute to these inequalities. Gender biases in AI systems can lead to lower service quality for women, reinforcing stereotypes and erasing marginalized gender identities. This underscores the need for a gender-responsive approach in deploying AI and data science tools throughout the vaccine development and administration process to mitigate and address gender inequalities that impact vaccine deployment equity, coverage, and effectiveness.

To address gender equity concerns, it is crucial to implement gender-responsive strategies throughout vaccine delivery and administration and the development of AI and data science tools to accelerate and enhance deployment. Investments in gender-responsive data practices can help fill data gaps required to build equitable techniques.[83] This includes coordinating with gender experts to assess datasets for under-representation of gender concerns and ensure that tools do not replicate existing gender inequities.[84] Similarly, leveraging tools on participatory design is crucial. For example, providing accessible AI and data literacy training for gender and immunization experts to support the integration of gender expertise into AI systems.[84] Encouraging technology development partners to prioritize gender diversity, advocate for gender-responsive techniques, and establish responsible AI codes and principles with a gender lens is integral.[84] Additionally, incorporating gender dimensions in the testing of AI and data science tools, including developing deployment plans that consider gender-related barriers to vaccine delivery and administration can ensure a holistic approach to gender equity in AI for vaccine deployment.

Technical considerations

Implementation and deployment of data science techniques as outlined in the earlier sections can pose technical challenges for organizations. In addition to the area specific challenges introduced in the respective sections, the following subsections examine cross-cutting, technical considerations regarding the use of AI and data science techniques.

Lack of data standardization

Many of the use cases outlined above require data input from various sources. This can range from different systems within the same organization, such as input from electronic health record platforms, and enterprise resource planning software, to systems that are distributed between various types of organizations and countries. With a growing number of systems and organizations involved, structuring, standardizing, and cleaning this input data becomes increasingly difficult. In the United States, for example, uniquely designed immunization information systems for each state had to be connected individually to one centralized, overarching vaccine tracking system during the COVID-19 pandemic.[85] In an international context, this can be exacerbated by multiple languages, data capturing processes, and varying terminology. The broad adoption of a joint standard, such as the healthcare data standards developed by Health Level Seven International, can address these challenges.[86]

Unstructured data

Especially in the areas of social listening and information campaigns, data science solutions need to work with text data from various sources, making it difficult to separate valuable information from noise. In many vaccine-related settings, information is still captured by hand, which further complicates this challenge.[86,87]

Data security

The administration of vaccines necessitates the collection, recording, and storing of personally identifiable information (PII), and protected health information. Due to the critical nature of this information, technical and procedural guardrails and safety mechanisms should be implemented to protect organizations, individuals, and communities against data security breaches. This becomes even more critical in the context of transferring data to, and saving data on, a centralized platform.[86] It is recommended that organizations handling vaccine-related PII adhere to established standards around data security, such as GDPR.

Missing prerequisites

The feasibility of deploying AI and data science solutions in different contexts depends greatly on the digital readiness of countries, such as the coverage and capacity of communications networks, the extent to which digital solutions are already in use, and the level of digital literacy among healthcare workers and the population.[88] For example, the success of CCE temperature loggers presented in previous sections might depend on the coverage and quality of mobile networks in a country. Further challenges might include language barriers or the lack of trained personnel to monitor autonomous systems. Language barriers can be mitigated, to a degree, through automated translation capabilities of AI-based tools and services. Lack of connectivity or access to technology has proven stubbornly difficult to address, and clearly, advanced technology tools are of little value without the necessary infrastructure. Limited access to technology poses a more fundamental challenge, as activities occurring offline—and hence not digitally recorded—cannot contribute to the training of ML models, nor can the insights generated by these models be applied to influence real-world conditions. For instance, when misinformation circulates through offline channels, the detection capabilities of AI and data science tools are significantly diminished, as they lack access to the undigitized information. Consequently, if an AI model identifies a particular piece of information as being the most widely spread, its analysis may be skewed due to the incomplete inclusion of all relevant data. This scenario underscores the importance of comprehensive data capture in enhancing the accuracy and applicability of AI-driven insights. Furthermore, if the target audience does not use digital services or platforms, the engagement phase is limited to either word of mouth or printed media.

Model quality and transferability

Mathematical models can yield great results in predicting complex systems like the vaccine life cycle by analyzing observed data to detect patterns and trends. However, they often do not account for the underlying physical and biological principles (i.e., the real-world meaning of input variables), and their reliability can falter in new or changing scenarios. This is a challenge in vaccine-related predictions, where emerging conditions and novel variables can play a significant role.

Mechanistic models, conversely, are grounded in the natural laws governing disease transmission. They incorporate factors like population differences and disease progression, making them more dynamic and relevant for understanding vaccine effects in varied conditions. These models can better address the complexities of vaccine response and efficacy, offering insights that statistical models might miss due to their data-dependent nature. In vaccine forecasting and public health planning, a combination of both types of models is often necessary for accurate and effective strategies.[48]

Transferability in AI and data science solutions refers to the capacity of algorithms to acquire and utilize knowledge from one learning experience (for example, patterns around vaccine uptake discovered in one country or region) and apply it effectively to new, unseen situations and tasks. To this end, it is important to establish appropriate objectives for learning transferable knowledge and adapting it to new tasks and domains.[89]

Model explainability

Due to the critical nature of the tasks and potential immediate effects on human health outcomes, explainability is a crucial consideration when it comes to the application of AI in healthcare, such as the administration of vaccines. In this context, explainability refers to AI solutions allowing humans to comprehend and retrace the steps taken to produce a certain outcome, such as a classification or prediction, and recreate these steps if necessary. While this can often be achieved by understanding the underlying mathematical models for traditional ML algorithms, such as regression models, explaining the inner workings of modern approaches, such as artificial neural networks, is a more difficult challenge due to the complexity of the models.[90]

To create transparency about the inner workings of ML algorithms, Google introduced the framework of Model Cards as a dynamic way of documentation. Model cards can contain information about the algorithm or ML approach that is used, its capabilities and limitations, model performance, and known biases.[91]

The explainability of AI models in the health sector is further complicated by differing mental models and sources of truth between software developers and healthcare professionals (data vs. patient and model interpretability vs. clinical plausibility). Thus, the perspectives of both health professionals and AI developers need to be incorporated in the design and development of AI solutions for the health sector and, by extension, the vaccine life cycle.[92]

Model interpretability

Interpretability refers to humans' ability to contextualize the output of an ML algorithm or AI system, apply domain expertise and social values to it, and draw the relevant conclusions from it. Interpretation of rule-based software systems, while labor- and time-intensive, can be a straightforward process, as a user or administrator can simply retrace the steps of the system starting with a given input to verify the output. Due to their mathematical complexity and the often high number of parameters, interpreting the output of ML algorithms is more challenging and requires expertise in mathematics, software engineering, and the given domain in which the system is operating.[93]

Due to the potential health implications of vaccination space, interpreting model outputs correctly and drawing the right conclusions from them are crucial and solutions should be held to high standards accordingly.

Model drift

Model drift describes a decrease in the performance of an ML model over time, which can be caused by either a change in the definition of the variable that is predicted/ modeled (conceptual drift), or a change in the distribution of the features upon which the algorithm built the model (data drift).[94] Model drift can also result from the assumption that past dependencies between variables and the distribution of variables still hold after a given time, while in reality, these factors might have changed. Lastly, if a model is trained in one region or context, but applied in a similar, yet slightly different setting, the results can no longer be expected to be correct.[95]

As an example in the vaccine space, past datasets used for training purposes might state the number of unregistered individuals in a region at a certain number, causing an algorithm to add a buffer to its forecast of vaccine demands. If the number of unregistered individuals changes in response to external events – a natural disaster, say – the model will begin producing inaccurate results without the model being retrained with the adjusted figures.

To detect model drift, the outputs of the model must be continually monitored and validated.[94] Besides retraining the entire model if drift is detected, problematic features that are detectable via detailed evaluation and root cause analysis could be excluded from the input.[95]

Data privacy by design

The privacy aspects in the previous section should be considered in the design, implementation, operation, and maintenance of technical solutions. Many existing guidelines for responsible AI development recommend collecting and handling training data responsibly, computing usage statistics locally on the user's device if possible, and ensuring that ML models themselves, as well as their outputs, adhere to privacy safeguards.[93]

In the collection process, this can for example include evaluating the necessity of a model being trained on sensitive information, minimizing the use of such information where possible, and applying techniques such as anonymization and randomization, including metadata. Calculating usage statistics on the user's device rather than sharing the raw data ensures the relevant information is available for the algorithm to function properly without revealing potentially sensitive user information.[93]

Misinformation risks of generative AI

The possibility of misinformation poses a major challenge when interacting with, or building applications based on, generative AI, such as LLMs. Misinformation is typically the result of two potential shortcomings of LLMs: hallucination and manipulation. Hallucination refers to the risk of an LLM producing seemingly valid, yet factually incorrect content. Due to the typically utilized chat interface that allows back-and-forth communication between the LLM and the user, the model can mislead the user to overestimate its capabilities and trustworthiness, thus manipulating the user to accept - and worse, potentially act on - the false information the model provided. This challenge becomes even more critical considering the potential for ill-intentioned actors leveraging LLMs to spread misinformation and erode public trust.

As with any ML algorithm, LLMs can reflect the biases of the (text) data they were initially trained on.[96] While these biases in the training data can be intentional, they can also result from insufficiency in the data collection process or data processing. Accordingly, if decision makers base their decisions on content created by an LLM, they might inadvertently reflect the biases or shortcomings of the data the LLM was trained on.

Another current limitation of many generic generative AI algorithms, such as ChatGPT, lies in their opaqueness—that they typically do not disclose the sources from which the information they convey has been extracted. This lack of information makes it even more difficult for users to fact-check the content or even retract how the model arrived at a given conclusion or statement. Accordingly, the level of transparency about the information sources can be an important criteria when selecting LLM applications.

The two-way communication between the LLM and the user, together with many companies' privacy policies, can pose significant privacy risks. As of July 2023, the privacy policies of both Google and OpenAI for their generative AI services, BARD and ChatGPT, state that the companies can access and share conversation data for business transfers and legal purposes.[96,97] Given the criticality and private nature of patient information in the vaccine space, the risk of data breaches should be considered when designing, implementing, and operating solutions using LLMs.

Since generative AI ultimately applies mathematical operations rather than (human) creativity to produce content, this content (even when conveying factually correct information) can oversimplify situations and lack human emotions, making it monotonous and less engaging.[96] This behavior could negatively impact the performance of LLM-created content in improving vaccine acceptance or in addressing concerns regarding vaccination.

Adoption of AI and data science solutions

At its best, the implementation of an AI, ML, or data science solution is to support a vaccine operator in meeting their service delivery objectives. However, the successful development and implementation of AI and data science solutions greatly depend on the various people involved in the process. Adoption of these solutions is dependent on far more than technical capability and resolution of ethical questions. It also relies on trust in the solution, solution literacy (including the capacity to receive and respond to training), and solution sustainability. Therefore, the implementation, usage, and continuous improvement of AI and data science tools in the vaccine life cycle should be a collaborative process and stakeholders, ranging from data professionals developing new solutions to health professionals and policymakers, have to be aligned and enabled to contribute to the various components, ensuring all relevant viewpoints are considered.

Trust, acceptance, and expectations of AI and data science tools

As outlined in previous sections, users' trust in any new tool or technology is essential for its acceptance and ultimately for its adoption and widespread usage. The lack of transparency of how AI algorithms operate to transform input variables into outputs can harm trust and users' willingness to use a solution, which holds especially true if models produce wrong results, such as during development or user testing.

The misunderstanding of AI's role and potential can lead to concerns about job displacement in the health workforce. AI, even as it evolves, is unlikely to render healthcare jobs obsolete. Instead, it necessitates a re-engineering of existing roles, where human expertise and AI tools work in tandem, acknowledging the complex and unpredictable nature of medical processes that AI alone cannot fully address. Fears of displacement may lead to resistance in adoption. Supporting health workers to understand how AI tools assist them in their daily work is key to overcoming those concerns.

Furthermore, overestimation of AI's current capabilities can lead to disillusionment among the public if AI does not meet these inflated expectations. This issue underscores the need for greater public dialogue and education about AI's realistic capabilities and limitations in healthcare settings. Such dialogue can help align public expectations with the actual potential and constraints of AI in the vaccine space, facilitating smoother integration and acceptance.[98]

Interconnected with trust in, and acceptance of, the AI and data science tools solution, as well as expectation management is the alignment to appropriate choice of implementation.

Solution literacy

Developing solution literacy involves establishing community data literacy, expanding training and support for implementation workers, and responding and developing response to changing employment environments within the local ecosystem.

Data literacy

With the increasing digitization of medical records and the proliferation of health-related data, community health workers face a vast amount of information. Data literacy is crucial for them to effectively navigate and utilize this information to extract insights, identify patterns, and make more informed decisions. By analyzing data effectively, healthcare professionals can identify best practices, evidence-based treatments, and potential complications. Furthermore, data literacy empowers healthcare professionals to critically evaluate research studies, understand the limitations and biases in datasets, and use statistical data to guide their decision-making. Professionals with data literacy skills can effectively communicate data-driven information to patients, helping them understand treatment options such as a specific vaccine.[99]

Through digital skills, healthcare workers can actively participate in streamlining workflows, automating routine tasks, and optimizing processes, to increase efficiency, save resources, and enable more effective patient care. Especially in low-resource settings with a low ratio of healthcare workers to patients, this can lower the demand on professionals, while ensuring adequate patient care.[100]

Sufficient capacity

Besides the necessary IT infrastructure, financial resources and technical skills are required to design, implement, operate, and maintain any ML and AI solution. The costs are determined by multiple factors, starting with the cost of capturing the data as a foundation for the training and operation of the algorithms.[101] For example, along the vaccine deployment life cycle this could include the opportunity cost of medical staff spending time on data capture instead of treating patients, additional personnel surveying vaccination rates in remote locations, or digital devices to support them in this process.

Furthermore, costs are determined by the research preceding the implementation stages of a project. In this phase, experts from various fields determine the feasibility of the project, for example, through the evaluation of various approaches or tools and the implementation of a prototype solution. When developing a prototype into a production-ready system, additional costs for cloud resources, integration of the solution into the existing infrastructure, support, and maintenance occur.[102]

Besides subject matter experts, varying business and technical skill sets are essential for the success of an AI and ML project. Depending on the size and complexity of the project and the technical solution, this might include a (business) analyst, data engineer, ML engineer, and project manager(s).[102]

Training and support of the health workforce

In advancing the digital literacy of healthcare workers, the spectrum of training content may range from foundational data entry techniques and tool-specific instructions to the fundamentals of AI and data science. While basic training likely suffices in most cases, health workers that closely collaborate with technical partners might benefit from an in-depth understanding to be able to identify AI use cases and advise technologists on domain specific requirements, etc. This training should be complemented by hands-on experience with AI tools, fostering a blend of theoretical understanding and practical application. The cultivation of a learning culture within healthcare organizations is also vital, encouraging continuous skill development and providing easy access to relevant resources. Emphasizing the practical applications of AI in healthcare, the training should demonstrate the tangible benefits these technologies bring to specific roles, addressing any concerns regarding job displacement, ethical considerations, and data privacy.

Motivation is key to encouraging healthcare workers to embrace digital tools. Clear communication about the benefits of AI, including how it simplifies tasks and enhances patient outcomes, is essential. Creating a culture that values learning and innovation, where the adoption of AI is recognized and rewarded, fosters an environment conducive to technological acceptance. Ensuring that AI tools are relevant and useful for their specific tasks enhances individuals' engagement with these technologies. Autonomy in decision-making and problem-solving, facilitated by AI tools, empowers healthcare workers, fostering a sense of ownership and involvement. Acknowledging and celebrating the successful use of AI, for example in the data capturing process during an immunization appointment, boosts morale and encourages continued engagement with these technologies. Lastly, open channels for feedback and concerns about AI usage should be established, allowing for continual adaptation and improvement of AI implementation strategies.[100]

Responding to impacts on employment scope

A study by McKinsey & Company concluded that the introduction of AI in healthcare is expected to change the nature of health work significantly and that AI's role in automating routine, administrative tasks in the health sector can free up to 70% of a healthcare practitioner's time (based on analyses in Europe), allowing them to focus more on patient care, should this not lead to staff reductions at the same time. The integration of AI and the need to embed digital and AI skills within healthcare organizations can shift the focus of healthcare education away from memorizing facts to fostering innovation, entrepreneurship, continuous learning, and multidisciplinary collaboration.[103]

Ensuring solution sustainability

To ensure that ML models are applied and operated sustainably, social, economic, and environmental dimensions should be considered. Social sustainability depends on the real-life value and benefits an application can provide, measured by the improvements of a given situation that is being addressed.[104] In the context of vaccines, this could, for example, be measured by the development of stock-out rates of a specific vaccine before and after ML-based demand forecasting is applied, or evaluating the development of vaccine coverage once a solution has been implemented.

In the initial conceptual stages of a project, the formulation of project statements and a comprehensive measurement plan is key. This approach enables stakeholders and implementing organizations to methodically align technical solutions with tangible real-world outcomes, thereby facilitating positive change in a specific domain or for a particular use case. In the case of an AI and data science project, the DataKind Playbook recommends the following template to form a project statement that will determine the impact map and success metrics.

We want to **(do this kind of analysis)**

Using **(this kind of data)**

So that **(we respond to this organizational need)**

So that **(we see this change in the world)**

Figure 4: DataKind Playbook - template for project statement[105]

Methodologies, such as the Logical Framework Approach, can be employed to delineate the components of a project and its activities, establishing their interrelationships. This framework aids in identifying the metrics through which the project's anticipated results can be systematically monitored, which lays a solid foundation for ongoing evaluation of the project's or solution's impacts.[106]

The United Nations Peace and Development Sub-Fund is similarly based on a results-based management strategy.[107] It defines evaluations as a systematical assessment of the results chain, processes, and contextual factors of a project based on the following criteria:

+ Relevance: Measures the extent to which both the objectives and the design of a solution respond to the desired beneficiaries' needs.
+ Effectiveness: Analyzes the progress towards objectives along the results chain.
+ Efficiency: Measures how well the solution transforms available resources into the intended outputs.
+ Impact: Addresses the ultimate significance and potentially transformative social, environmental, and economic effects of the solution, as well as potential consequences of the intervention.
+ Sustainability: Assesses how the net benefits of a solution can be maintained from a financial, economic, social, and environmental perspective.

By considering these factors in each stage of a project, the intended and unintended results, as well as their implications can be understood and managed.

Economic sustainability is largely determined by the operating cost of a solution in comparison to the observed economic benefits.[104] Regarding vaccines, this could be calculated as the increased productivity of a community due to decreased medical leave against the cost of necessary hardware infrastructure and personnel to operate a model.

The high energy consumption of training large, complex ML models can have a detrimental impact on the environment and should be considered when determining the overall net benefit of a solution. The application of more energy-efficient hardware, such as data centers and graphics processing units, as well as optimizing models to require fewer computing resources are potential strategies to ensure environmental sustainability.

IV. Recommendations and conclusion

AI and data science techniques can be applied across all stages of the vaccine deployment life cycle, from simulation-based demand forecasting of immunization materials to real-time route optimization and understanding of concerns around vaccination through social listening. With this wide range of potential applications, the specific focus areas for potential projects should not be determined by technological feasibility but by understanding the breadth of available solutions and prioritizing the most crucial needs of a particular target group, in line with existing national and international strategies, standards, and guidance. Potentially competing interests of different groups may have to be weighed and considered in the conception of new projects. While administrators and analysts might, for example, be interested in high-quality, real-time data regarding the immunization status of a targeted community, this might put an additional workload, without immediate benefit, on the already busy community health worker tasked with this data collection.

The cross-cutting issues of data quality and health workforce management emerged as the dominant themes in both the literature research and stakeholder survey undertaken for this landscape report. There are significant opportunities to improve data quality with flexible and accessible solutions like DOT that can be configured-for-purpose.

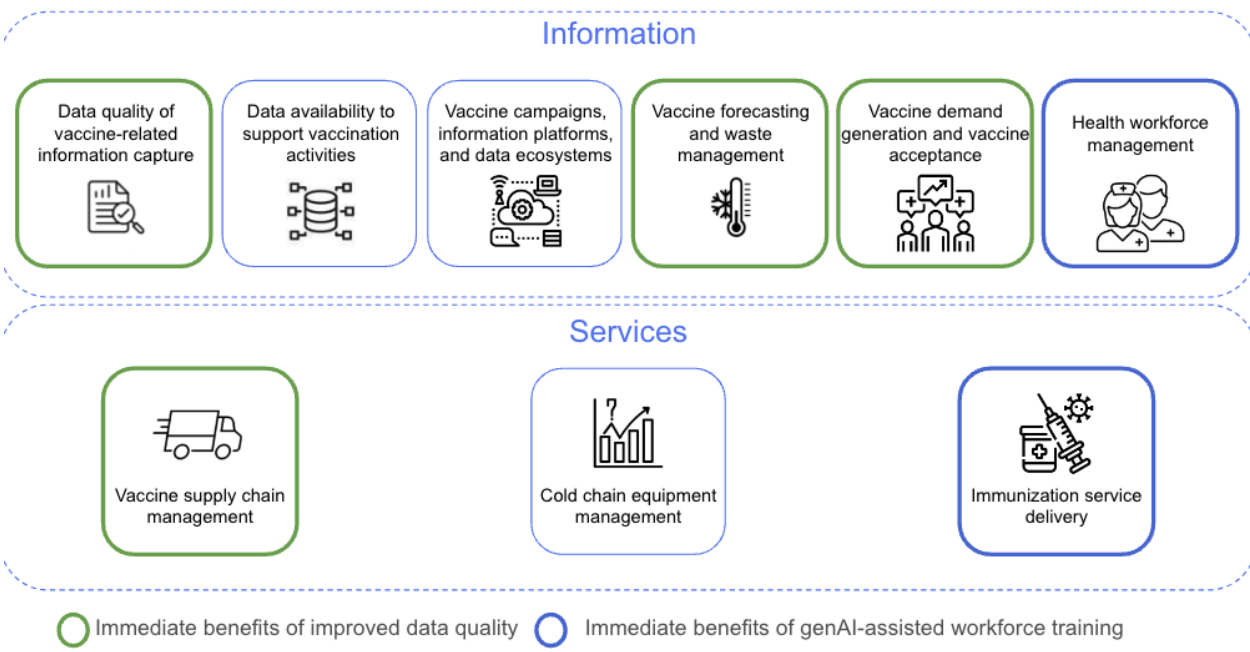

Figure 5: Areas affected by improved data quality and more efficient workforce training

To simultaneously increase the availability and the quality of data and support health workforce management, a multi-dimensional approach is recommended:

+ Improve the data capturing process: Preventing quality issues is most effective and efficient at the data collection stage. In doing so, organizations can prevent data quality issues spreading across their systems and to other organizations' systems, reducing the workload for administrators and IT professionals who would otherwise have to correct errors at multiple points.

+ Increase the data literacy of health workers: Besides using custom-built forms in data collection applications to guide staff through the data capturing process, increasing data literacy of health workers can improve both the quantity and quality of the data they collect. In-person or online workshops can be supplemented with generative AI-based applications, such as conversation-based knowledge retrieval from training resources. Furthermore, LLM chat interfaces could support healthcare professionals in the operational aspects of their work by providing access to information from training resources on a range of topics, such as vaccine administration.
+ Provide flexible tools for data quality control: Even with improvements in the data capturing process, data quality issues can occur due to out-of-control factors, such as compromised sensors or data that has been provided by third parties. Dedicated quality assurance tools, such as DOT can be deployed to detect these cases.

Initiatives on AI and data science should be a joint effort between subject matter experts, such as community health workers, policymakers, communities, technical contributors, and funders, to ensure the right challenges are addressed with the appropriate solutions. A leading example of this approach is the initiative undertaken by Jacaranda Health, who developed and launched an ML algorithm designed to discern the intent behind inquiries made by expecting and new mothers on their SMS-based digital health platform, PROMPTS. The challenges of working in low resource settings, such as data scarcity and the underrepresentation of users' native languages in many pre-existing solutions, was addressed by evaluating and comparing multiple models not only related to their output quality, but also with regards to cost and data efficiency, to determine the suitability for the given use case.[108]

In this context, an appropriate solution could be technologically less advanced and yet lead to better real-world outcomes due to higher user acceptance, more local expertise for maintenance and user support, or better interoperability with existing systems. Given the process dependencies and overlap of different stakeholders and organizations, the potential of any software solution must be evaluated in light of the overall ecosystem of processes and IT infrastructure. Although individual solutions might deliver the expected results, low interoperability with other systems might lead to an overall negative impact. From a technical perspective, fragmented IT landscapes on different administrative levels and missing standards regarding information storage and transfer pose significant challenges regarding the establishment of a transparent and efficient infrastructure. Furthermore, agreeing on and adhering to privacy and security protocols can strengthen data privacy and ensure the protection of sensitive patient information. Accordingly, advocating for the use of data standards for digital health records and incentivizing data sharing (within the boundaries of fair competition) between stakeholders should be a priority.

Funders hold significant influence in harnessing the potential of AI and data science for vaccine deployment. To facilitate this, they can take several proactive steps. Firstly, funders can evaluate their current health and immunization portfolios to pinpoint areas where AI and data science could add immediate value. Based on this analysis, they could engage stakeholders to adapt ongoing projects accordingly or develop funding opportunities addressing these needs.

Additionally, funders can allocate resources to support research and analysis, such as landscape assessments, to identify use cases and deepen understanding of tool capabilities and limitations.

The scalability of solutions is crucial for demonstrating their value and impact. When proof-of-concept AI and data science solutions demonstrate promise, funders should prioritize those with high demand, impact, and the potential for scaling into production-level tools to facilitate their integration into real-world vaccine deployment scenarios. Funders can also play a pivotal role in fostering partnerships across public, private, and philanthropic sectors, fostering collaborative efforts to maximize the utility of AI in vaccine deployment.

For funders supporting initiatives in LMICs or low-resource settings, targeted investments should be considered to bolster local capacities for implementing AI and data science techniques effectively. By strategically investing in these areas, funders can ensure the widespread adoption and sustainable use of AI-driven solutions, thereby enhancing global vaccine deployment efforts.

With the constant and rapid technological advancement in the fields of AI and data science, any project or initiative in this area should be embedded in a continuous learning process of understanding the capabilities, risks, and challenges of new technologies, defining mitigation strategies, and carefully selecting appropriate use cases.

To ensure the long-term sustainability of any AI and data science application, it is paramount to empower local

Insights from the Wellcome Trust: Supporting the integration of AI and data science in vaccine deployment. May 2024.

The Wellcome Trust (Wellcome) supports science to solve urgent health challenges. We fund discovery research into life, health, and wellbeing, and we're taking on three worldwide health challenges: mental health, infectious disease, and climate and health.

These are big, complex global health challenges, and we know Wellcome won't solve them alone. Making progress will require building understanding and cooperation among different actors in the global health ecosystem – which includes philanthropic funders like Wellcome, but also governments, policymakers, researchers, and industry.

This report aims to get those various stakeholders on the same page when it comes to the opportunities, limitations, and ethical considerations for using AI for vaccine deployment. DataKind has mapped these out along the full vaccine deployment life cycle – which is critical for understanding where these tools are currently used, what the gaps are, and where there are solutions that might need to be prioritized.

Throughout the report, it's clear that investment – both public and private – is needed to demonstrate value and encourage use of AI in the most promising areas. The report gives helpful recommendations for where energy and funding should be directed in order to address current blockers around data collection and quality, and data literacy for health workers.

There is potential for AI and data science to improve vaccine coverage and reduce inequity in access. But – as the report explains – that's by no means inevitable. So alongside improving the underlying data and developing skills in the workforce, action is required to ensure these tools are designed, built, and governed in trustworthy ways. From ensuring training data is representative of the population AI tools are being developed for, to ensuring effective public voice in policy and regulation – many of the solutions go beyond vaccine deployment and speak to fundamental issues about the future of AI in health.

stakeholders on all levels through education, training, and ownership of the project from conception to implementation and operation, taking a human-centered design approach to creation, implementation, and sustainability of the solution.

References

1. The United Nations General Assembly. (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Resolution A/RES/70/1. New York. Retrieved February 19, 2023, from https://sdgs.un.org/2030agenda

2. World Health Organization. (2019, October 29). *Vaccines and Immunization*. Retrieved November 8, 2023, from https://www.who.int/health-topics/vaccines-and-immunization#tab=tab_1

3. Patel, M., Dumolard, L., Nedelec, Y., Sodha, S. V., Steulet, C., Gacic-Dobo, M., Kretsinger, K., McFarland, J., Rota, P. A., & Goodson, J. L. (2019, December 6). Progress Toward Regional Measles Elimination — Worldwide, 2000–2018. *Morbidity and Mortality Weekly Report*. https://doi.org/10.15585/mmwr.mm6848a1

4. Gavi, the Vaccine Alliance. (2023, October). *Facts and figures*. https://www.gavi.org/programmes-impact/our-impact/facts-and-figures

5. UNICEF. (2022, July 14). C*OVID-19 pandemic fuels largest continued backslide in vaccinations in three decades*. Retrieved November 8, 2023, from https://www.unicef.org/press-releases/WUENIC2022release

6. UNICEF. (2023, July 17). *Vaccination and Immunization Statistic*s - *UNICEF Data*. Retrieved November 8, 2023, from https://data.unicef.org/topic/child-health/immunization/

7. Gavi, the Vaccine Alliance. (2023, January 19). *The US$ 500 million challenge: Gavi launches INFUSE 2.0 to scale up innovations in immunisation*. Retrieved November 8, 2023, from https://www.gavi.org/news/media-room/us-500-million-challenge-gavi-launches-infuse-20-scale-innovations-immunisation

8. Chiabarra, C. (2023, September 7). *Healthcare Workers Deserve Better: Can Artificial Intelligence Help?* Forbes. Retrieved January 8, 2024, from https://www.forbes.com/sites/forbestechcouncil/2023/09/07/healthcare-workers-deserve-better-can-artificial-intelligence-help/?sh=6a614f78704e

9. Welch, A. (2023, September 13). *Artificial intelligence is helping revolutionize healthcare as we know it*. Johnson & Johnson. Retrieved January 8, 2024, from https://www.jnj.com/innovation/artificial-intelligence-in-healthcare?&utm_source=google&utm_medium=cpc&utm_campaign=GO-USA-ENG-PS-Corporate+Equity-GP-EX-RN-NB_STORIES_ARTIFICIAL+INTELLIGENCE&utm_content=AI+-+Healthcare&utm_term=use+of+ai+in+health+care&gcl

10. Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021, July). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Health Journal, 8*(2), 188 - 194. https://doi.org/10.7861/fhj.2021-0095

11. Parums, D. V. (2023, June 1). Editorial: Infectious Disease Surveillance Using Artificial Intelligence (AI) and its Role in Epidemic and Pandemic Preparedness. *Medical Science Monitor*. https://doi.org/10.12659/MSM.941209

12. Kumar, A., Mani, V., Jain, V., Gupta, H., & Venkatesh, V. (2023, January 1). Managing healthcare supply chain through artificial intelligence (AI): A study of critical success factors. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2022.108815

13. Wellcome Trust. (2023, June 26). *Unlocking the potential of AI in drug discovery*. Retrieved January 22, 2024, from https://cms.wellcome.org/sites/default/files/2023-06/unlocking-the-potential-of-AI-in-drug-discovery_report.pdf

14. Kit-Kay, M., & Pichika, M. R. (2019, March). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today,* 24(3), 773-780. https://doi.org/10.1016/j.drudis.2018.11.014

15. McCarthy, J. (2007, November 12). *What is artificial intelligence?* Retrieved February 19, 2024, from https://www-formal.stanford.edu/jmc/whatisai.pdf

16. Gavi, the Vaccine Alliance. (2016, May). *Strengthening the Immunisation Supply Chain*. Retrieved November 10, 2023, from https://www.gavi.org/sites/default/files/publications/GAVI%20Supply%20Chain%20Strategy%20FS.pdf

17. World Health Organization. (n.d.). *Immunization Handbook. Unit 4: Cold chain and logistics management*. Retrieved November 10, 2023, from https://www.who.int/docs/default-source/searo/india/publications/immunization-handbook-107-198-part2.pdf

18. World Health Organization, Regional Office for the Eastern Mediterranean. (n.d.). *Health workforce: management systems*. Retrieved November 10, 2023, from https://www.emro.who.int/health-workforce/management/index.html#:~:text=Health%20workforce%20management%20systems%20also,their%20responsibilities%20efficiently%20and%20effectively

19. World Health Organization. (n.d.). *Immunization campaigns*. Retrieved November 10, 2023, from https://www.who.int/teams/immunization-vaccines-and-biologicals/essential-programme-on-immunization/implementation/immunization-campaigns

20. National Institute of Standards and Technology. (n.d.). *Computer Security Resource Center: Glossary*. Retrieved November 10, 2023, from https://csrc.nist.gov/glossary/term/availability

21. Bloland, P., & MacNeil, A. (2019, April 4). Defining & assessing the quality, usability, and utilization of immunization data. BMC Public Health, 19(1). https://doi.org/10.1186/s12889-019-6709-1

22. World Health Organization (WHO). (2022, October 19). *Shortage of cholera vaccines leads to temporary suspension of two-dose strategy, as cases rise worldwide*. World Health Organization (WHO). Retrieved January 22, 2024, from https://www.who.int/news/item/19-10-2022-shortage-of-cholera-vaccines-leads-to-temporary-suspension-of-two-dose-strategy--as-cases-rise-worldwide

23. Mangan, J., Lalwani, C., & Calatayud, A. (2020, December 7). Global Logistics and Supply Chain Management. John Wiley & Sons. http://books.google.ie/books?id=32UHEAAAQBAJ&pg=PR2&dq=978-1-119-70299-3&hl=&cd=1&source=gbs_api

24. Kennedy, A. (2022, October 10). *The 7 Rights of Logistics*. AK Logistics and Supply Chain. Retrieved September 25, 2023, from https://aklogisticsandsupplychain.com/2022/10/10/the-7-rights-of-logistics/

25. Sun, X., Andoh, E. A., & Yu, H. (2021). A simulation-based analysis for effective distribution of COVID-19 vaccines: A case study in Norway. *Transportation Research Interdisciplinary Perspectives*, *11*, 100453. https://doi.org/10.1016/j.trip.2021.100453

26. Chu, C. Y., Park, K., & Kremer, G. E. (2020, August). A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. *Advanced Engineering Informatics*, *45*. https://doi.org/10.1016/j.aei.2020.101053

27. Sarkar, B. (2023, December 22). *AI for Avoiding Supply Chain Disruptions – Two Use Cases*. Retrieved January 6, 2024, from https://emerj.com/ai-sector-overviews/ai-for-avoiding-supply-chain-disruptions-two-use-cases/#:~:text=Logistics%20and%20supply%20chain%20leaders,sizes%20or%20diversifying%20supplier%20mix

28. Castro, N. (2022, May 7). *American Airlines Delivers the Goods, with Data Science Workstations*. Retrieved January 6, 2024, from https://blogs.nvidia.com/blog/american-airlines-data-science-workstations/

29. Centers for Disease Control and Prevention (CDC). (2023, January). *Vaccine Storage and Handling Toolkit*. Retrieved January 6, 2024, from https://www.cdc.gov/vaccines/hcp/admin/storage/toolkit/storage-handling-toolkit.pdf

30. Schlepers, R., Potter, L., & Kamp, L. V. (2017, May 16). *Project Report — Smart Time Temperature Indicator*. Medecins Sans Frontieres. Retrieved November 6, 2023, from https://static1.squarespace.com/static/5fe75ec29a67a4363ac1bbed/t/5fe9d40cd85eb525305ddc65/1609159739670/MSF-Project-Report-Cold-Chain-Indicator.pdf

31. World Health Organization (WHO). (n.d.) *E006: Temperature monitoring devices*. Retrieved January 24, 2024, from https://apps.who.int/immunization_standards/vaccine_quality/pqs_catalogue/categorypage.aspx?id_cat=35

32. UNICEF. (2020, October 23). *General procurement guideline for cold chain equipment*. Retrieved January 24, 2024, from https://www.unicef.org/supply/media/4311/file/General-Procurement-Guidelines-Cold-Chain.pdf

33. Nelson C. M., Wibisono, H., Purwanto, H., Mansyur, I., Moniaga, V., Widjaya, A. (2004, February 1). Hepatitis B vaccine freezing in the Indonesian cold chain: evidence and solutions. *PubMed Central (PMC)*. https://ncbi.nlm.nih.gov/pmc/articles/PMC2585906/

34. Techathawat, S., Varinsathien, P., Rasdjarmrearnsook, A., & Tharmaphornpilas, P. (2007). Exposure to heat and freezing in the vaccine cold chain in Thailand. *Vaccine,* 25(7), 1328–1333. https://doi.org/10.1016/j.vaccine.2006.09.092

35. World Health Organization (WHO). (2021, September 21). *Cold chain equipment and dry store temperature mapping tool*. Retrieved November 8, 2023, from https://www.who.int/publications/m/item/cold-chain-equipment-and-dry-store-temperature-mapping-tool

36. Chandra, A. A., & Lee, S. R. (2014). A Method of WSN and Sensor Cloud system to Monitor Cold Chain Logistics as Part of the IoT Technology. *International Journal of Multimedia and Ubiquitous Engineering*, *9*(10), 145-152. https://gvpress.com/journals/IJMUE/vol9_no10/15.pdf

37. Wood, L. H., Atkinson, W. L., Tan, L.J., & Wexler, D. L. (2023, June 16). *Vaccinating Adults: A Step-by-Step Guide*. Retrieved November 17, 2023, from https://www.immunize.org/about/history/pub-archives/adult-guide/

38. Carvalho, T. P., Soares, F., Vita, R., Francisco, R., Basto, J., & Alcala, S. (2019, November). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering,* 137. https://doi.org/10.1016/j.cie.2019.106024

39. Velraj, V. (2020, March 2). *What Artificial Intelligence means to Cold Chains*. Retrieved November 8, 2023, from https://www.linkedin.com/pulse/what-artificial-intelligence-means-cold-chains-vishnu-vardhan-velraj

40. Rajgopal, J., Connor, D. L., Assi, T. M., Norman, B. A., Chen, S. I., Bailey, R. R., Long, A. R., Wateska, A. R., Bacon, K. M., Brown, S. T., Burke, D. S., & Lee, B. Y. (2011, July). The optimal number of routine vaccines to order at health clinics in low or middle income countries. *Vaccine, 29*(33), 5512–5518. https://doi.org/10.1016/j.vaccine.2011.05.044

41. Immunization Agenda 2030. (2022). *A global strategy to leave no one behind*. Retrieved January 24, 2024, from https://www.immunizationagenda2030.org/images/documents/VisionStratagiy/BLS20116_IA_Visual-ID-DesignLayout_spread_009_WEB.pdf

42. Lazarus, J. V., Abdool Karim, S. S., van Selm, L., Doran, J., Batista, C., Ben Amor, Y., Hellard, M., Kim, B., Kopka, C. J., & Yadav, P. (2022, April). COVID-19 vaccine wastage in the midst of vaccine inequity: causes, types and practical steps. BMJ Global Health, 7(4), e009010. https://doi.org/10.1136/bmjgh-2022-009010

43. Mvundura, M., Ng, J., Reynolds, K., Theng Ng, Y., Bawa, J., Bambo, M., Bonsu, G., Payne, J., Chua, J., Guerette, J., Odei Antwi-Agyei, K., Ribe, S., Chinavane, D., Arhin-Wiredu, K., Shah, A., Sitoe, J., Yunus, S., Powelson, J., Amponsa-Achiano, K., Robertson, J. (2023, June). Vaccine wastage in Ghana, Mozambique, and Pakistan: An assessment of wastage rates for four vaccines and the context, causes, drivers, and knowledge, attitudes and practices for vaccine wastage. *Vaccine, 41*(28), 4158–4169. https://doi.org/10.1016/j.vaccine.2023.05.033

44. Vaccine Impact Modelling Consortium. (n.d.). *Vaccine Impact Modelling Consortium*. https://www.vaccineimpact.org/. Retrieved January 24, 2024, from https://www.vaccineimpact.org/

45. World Health Organization. (2023, March 8). *How to use the WHO vaccination session planner & vaccine wastage calculator - Instruction guide*. World Health Organization (WHO). Retrieved December 14, 2023, from https://cdn.who.int/media/docs/default-source/immunization/supply-chain/session-planning-tool-guide.pdf?sfvrsn=efd7a124_2

46. Mueller, L., Haidari, L. A., Wateska, A. R., Phillips, R. J., Schmitz, M. M., Connor, D. L., Norman, B. A., Brown, S. T., Welling, J. S., & Lee, B. Y. (2016, July 12). The impact of implementing a demand forecasting system into a low-income country's supply chain. *Vaccine*, *34*(32), 3663-3669. https://doi.org/10.1016/j.vaccine.2016.05.027

47. Berry, T., Blakeley, P., Bradshaw, A., Li, H. L., Seidman, G., Sfeile, E., & Triki, C. (2023, May). *A Vision for Global Health: How Demand Forecasting, Supply Planning and Market Shaping Can Deliver a New Model for Global Vaccine Manufacturing*. Retrieved November 14, 2023, from https://assets.ctfassets.net/75ila1cntaeh/5JYzbOPJMQbZ3wZmquUB7A/0538bc7e3e75c8431c625119e0262273/Tony_Blair_Institute_GHSC_One_Shot_Vaccine_Manufacturing_May_2023.pdf

48. Bluedot. (2022). *Revealing the hidden treat: Three innovative strategies for assessing disease burden today, tomorrow, and beyond.* https://4002881.fs1.hubspotusercontent-na1.net/hubfs/4002881/2023/Hidden%20Threat%20-%20Underdiagnoses%20and%20Underreporting%20-%20BlueDot%20eBook.pdf?__hstc=4570412.6c0faf5d704ea8401851e440152413af.1701275734906.1701275734906.1701275734906.1&__hssc=45704

49. Allam, Z. (2020). The Rise of Machine Intelligence in the COVID-19 Pandemic and Its Impact on Health Policy. *Surveying the Covid-19 Pandemic and Its Implications,* 89–96. https://doi.org/10.1016/b978-0-12-824313-8.00006-1

50. Hariharan, R., Sundberg, J., Gallino, G., Schmidt, A., Arenth, D., Sra, S., & Fels, B. (2020, October 30). An Interpretable Predictive Model of Vaccine Utilization for Tanzania. *frontiers in Artificial Intelligence*, *3*(559617). https://doi.org/10.3389/frai.2020.559617

51. Yale Institute of Global Health & UNICEF (2020, December). *Vaccine Messaging Guide*. Retrieved January 8, 2024, from https://www.unicef.org/media/93661/file/Vaccinemessagingguide.pdf

52. Thomson, A., & Finnegan, G. (2020). *Vaccine Misinformation Management Field Guide*. https://vaccinemisinformation.guide/. https://vaccinemisinformation.guide/

53. Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., Ali, A., & Sheikh, A. (2021, April 5). Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. Journal of Medical Internet Research, 23(4), e26627. https://doi.org/10.2196/26627

54. Zarei, M. R., Christensen, M., Everts, S., & Komeili, M. (2023, April 13). *Vax-Culture: A Dataset for Studying Vaccine Discourse on Twitter*. Retrieved November 8, 2023, from https://arxiv.org/abs/2304.06858

55. Kuhn, J. (2023, June 16). *AI And Large Language Models: The Future Of Healthcare Data Interoperability*. Retrieved November 20, 2023, from https://www.forbes.com/sites/forbestechcouncil/2023/06/20/ai-and-large-language-models-the-future-of-healthcare-data-interoperability/?sh=416c9f8d67dc

56. Lee, K. Y., Dabak, S. V., Kong, V. H., Park, M., Kwok, S. L. L., Silzle, M., Rachatan, C., Cook, A., Passanante, A., Pertwee, E., Wu, Z., Elkin, J. A., Larson, H. J., Lau, E. H. Y., Leung, K., Wu, J. T., & Lin, L. (2023, May 25). Effectiveness of chatbots on COVID vaccine confidence and acceptance in Thailand, Hong Kong, and Singapore. *Npj Digital Medicine, 6*(1). https://doi.org/10.1038/s41746-023-00843-6

57. Karabacak, M., & Margetis, K. (2023, May 21). Embracing Large Language Models for Medical Applications: Opportunities and Challenges. *Cureus*. https://doi.org/10.7759/cureus.39305

58. McKee, B. (2019, December 11). *Leading The Healthcare Workforce With Data Science*. Retrieved November 14, 2023, from https://www.healthitoutcomes.com/doc/leading-the-healthcare-workforce-with-data-science-0001

59. Rogach, Y., Darden, A. T., & Robison, C. (2022, November 10). *Is the future of healthcare workforce optimization in data science?* Oracle Blogs. Retrieved November 14, 2023, from https://blogs.oracle.com/cloud-infrastructure/post/healthcare-workforce-optimization-data-science

60. Utazi, E., Thorley, J., Alegana, V. A., Ferrari, M. J., Takahashi, S., Metcalf, J. E., Lessler, J., Cutts, F. T., & Tatum, A. J. (2019). Mapping vaccination coverage to explore the

effects of delivery mechanisms and inform vaccination strategies. *Nature Communications*, *10*(NA), NA. https://www.nature.com/articles/s41467-019-09611-1

61. Borenstein, J., & Weintraub, R. (2020, December 21). *Rolling Out the Covid Vaccine Is a Huge IT Challenge*. Harvard Business Review. Retrieved November 17, 2023, from https://hbr.org/2020/12/rolling-out-the-covid-vaccine-is-a-huge-it-challenge

62. Jalloh, M. F., Namageyo-Funa, A., Gleason, B., Wallace, A. S., Friedman, M., Sesay, T., cansey, D., Feldstein, L., Conklin, L., Hersey, S., Singh, T., & Kaiser, R. (2020, September 3). Assessment of VaxTrac electronic immunization registry in an urban district in Sierra Leone: Implications for data quality, defaulter tracking, and policy. *Vaccine, 38*(39), 6103 - 6111. https://doi.org/10.1016/j.vaccine.2020.07.031

63. Nshunju, R. J., Ezekiel, M., Njau, P., & Ulomi, I. (2018, July). *Assessing the effectiveness of a web-based vaccine information management system on immunization-related data functions - An implementation research study in Tanzania*. Retrieved January 24, 2024, from https://publications.jsi.com/JSIInternet/Inc/Common/_download_pub.cfm?id=21683&lid=3

64. Centers for Disease Control and Prevention (CDC). (2023, February 21). *North Star Architecture Technologies*. Retrieved November 15, 2023, from https://www.cdc.gov/surveillance/data-modernization/technologies/north-star-architecture.html

65. Murray, M., Britten, B., Zitomer, J., Millspaugh, A. M., Paseltiner, D., Mader, B., & Kathol, S. (2022, November). *A Prototype of Modernized Public Health Infrastructure for All: Findings from a Virginia Pilot*. Retrieved November 15, 2023, from https://github.com/CDCgov/phdi/blob/main/publications/DMI_VAWhitePaper_V3.pdf

66. Obi-Eyisi, N. (2023, August 31). *LLM is a Game Changer in Data Engineering*. Retrieved January 09, 2024, from https://afroinfotech.medium.com/llm-is-a-game-changer-in-data-engineering-6745904057b2

67. Mishra, O. (2023, August 7). *Using langchain for Question Answering on Own Data*. Retrieved January 9, 2024, from https://medium.com/@onkarmishra/using-langchain-for-question-answering-on-own-data-3af0a82789ed

68. Radonjic-Simic, M., Mahrt, C., Niemand, S., Speck, A., & Windrich, M. (2022, December 31). Decentralized Open Platform for Vaccination—A German Example: COVID-19-Vacc. *Journal of Open Innovation: Technology, Market, and Complexity*, *186*(7). https://doi.org/10.3390/joitmc7030186

69. Winn, Z. (2020, June 25). *Improving global health equity by helping clinics do more with less*. Retrieved February 19, 2024, from https://news.mit.edu/2020/macro-eyes-vaccine-chain-health-equity-0626

70. Liu, G., Carter, B., & Gifford, D. K. (2020, November 26). Predicted Cellular Immunity Population Coverage Gaps for SARS-CoV-2 Subunit Vaccines and their Augmentation by Compact Peptide Sets. *Cell systems*, *12*(1), 102-107.e4. https://doi.org/10.1016/j.cels.2020.11.010

71. Arora, G., L., Joshi, J., Mandal, R. S., Shrivastava, N., Virmani, R., & Sethi, T. (2021, August 18). Artificial Intelligence in Surveillance, Diagnosis, Drug Discovery and Vaccine Development against COVID-19. *Pathogens*, *10*(8), 1048. https://doi.org/10.3390/pathogens10081048

72. Hall, S., Kaplow, L., Sun, Y. S., & Holt, T. Z. (2021, April 23). *COVID-19 vaccine rollout in LMIC*. Retrieved November 17, 2023, from https://www.mckinsey.com/industries/healthcare/our-insights/none-are-safe-until-all-are-safe-covid-19-vaccine-rollout-in-low-and-middle-income-countries

73. World Health Organization. (n.d.). *The Global Health Observatory*. Retrieved December 21, 2023, from https://www.who.int/data/gho

74. Global Health Data Exchange. (n.d.). *GHDx*. Retrieved December 21, 2023, from https://ghdx.healthdata.org/search/site/vaccines

75. PATH BID Initiative. (n.d.). *Recommendations and lessons learned - BID Initiative briefs*. bidinitiative.org. Retrieved January 24, 2024, from https://bidinitiative.org/wp-content/uploads/VAD_BID_LessonsLearned_Folder_v1_rev03.pdf

76. DataKind. (n.d.). *Get Back on the Road: How Riders for Health Intends to Use Computer Vision to Digitize Health Forms*. Retrieved January 8, 2024, from https://www.datakind.org/2020/10/27/get-back-on-the-road-how-riders-for-health-intends-to-use-computer-vision-to-digitize-health-forms/

77. Bigelow, S. J. (n.d.). *What is edge computing? Everything you need to know*. Retrieved January 22, 2024, from https://www.techtarget.com/searchdatacenter/definition/edge-computing

78. Ayyangar, M., Harris, M., Augustin, C., & Wang, I. (2020). *Building Confidence in Community Health Systems: DataKind's Data Observation Toolkit (DOT)*. Retrieved January 8, 2024, from https://www.datakind.org/2022/11/17/building-confidence-in-community-health-systems-datakinds-data-observation-toolkit-dot/

79. DataKind. (2023, March 21). *Data Observation Toolkit*. Retrieved January 8, 2024, from https://github.com/datakind/Data-Observation-Toolkit

80. Wu, S., Reis, A. A., Pujari, S., Muneene, D., & Reeder, J. (2021, November 27). Leveraging artificial intelligence to assist the ethical and science-based distribution of COVID-19 vaccines. *Journal of Global Health*, *11*. https://doi.org/10.7189/jogh.11.03124

81. Harrer, S. (2023, April). Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*, *90*, 104512. https://doi.org/10.1016/j.ebiom.2023.104512

82. Sallam, M. (2023, March 19). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, *11*(6), 887. https://doi.org/10.3390/healthcare11060887

83. Heidari, S., Durrheim, D. N., Faden, R., Kochhar, S., MacDonald, N., Olayinka, F., & Goodman, T. S. (2021, August). Time for action: towards an intersectional gender approach to COVID-19 vaccine development and deployment that leaves no one behind. *BMJ Global Health*, *6*(8), e006854. https://doi.org/10.1136/bmjgh-2021-006854

84. Smith, & Rustagi. (2021, March 31). *When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity (SSIR)*. Stanford Social Innovation Review. Retrieved November 21, 2023, from https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity

85. Agrawal, G., Azimi, T., Heller, J., Kumar, P., Mysore, M., Patel, P., Sabow, A., Singhal, S., & Truesdale, J. (2021, January 26). *COVID-19 vaccine-deployment risks*. Retrieved November 27, 2023, from https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-risks-and-challenges-of-the-global-covid-19-vaccine-rollout

86. Krylov, A. (2023, March 23). *Data Analytics in Healthcare: Challenges and Uses*. Retrieved November 27, 2023, from https://kodjin.com/blog/data-analytics-in-healthcare-challenges-and-solutions/

87. Awrahman, B. J., Aziz Fatah, C., & Hamaamin, M. Y. (2022, September 29). A Review of the Role and Challenges of Big Data in Healthcare Informatics and Analytics. *Computational Intelligence and Neuroscience, 2022,* 1–10. https://doi.org/10.1155/2022/5317760

88. Guermazi, B. (2021, January 13). *Digital technologies and vaccine deployment: Opportunities and challenges*. Retrieved November 27, 2023, from https://blogs.worldbank.org/digital-development/digital-technologies-and-vaccine-deployment-opportunities-and-challenges

89. Jiang, J., Shu, Y., Wang, J., & Long, M. (2021, September 22). *Transferability in Deep Learning: A Survey*. Retrieved November 29, 2023, from https://arxiv.org/abs/2201.05867v1

90. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020, November 30). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1). https://doi.org/10.1186/s12911-020-01332-6

91. Google. (n.d.). *The value of a shared understanding of AI models*. Retrieved January 8, 2024, from https://modelcards.withgoogle.com/about

92. Bienefeld, N., Boss, J. M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., Willms, J., & Keller, E. (2023, May 22). Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *Npj Digital Medicine, 6*(1). https://doi.org/10.1038/s41746-023-00837-4

93. GoogleAI. (n.d.). *Responsible AI practices*. ai.google. Retrieved January 8, 2024, from https://ai.google/responsibility/responsible-ai-practices/

94. Domino. (n.d.). *Model Drift*. Retrieved January 8, 2024, from https://domino.ai/data-science-dictionary/model-drift

95. Mardziel, P. (2021, November 3). *Drift in Machine Learning*. Towards Data Science. Retrieved January 8, 2024, from https://towardsdatascience.com/drift-in-machine-learning-e49df46803a

96. Motalebi, N., & Verity, A. (2023, September 22). *Generative AI for Humanitarians*. Retrieved January 8, 2024, from https://reliefweb.int/report/world/generative-ai-humanitarians-september-2023

97. OpenAI. (2023, November 14) *Privacy Policy*. Retrieved February 9, 2024, from https://openai.com/policies/privacy-policy

98. Khan, B., Fatima, H., Qureshi, A., Kumar, S., Hanan, A., Hussain, J., & Abdullah, S. (2023, February 8). Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. *Biomedical Materials & Devices, 1*(2), 731–738. https://doi.org/10.1007/s44174-023-00063-2

99. Silva, P. (2023, June 4). *From Data to Action: Why Health Professionals Need Data Literacy Skills*. Retrieved November 29, 2023, from https://www.linkedin.com/pulse/from-data-action-why-health-professionals-need-literacy-silva/

100. Velibor, B., & Indrasen, P. (2023, April). The role of artificial intelligence in increasing the digital literacy of healthcare workers and standardization of healthcare. NA. https://doi.org/10.13140/RG.2.2.30715.80165

101. Incze, R. (2019, September 12). *The Cost of Machine Learning Projects*. Retrieved January 8, 2024, from https://medium.com/cognifeed/the-cost-of-machine-learning-projects-7ca3aea03a5c

102. Hackernoon. (2023, October 3). *Machine Learning Costs: Price Factors and Real-World Estimates*. Retrieved January 8, 2024, from https://hackernoon.com/machine-learning-costs-price-factors-and-real-world-estimates

103. Spatharou, A., Hieronimus, S., & Jenkins, J. (2020, March 10). *Transforming healthcare with AI: The impact on the workforce and organizations*. Retrieved November 29, 2023, from https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai

104. Unceta, I. (2023, March 17). *We Need To Make Machine Learning Sustainable. Here's How*. Retrieved January 8, 2024, from https://www.forbes.com/sites/esade/2023/03/17/we-need-to-make-machine-learning-sustainable-heres-how/?sh=5472222a6d25

105. DataKind. (2021). *DataKind Playbook*. Retrieved January 29, 2024, from https://playbook.datakind.org/

106. American University, Washington D.C. (n.d.). *What is a LogFrame?* Retrieved January 30, 2024, from https://programs.online.american.edu/online-graduate-certificates/project-monitoring/resources/what-is-a-logframe

107. United Nations Peace and Development Sub-Fund. (2022, June). *2030 Agenda for Sustainable Development Sub-Fund:Evaluation Framework (ANNEX 2)*. Retrieved January 30, 2024, from https://www.un.org/sites/un2.un.org/files/2022/06/evaluation_framework_of_the_2030_agenda_sub-fund.pdf

108. Mwongela, S., Patel, J., Rajasekharan, S., Wotton, L., Ahmed, M., Hacheme, G., Shibwabo, B. & Butime, J. (2023). *Data-efficient learning for healthcare queries in low-resource and code mixed language settings*. Retrieved February 6, 2024, from https://pml4dc.github.io/iclr2023/pdf/PML4DC_ICLR2023_40.pdf

Appendix

Project partners

DataKind® is a global nonprofit that harnesses the power of data science and AI in the service of humanity. For more than a decade, DataKind has empowered social impact organizations with the tools, knowledge, and capacity to solve the world's biggest challenges. Named one of Fast Company's top 10 innovative nonprofits, DataKind creates tools that improve impact, increase efficiency, and develop greater insight to solve global challenges. DataKind works extensively in the areas of economic opportunity, humanitarian response, climate & environment, and frontline health systems. DataKind mobilizes more than 30,000 volunteer data scientists and technologists to support local organizations, lead data collaborations across sectors, and provide critical context. Today, DataKind maintains a presence in the UK, India, Singapore, Kenya, and major US cities. For more information on our partners and programs, visit: www.datakind.org.

Gavi, the Vaccine Alliance is a public-private partnership that helps vaccinate half the world's children against some of the world's deadliest diseases. Since its inception in 2000, Gavi has helped to immunize a whole generation – over 981 million children – and prevented more than 16.2 million future deaths, helping to halve child mortality in 73 lower-income countries. Gavi also plays a key role in improving global health security by supporting health systems as well as funding global stockpiles for Ebola, cholera, meningococcal, and yellow fever vaccines. After two decades of progress, Gavi is now focused on protecting the next generation, above all the zero-dose children who have not received even a single vaccine shot. The Vaccine Alliance employs innovative finance and the latest technology – from drones to biometrics – to save millions more lives, prevent outbreaks before they can spread, and help countries on the road to self-sufficiency. For more information about Gavi, visit the website: https://www.gavi.org/our-alliance and Twitter: https://twitter.com/gavi.

Stakeholder table

Besides the core project team, the following stakeholders contributed to this landscape report by providing subject matter expertise, or example datasets, and participating in the survey:

Organization name and website	Type	Description
Gavi, the Vaccine Alliance	International organization	Public-private partnership that helps vaccinate half the world's children against some of the world's deadliest diseases.
GiftedMom	Health service provider	Leading mobile health solutions provider in Africa.
Khushi Baby	Technical assistance/ knowledge	Digital health platform that delivers health services to the last mile with community health workers.

Nexleaf Analytics	Health service delivery	Nonprofit technology company that builds, scales, and supports wireless sensor devices and data analytics tools to improve global public health and the environment.
Novel-T	Data science/tech company	Nonprofit organization that helps new businesses get off the ground and existing businesses to innovate.
Parsyl	Data science/tech company	Venture-backed supply chain data platform that helps shippers and insurers understand the quality conditions of sensitive and perishable products as they move through the supply chain, from the first mile to the last.
Simprints	Technical assistance/ knowledge	Nonprofit tech company that builds technology to increase transparency and effectiveness in global development, with the goal of making sure that every vaccine, every dollar, and every public good reaches the people who need them most.
UNICEF	International organization	The United Nations agency responsible for providing humanitarian and developmental aid to children worldwide.
The World Health Organization (WHO)	International organization	The United Nations agency works to promote health, keep the world safe, and serve the vulnerable.

Table 2: Overview of stakeholders

Stakeholder survey

To supplement the literature research of this landscape analysis, a stakeholder survey was conducted to explore the types of immunization-related data collected by organizations and identify the potential AI and data science needs and opportunities across various stages of the vaccine deployment life cycle.

The survey included questions regarding respondent details, such as name, organization, and area of operations, as well as the following questions:

- How do you use immunization-related data in your work and your organization's products?
- What immunization-related datasets do you use? Please briefly describe which information is captured in the datasets.
- Where do you get this data (per dataset)? If you use open and publicly available data, please provide links, if possible.
- Where do you store this data (per dataset)?
- What challenges do you regularly encounter when collecting, using, and analyzing data? For example, relating to access, coverage, quality, archiving, triangulating, interpretation, etc.
- What best practices do you follow to solve these challenges?
- What AI and data science tools or solutions are you aware of in the thematic area(s) of your/your organization's work? Please provide examples.

- What existing AI and data science tools do you use?
- How does the use of these tools impact your work? For example, allows your organization to support new use cases, boosts efficiency, increases quality, automates repetitive tasks, etc.
- Are you considering adding or exploring new AI and data science tools and technologies to your products, processes, or services? If yes, please provide a description of what these tools and technologies are.
- What, if any, is the evidence of the impact of these tools on health outcomes? If there is no recorded evidence of impact on health outcomes, please write "N/A" below.
- What are the primary challenges faced by your organization when applying AI and data science technologies across the immunization life cycle? Please select all that apply and provide examples if selecting "other".
- What do you think AI and data science should do, or has the potential to do, to improve efficiency in the thematic area(s) of your/your organization's work? Please describe the process and outputs.
- Based on your experience, please describe any successful best practices or innovative approaches for applying AI and data science technologies across the vaccine deployment life cycle.
- Please list the tasks that are done frequently or repeatedly in the thematic area(s) of your/your organization's work, if any. If there are no tasks, please write "N/A" below.
- What factors do you think could increase acceptance and use of AI in the thematic area(s) of your/your organization's work?
- What factors do you think could limit the acceptance and use of AI in the thematic area(s) of your/your organization's work?
- Are there any legal or ethical considerations that restrict the use of AI and data science technologies? Please provide examples.
- Do you follow any specific processes or frameworks to ensure data protection and data security? If yes, please provide examples.
- What are the main challenges faced by your organization in ensuring data protection, privacy, and security while collecting, accessing, or sharing data?
- Do you face any resource constraints that impact the use of AI and data science technologies? E.g., financial, technical, or human resource constraints. If yes, please provide examples.
- Do you have challenges related to expertise in AI, data science, data engineering, or analytics? If yes, please provide examples.
- In your opinion, what additional capacity or resources would be beneficial to support the use of AI and data science technologies for immunization activities?
- What are the priority gender-related barriers and inequalities in your thematic area of work and the immunization life cycle?
- How can AI and data science solutions be leveraged to address these gender-related barriers and inequalities?
- Are there any additional equity-related issues that affect the use of AI and data science tools or solutions in your context?
- Please provide any additional comments on the use of AI and data science technologies.

Evaluation of the stakeholder survey

The survey included ten organizations evenly selected across thematic areas, with three to six respondents per area. Respondents could select multiple areas to describe their organization's domain. Additionally, organizations reported activities in health policy, vaccine delivery verification (including time, place, and unique ID of dosage), stock management strengthening, and immunization program performance monitoring.

Survey findings indicated that most organizations primarily use immunization data for analytics and reporting purposes, as shown in Figure 6.

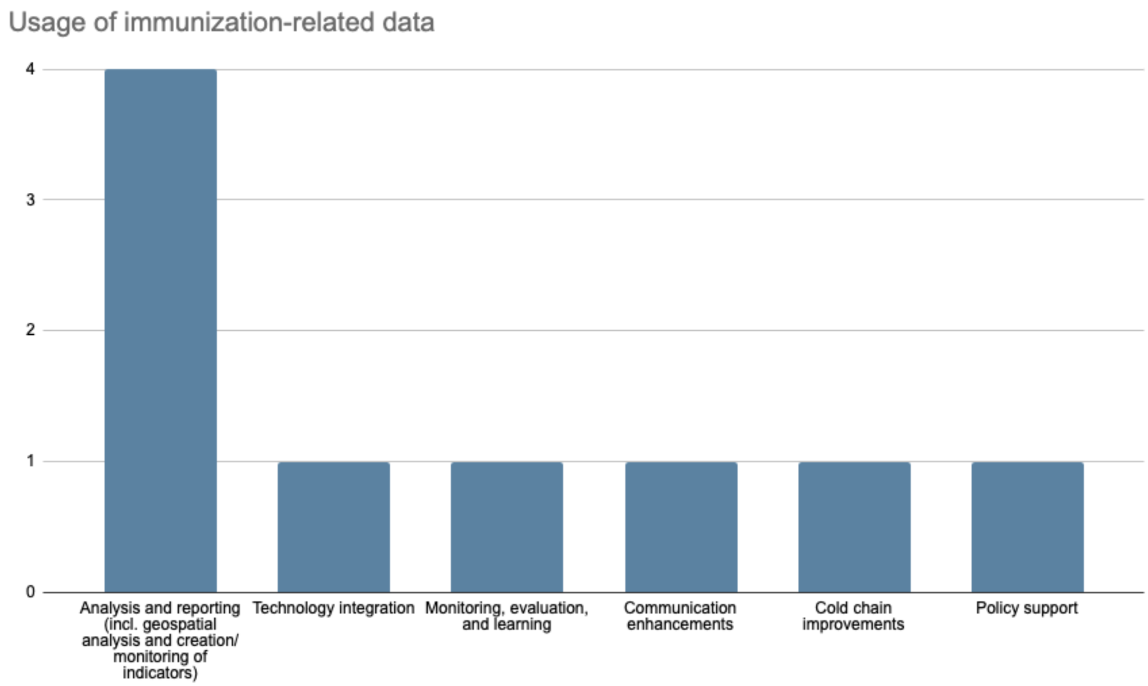

Figure 6: Use cases of immunization data across surveyed organizations

Based on the specific use case, organizations use diverse datasets, including supply chain, stock and consumption data, appointment and vaccination records, and policy and governance documents. These datasets are sourced manually at the point of care (i.e., health facilities), through sensors, by sharing with partnering organizations, or through accessing publicly available sources (e.g., the WHO/UNICEF Estimates of National Immunization Coverage datasets).

The majority of organizations use cloud data storage. Two respondents save information on local machines or opt out of saving any data, while only one organization uses an on-premise server. Regardless of how the data is stored, all respondents implement specific processes or frameworks to ensure data protection and security, including compliance with GDPR, anonymization of records, and application of information and communication technology security testing.

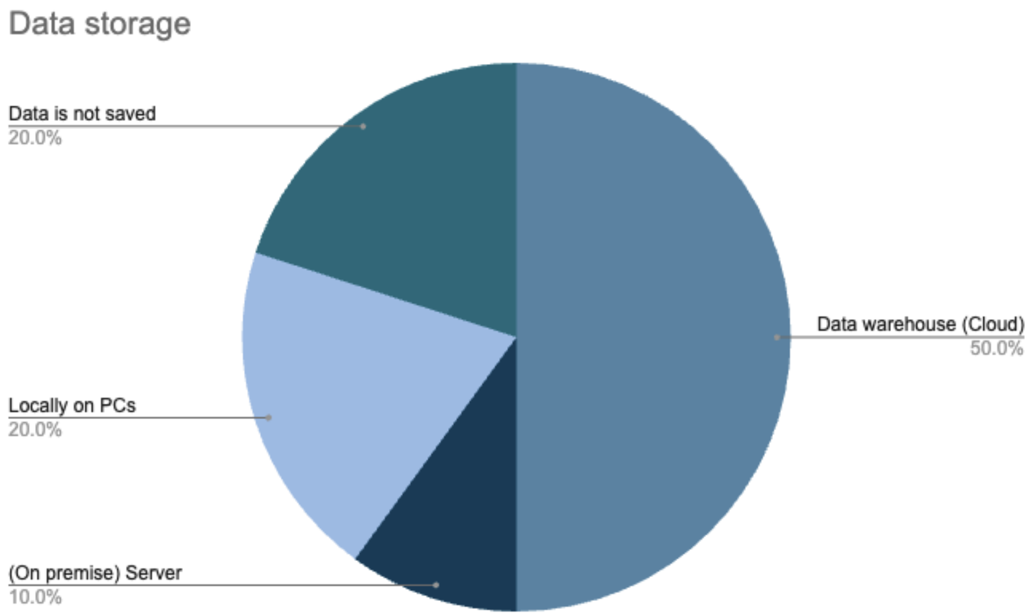

Figure 7: Data storage at the different organizations

Respondents overwhelmingly cited data quality issues, including timeliness, completeness, and accessibility, as significant challenges in data collection, analysis, and use. Less frequently mentioned challenges include data interpretation, dataset interoperability, and synchronization and validation issues. Efforts to address data quality issues involve implementing feedback loops, enhancing data collection processes, and training, and increasing the utilization of digital platforms such as the electronic logistic management information systems (eLMIS).

Knowledge of AI and data science varies widely, with four respondents reporting no knowledge, while others reporting familiarity with, and application of, advanced concepts and tools, such as convolutional neural networks and biometric matching algorithms. These tools are perceived to offer added benefits, primarily in improved planning and decision-making, and increased efficiency. Six out of ten organizations are considering adding or exploring new AI and data science tools and technologies in their products, processes, or services, while the rest are undecided. Although none of the respondents ruled out exploring these tools, only two specified concrete solutions or techniques they plan to deploy. Another two stated to still be in the exploratory phase, evaluating how best to integrate AI and data science into their organizations.

While one respondent specified the evidence and impact of AI and data science tools on health outcomes, stating reduced healthcare worker burden, time savings, and decreased data anomalies, the concrete added benefit seems to be generally difficult to pinpoint at this stage. This suggests that, although the general potential of AI and data science tools seems to be well understood, organizations are still identifying how to translate this potential into tangible improvements. For the majority of respondents, data quality and data availability are the primary obstacles to adopting AI and data science technologies, as shown in Figure 8.

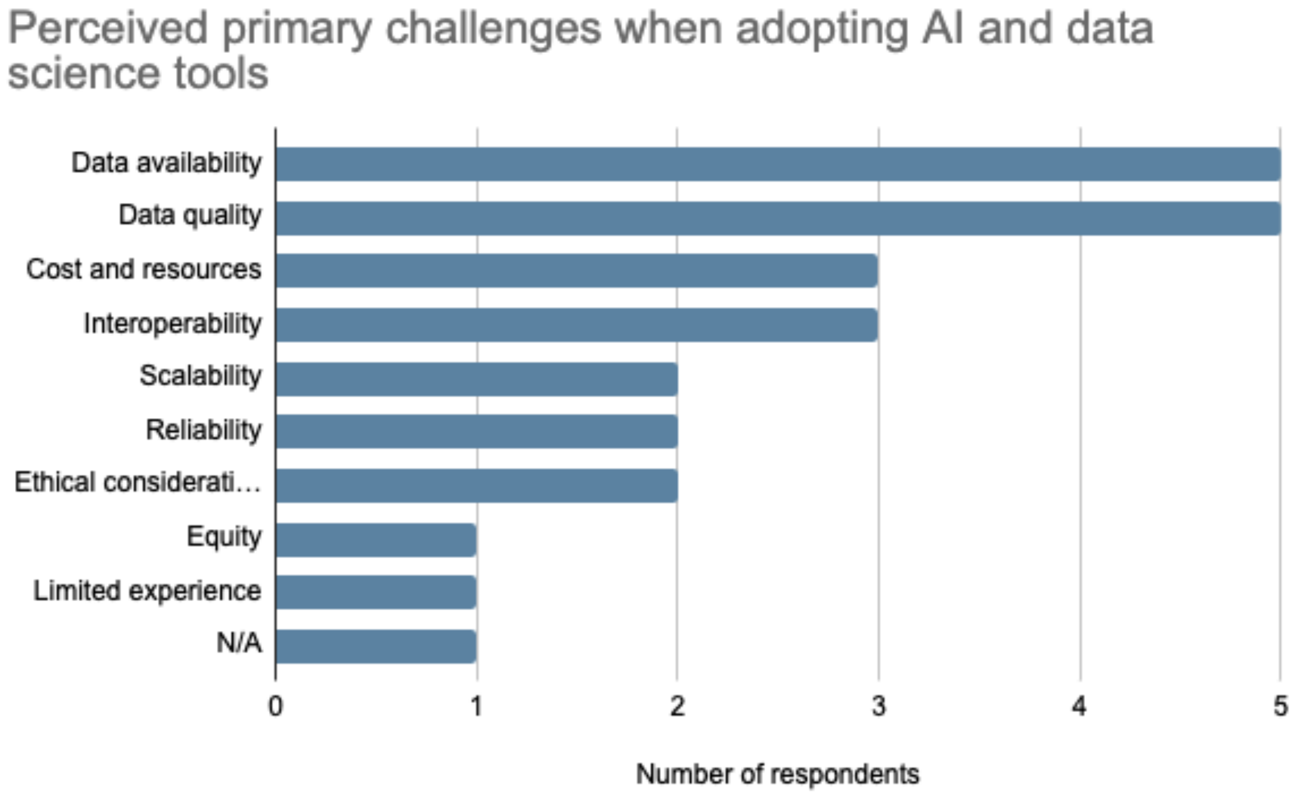

Figure 8: Perceived challenges for organizations in the adoption of AI and data science tools

Acknowledgements

This report would not have been possible without the efforts of the following:

Feedback and insights:
- Gavi, the Vaccine Alliance
- The Wellcome Trust
- INFUSE Pacesetters - GiftedMom, Khushi Baby, Nexleaf Analytics, Novel-T, Parsyl, Simprints
- Gavi Alliance Partners - UNICEF, World Health Organization
- DataKind - Mitali Ayyangar, Caitlin Augustin, John Harnisher, Matthew Harris, Lauren Woodman

Financial support:
- The Wellcome Trust

Research, writing, and report design:
- Jan Peter, Mignonne Fowlis, Melissa DiLoreto

For questions about this report:

DataKind®

contact@datakind.org
www.datakind.org