Powered by
NAACP | sanofi

Building a Healthier Future: Designing AI for Health Equity

S. CRAIG WATKINS, PHD,
ACE YOUR HEALTH AI TASK FORCE CHAIR

MATT KAMMER-KERWICK, PHD
NICOL TURNER LEE, PHD
RUBEN RATHNASINGHAM, PHD
TAYLOR WILDS, PHD CANDIDATE
CRIS WOOLSTON, PHD
ACE YOUR HEALTH AI TASK FORCE MEMBERS

NOVEMBER 2025

CONTENTS

Foreword

Artificial intelligence has entered the national dialogue in ways that demand attention, understanding, and accountability. For many in our communities, it is both a source of hope and one of concern. We are hopeful because of its promise to revolutionize medicine, close gaps in access, and finally eliminate some of the health challenges that have long plagued Black Americans. We are concerned because history has shown that when innovation moves faster than inclusion, the communities most in need of progress are often the last to benefit.

Our mission is rooted in the need for health equity as the basis of a fair and just society, ensuring that everyone can achieve their highest and healthiest potential. We recognize that health disparities arise from differences in access to care, the quality of care received, and life opportunities that influence health outcomes. Our goal is to make sure that technology does not deepen inequalities but instead drives meaningful change and system transformation.

Our members are asking the right questions. During our national convenings, we adopt NAACP-wide resolutions through a democratic process in which every member can bring forward, advocate for, and vote on an issue of particular concern to our community. Over the past two conventions, the number of proposed AI resolutions has grown exponentially. How can AI be used to prevent disease, improve care, and make health systems more responsive? How do we ensure that it does not replicate the same biases that have harmed us in the past? Curiosity about both the benefits and potential risks of this technology is growing across our branches and partners. That curiosity is a form of power! It signals engagement, ownership, and the desire to shape this next era of health care rather than be shaped by it.

This paper is an answer to that moment. It is the start of our work, not the culmination of it. The NAACP is proud to partner with Sanofi in launching this first phase of a broader effort to ensure that the design and governance of Health AI reflect equity from the ground up. What follows is not just a framework, but a call to action to make technology serve people, not the other way around.

The steps following this paper will take us deeper: cultivating shared knowledge that fuels AI and digital fluency; convening policy leaders and regulators to embed equity into rulemaking; engaging venture and capital markets to invest in technologies that serve diverse communities; and developing pathways for our members and institutions to participate directly in the design and oversight of health innovation.

AI will define the future of health care, but we intend to define what justice looks like within it.

Derrick Johnson
President and CEO, NAACP

Executive Summary

The deployment of artificial intelligence (AI) in health care must consider several ethical issues, including bias, disparate impacts, and the worsening of health equity. We define health equity as ensuring that everyone has a fair chance to be healthy, regardless of who they are, where they live, or how much money they have. Ensuring that applications of AI in health care do not worsen health outcomes should be a priority of clinicians, patient advocate groups, industry, technology developers, and policymakers, just to name a few.

While the use of AI in health care offers opportunity to accelerate advances in medicine, its application in such a high-stakes context also raises important questions about the technology and its social impacts. AI is used in a variety of ways in health care, with new uses consistently introduced. For example, AI is can be used in health care to:

- Predict the onset of various diseases.
- Conduct diagnostic reasoning.
- Manage patients' health and recovery after a minor or major medical procedure.
- Provide answers and support to patients when they are outside of the clinic.
- Take and summarize clinical notes.
- Deliver just-in-time insulin to diabetic patients, reducing the need for manual and painful needle injections.
- Make administrative management of the Electronic Health Record more efficient.
- Coach clinicians in good patient communication skills and bedside manner.
- Conduct protein modeling and design.
- Analyze multi-modal physiological data to give personalized biomarker forecasts.
- Help identify treatments for disease in research and development.
- Support manufacturing of treatments.
- Provide increased access to health care in rural communities.
- Conduct risk assessment and underwriting for insurers, as well as claims processing, prior authorization and utilization management, and customer service and support.

Notably, this list is just a small sample of what AI will be capable of. In fact, AI is doing some of these things today, but not as a matter of routine across large portions of healthcare systems around the world.

AI as a Socio-Technical System

Public, industry, and academic debates about AI have treated the technology as a technical problem that requires technical solutions. Questions guided by this approach ask, for example, “What are the technical specifications for an algorithmic model?” or “Does an AI system perform the task for which it was created?” These questions are generally regarded as the purview of experts from fields like engineering and computer science. But the deployment of AI in high-stakes contexts like health care means that AI is not simply a technical problem that demands technical solutions. The use of AI to predict the onset of disease or develop a new drug for high blood pressure is also an inherently social and ethical matter that requires sociological and ethical solutions, too. Questions guided by this approach ask, for example, “Are the prediction errors of an AI system equally distributed?” or “How do we ensure that the models for drug development work reliably across diverse patient populations?” Treating AI like a socio-technical system rather than simply a technical system forges opportunities to engage diverse expertise to support a substantive consideration of the social and ethical dilemmas posed by AI.

Thematic Examples of AI Applications in Health and Industry

The full white paper is organized around six thematic examples that illustrate the dynamic interactions between health, equity, and AI. These include:

- Patient Engagement with AI
- AI in Health Care: The View from Safety-Net Health Care Professionals
- The Use of Large Language Models in Health Care
- Synthetic Data, Health Equity, and Pharma's AI Responsibilities

- How Pharmaceutical Companies Should Approach Health, Equity, and Artificial Intelligence/Machine Learning (AI/ML)
- AI and the Future of Clinical Trials

We also present a maternal health and AI “case illustration” to underscore many of the key themes and AI governance principles highlighted throughout the white paper.

Two criteria determined our selection of these thematic examples: (1) relevance to current and anticipated applications of AI in health care and (2) how each example underscores certain nuances regarding the equity stakes in its application of AI. The thematic examples are not meant to be exhaustive. Next, we highlight three examples.

AI in Health Care: The View from Safety-Net Health Care Professionals

Researchers at The University of Texas at Austin surveyed 230 safety-net health care providers to get their perspectives on AI (Kammer-Kerwick, et al., 2025). Safety-net service providers—health care workers who primarily serve uninsured patients—represent an important but often overlooked perspective related to the use of AI in health care. One of the longstanding aspects of inequities in health care is the lack of physical proximity to critical and specialized forms of care among marginalized groups. Safety-net providers identified AI as a potential remedy, noting that the technology’s ability to help collect and analyze digital and medical biomarkers while the patient is away from the clinic could be transformational. Safety-net providers also expressed concerns about AI, including insufficient technical and financial resources to support AI training and the build of the technical infrastructure required to integrate AI into the clinical workflow (see Table A1).

Table A1. Top Benefits and Barriers to Health AI, Findings From Safety-Net Health Care Providers

Top Benefit of Health AI		Top Barrier to Health AI Integration	
Streamlined administrative tasks	38.8%	Concerns about data privacy/security	31.9%
Enhanced patient outcomes	24.2%	Insufficient staff training/knowledge	19.5%
Improved diagnostic accuracy	23.3%	Lack of funding for AI implementation	17.3%

Synthetic Data and Health Equity

An emergent solution regarding the lack of data representativeness in AI systems—an inherent form of bias—is synthetic data. More precisely, synthetic data is artificially generated data that mimics real-world data but is not collected from actual people or real environments. The core problem synthetic data solves is non-generalizability. AI models trained on narrow or biased datasets consistently underperform across demographic subgroups, care settings, and devices, which systematically amplifies existing health disparities (Gallon, 2024; Daneshjou et al., 2022). In dermatology, for instance, training datasets historically underrepresent darker skin tones—less than 15% representation despite comprising approximately 40% of the U.S. population—driving measurable performance gaps of 20–50% for certain conditions (Phillips et al., 2020; Groh et al., 2021; Adamson & Smith, 2018). To date, safety-net, rural, and resource-constrained health care sites have virtually no capacity for building datasets that align with the populations that they serve. Synthetic data has emerged as a viable, though not perfect solution, which can be used to validate and tune models without transferring sensitive patient data. The broader implication is democratization of AI development.

While synthetic data offers powerful tools for enhancing equity and privacy, it is not immune to reproducing or even amplifying biases that exist in source datasets or in the design and deployment processes. To avoid these pitfalls, it is essential that synthetic data initiatives are co-designed and continuously guided by multi-stakeholder engagement, including patients, clinicians, and local advocacy organizations.

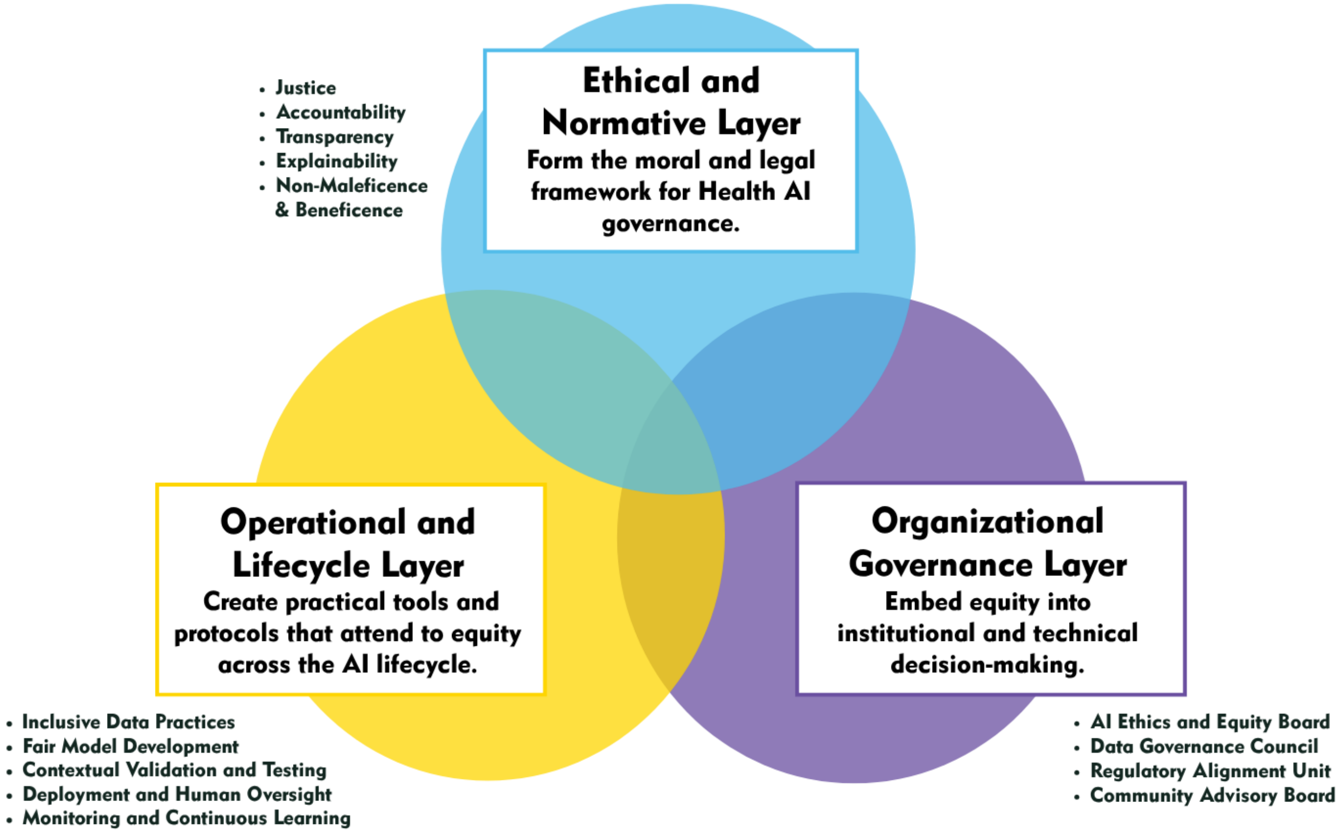
The Use of Large Language Models in Health Care

No AI technology is demanding more investment dollars, media attention, or consumer engagement than large language models (LLMs). Since the breakthrough success of ChatGPT in 2022, LLMs have been adopted in a wide range of clinical contexts, including orthopedics (Jin & Zhang, 2024), drug discovery (Lu et al., 2025), protein modeling and design (Madani et al., 2023), and the management of chronic diseases like diabetes, just to name a few. Currently, there are no formal policies or guidelines on how clinicians should use LLMs. Most clinicians understand the potential value of LLMs in health care but express concern about the ability of these systems to generate outputs that are based on clinically validated evidence. Although LLMs are being used to support diagnostic reasoning and offer mental health support to those in need, there are risks associated with the use of LLMs in health care, including the tendency of these systems to "hallucinate," that is, to make things up and add in bias against populations that may not be adequately represented in the datasets and training models used to develop frontier LLMs.

AI Governance for Health Equity

The ACE Your Health AI Task Force presents an AI governance framework (Figure A1) that features three layers: the Ethics and Normative Layer, the Organizational Governance Layer, and the Operational and Lifecycle Layer. AI governance refers to the framework of policies, processes, and technologies that ensure AI systems are developed, deployed, and managed responsibly. The main goal of the framework that we propose is to identify actionable pathways that strengthen the capacity of distinct stakeholders to advance. equitable outcomes in a steadily evolving Health AI landscape. The AI governance principles that we highlight apply to multiple stakeholders, including clinicians, data scientists, AI engineers, and IT leadership, among others.

Figure A1. AI and Health Equity Governance Framework



ETHICAL AND NORMATIVE LAYER

This layer forms the moral and legal foundation of AI governance in health care.

Select governance principle for this layer and goal: AI Transparency and Explainability to ensure that all AI systems are interpretable and understandable to clinicians, patients, and oversight bodies.

Stakeholders: Data Scientists, Product Managers, and Clinicians

Sample Actions:

- Mandate Model Cards and Data Sheets for Datasets for every deployed model.
- Require documentation of performance across subgroups.
- Make transparency reports publicly accessible.

Impact:

- Enables external auditing and community review.
- Reduces “black box” risk, that is, AI systems making health-related decisions with no transparency regarding how decisions are made.

ORGANIZATIONAL GOVERNANCE LAYER

This layer focuses on the structures that integrate equity into institutional and technical decision-making.

Select governance principle for this layer and goal: Form a Data Governance Council to ensure data integrity, representativeness, and lawful use across all AI systems.

Stakeholders: Data Stewards, IT Leadership

Sample Actions:

- Create a data provenance registry to document data set sources and demographics.
- Require bias detection and rebalancing protocols for underrepresented groups.

Impact:

- Improve data representativeness and integrity.
- Reduces hidden biases in datasets.

OPERATIONAL AND LIFECYCLE LAYER

This layer focuses on creating practical tools and routines that turn equity into action throughout the AI development process.

Select governance principle for this layer and goal: Fair Model Development and Validation to design, train, and test models with measurable fairness standards and equitable performance across population subgroups.

Stakeholders: AI Engineers, Fairness Auditors

Sample Actions:

- Apply fairness metrics to ensure parity in algorithmic outputs.
- Test models in diverse clinical settings (i.e., urban, rural).
- Require external validation before clinical integration.

Impact:

- Improves cross-groups predictive performance generalizability.
- Enhances clinical relevance across diverse environments.

Additional Actions for AI Governance

Community-Based Partnerships

Partnering with equity-oriented, health-based groups like the Association of Black Cardiologists and the National Medical Association could foster opportunities to leverage combined resources to advance AI governance. Working with cardiologists, for example, to develop datasets and AI algorithms that consider culturally specific social determinants of health could address the operational and lifecycle layer of AI governance. Engagement with health care professionals and patient advocacy groups that work specifically with populations that disproportionately suffer from cardiovascular disease could address the organizational governance layer.

Community and Stakeholder Engagement

Another specific action focuses on community and stakeholder engagement. Research cited in this white paper from the Pew Research Center, The University of Texas at Austin, and the Kaiser Family Foundation finds that there is a significant public trust gap regarding AI. This is driven by many factors, including a lack of understanding about what AI is and how AI systems function. As a result, there is a significant need for developing educational programs designed to build greater AI literacy in communities.

Health Data and Civil Rights

Data is the primary source of currency in the AI economy. The behavioral, medical, and psychosocial data that is generated from patients is extremely valuable. It is the resource that builds the datasets that train the AI and machine learning models that power Health AI systems. The Task Force believes that any future effort to build a more equity-focused approach to Health AI should consider data rights, data ownership, and data sharing agreements that afford patients greater power and influence in the AI economy. Forthcoming debates about the application of AI in health care should emphasize patient rights, which fall along several dimensions including:

- The right to know if and how AI is being used in their care.
- The right to know to what extent the AI being used has been validated and trained on representative data.
- The right to know what risks may be posed by the AI tool being used.

Using AI to Build a More Diverse Health Care Workforce

According to the U.S. Bureau of Labor Statistics (2025), total employment in the U.S. economy is expected to grow 3.1%, which is low compared to the 13% growth between 2014–2024. Jobs in the health care and social sector are projected to grow the most, 8.4%. A deeper probe of the employment projections for the next decade indicates that occupations such as nurse practitioners, physical therapist assistants, and physician assistants will grow the most. An aging population will drive labor demand in this sector as will the growing prevalence of chronic diseases like heart disease, cancer, and diabetes.

There is widespread concern that, as AI advances to perform several of the skills long associated with human intelligence such as decision-making and data analysis, the technology will lead to massive unemployment or what some refer to as the “job apocalypse” (VandeHei & Allen, 2025). But the future of AI does not have to be dystopian. MIT economist David Autor (2024) argues that AI could help workers in the middle segments of the labor market develop deeper expertise and skills that maintain—and even increase—their value within their industries, specifically, and the broader economy, generally. We believe this is especially the case with regards to health care professionals such as nurse practitioners, physical therapist assistants, and physician assistants.

The challenge ahead is, first, to identify which human skills will retain value in an AI-augmented society and, second, to ensure equal opportunity to develop those skills. This will require innovative policies, education, and workforce development programs to ensure that more rather than fewer people can develop AI and future-ready-skills.

How to Read This White Paper

The white paper is organized into five sections, which do not need to be read in any sequential order.

- **Part One** serves as a primer and offers definitions of various components of artificial intelligence and a description of the AI development cycle.
- **Part Two** presents a series of thematic examples that highlight the intersections between AI, health care, and health equity. Read this section to learn about specific applications of AI in health care and equity issues that emerge as a result.
- **Part Three** will be of interest to those who focus on AI governance. This section identifies specific ethical principles that support action-oriented approaches to AI governance.
- **Part Four** illustrates many of the themes and governance principles discussed in previous sections of the paper to highlight some of the ethical and equity challenges that emerge when AI solutions are deployed in maternal health.
- **Part Five**, the Conclusion, addresses the issue of data rights, privacy, and trust in Health AI as well as the relationship between AI, equity, and the future health care workforce.

PART ONE

Introduction

Introduction

This white paper focuses on some of the critical challenges and opportunities that are emerging as AI evolves and will play an ever-increasing role in health care. Further, the white paper is part of a larger public health initiative, ACE Your Health, made possible by a collaboration between the NAACP and Sanofi. The ACE Your Health initiative was created to help foster deeper knowledge about the state of health equity and how best to build mechanisms that improve health outcomes for all. As part of its efforts to assess the future of health care and health, the ACE Your Health initiative convened an AI Task Force to probe the complex and ever-evolving role that AI will play in shaping health futures and health equity. We define health equity as ensuring that everyone has a fair chance to be healthy, regardless of who they are, where they live, or how much money they have.

The deployment of AI in health care must consider several ethical issues, including bias, disparate impacts, and the worsening of health equity. For example, when the data used to train AI models do not represent the full population, this is called sampling bias. Take, for example, a heart risk model that is trained on data from fitness wearable users, who are mostly younger, wealthier, and more active than average. When the same model is used on older, low-income patients in a safety-net clinic, it will likely perform poorly and miss many at-risk people because those patients were not adequately represented in the training data. Another form of bias in AI that can lead to disparate impacts is measurement bias. Imagine a health algorithm that uses the number of hospital visits as a proxy for who is more likely to need health care. Because lower-income patients cannot afford health care, they may visit a doctor less frequently than their higher-income counterparts. In cases like these, the algorithm might assign lower-income patients a lower risk score and they may not be recommended for health care services even though their need for the services is higher. These are examples of how bias in health applications of AI can lead to disparate impacts and health disparities.

The ACE Your Health AI Task Force and this white paper consider several AI and health-related questions, including:

- How will AI be used to deliver health care services to patients?
- How will some of the of main features of Health AI—precision medicine, predictive medicine, and personalized medicine—transform the state of health care and impact health outcomes?
- What are some the social and ethical issues shaping the development and deployment of Health AI systems?
- What is the risk that AI will automate and replicate the legacy biases that diminish the quality of care for marginalized populations?
- How can more robust AI governance frameworks ensure that the development and deployment of Health AI will advance rather than diminish equitable health outcomes?

The pace of AI development is accelerating but adoption in health care remains gradual in some instances and non-existent in many others. Some of the reasons for slow adoption in health care include inadequate training and workforce preparation, and low levels of trust among clinicians and patients that AI can deliver safe, reliable, and unbiased results. Soon, however, AI will routinely be used to:

- Predict the onset of various diseases.
- Conduct diagnostic reasoning.
- Manage patients' health and recovery after a minor or major medical procedure.
- Provide answers and support to patients when they are outside of the clinic.
- Take and summarize clinical notes.
- Deliver just-in-time insulin to diabetic patients, reducing the need for manual and painful needle injections.
- Make administrative management of the Electronic Health Record more efficient.
- Coach clinicians in good patient communication and bedside manner.
- Conduct protein modeling and design
- Analyze multi-modal physiological data to give personalized biomarker forecasts.
- Help identify treatments for disease in research and development.

- Support manufacturing of treatments.
- Provide increased access to health care in rural communities.
- Conduct risk assessment and underwriting for insurers, as well as claims processing, prior authorization and utilization management, and customer service and support.

Notably, this list is just a small sample of what AI will be capable of. In fact, AI is doing some of these things today, but not as a matter of routine nor across large portions of the health care systems around the world.

Artificial intelligence and machine learning algorithms are a fact of daily life in the modern world. From the recommendation engines that serve up content on various media streaming platforms to the wide array of content that drives most users' social media experiences, algorithmic decision-making is pervasive. AI has a long history that dates to the 1950s when a group of scientists began discussing the potential of machines to perform acts of intelligence that were similar to humans. The notion of high-functioning machines overwhelming humans goes back even further as evidenced in the 1872 novel, Erewhon, published by English writer Samuel Butler. Much more recently, the Pew Research Center (2023) reported that most Americans, for example, could not correctly identify all six examples of everyday uses of AI presented to them. AI obviously existed in 2023, but most Americans did not necessarily know how seamlessly the technology has been woven into everyday life. Today, the public is generally much more aware of AI than they have been. Several factors explain this shift including, for example, the integration of AI into the products—smartphones, social media, streaming platforms—that we use as a matter of routine.

The most notable breakthrough in terms of public awareness of AI has been the unprecedented success of the large language model (LLM), ChatGPT. Released to the public in November of 2022, ChatGPT became one of the most quickly adopted technological innovations in human history. Although Google (Vaswani et al., 2017) introduced the transformer architecture that powers LLMs like ChatGPT in 2017, the integration of the technology into a consumer-facing interface and experience had not been established. With ChatGPT, Open AI brought user-friendly LLMs to the masses. But increasing public awareness of AI does not necessarily translate into greater AI literacy and competency. Whereas literacy refers to understanding AI's basic components, competency refers to the ability to use AI systems in life-enhancing ways.

Health Disparities in the United States

In the U.S., for example, racial and ethnic minoritized populations generally experience higher rates of morbidity and mortality than their White counterparts. Black Americans suffer higher mortality rates from heart disease, cancer, and HIV/AIDS than any other racial or ethnic population in the U.S (Ndugga et al., 2024). Hispanic Americans are roughly twice as likely as non-Hispanic whites to die from diabetes (Centers for Disease Control and Prevention, 2015). In the U.S., Asian Americans have the longest life expectancy, 83.5 years. By comparison, American Indian and Alaskan Native (65.2 years) and Black Americans (70.8 years) have the lowest life expectancy. The life expectancy for White Americans is 76.4 years and 77.7 years for Latino populations.

Several factors explain these disparate health outcomes. Historically underserved communities are more likely than their White counterparts to be uninsured, which decreases access to health care services. Uninsured individuals are more likely than insured individuals to go without medical services and less likely to receive preventative care and services. Social and economic inequalities are also a key factor. There are consistent correlations between income, educational status, and health outcomes. (Kaplan, Everson, & Lynch, 2000). Populations with lower income and educational levels experience higher rates of morbidity and mortality compared to their higher income and more educated counterparts. Social and economic factors also shape the lifestyle and behavioral conditions that contribute to health disparities. For instance, lower-income populations may lack access to safe neighborhoods and walkable streets, thus leading to sedentary or low-physical activity lifestyles (Shuval et al., 2013) . Similarly, low levels of education may correlate with lower levels of health literacy, thus contributing to racial and ethnic gaps in health outcomes. The state of the U.S. healthcare systems is also a major contributing factor in disparate rates of morbidity and mortality—a topic that we turn to next.

The U.S. Healthcare System

Because this white paper focuses on AI and its relationship to the state of health care primarily in the U.S., it is important to recognize the legacy of systemic inequality that has defined U.S. health care. The examples that we present in the white paper carefully illustrate the threats of bias and discrimination, for instance, that diminish the performance and value of clinical applications of AI in the delivery of health services. In virtually all the examples we share, evidence of bias and disparate impacts predates the presence or application of AI, suggesting a much longer history of inequities in the U.S. healthcare system. Research from public health, clinician researchers, and the social sciences have identified the multiple ways in which racial, ethnic, gender, and geographic factors impact, for example, access to health care and differences in the quality of health care services received (Smedley, Stith, & Nelson, 2003).

Racial and ethnic marginalized populations as well as rural and other safety-net populations are less likely than White or urban-based populations to have access to health care and fewer choices in where to receive care. Some populations live in what researchers describe as health care deserts, suggesting that they lack physical proximity to health services. Research has highlighted, for example, the impacts of maternal health deserts on the declining utilization rates of maternal health services for women around the world (Sarikhani et al., 2024). The explosive demand for mental health services also highlights the extent to which some populations live in areas with little or no access to mental and behavioral health services, so-called behavioral health deserts (Kaiser Family Foundation, 2022). And even when underserved populations have access to health care, research consistently find that the quality of services that they receive are often inadequate. For example, Black Americans suffering end-stage renal disease when are less likely to receive hemodialysis and kidney transplantation compared to their White counterparts even though Black Americans have a higher prevalence of kidney disease.

Similar disparities in health services have been found for cardiovascular disease (Ayanian et al., 1993) and hospital emergency visits (Todd et al., 2000).

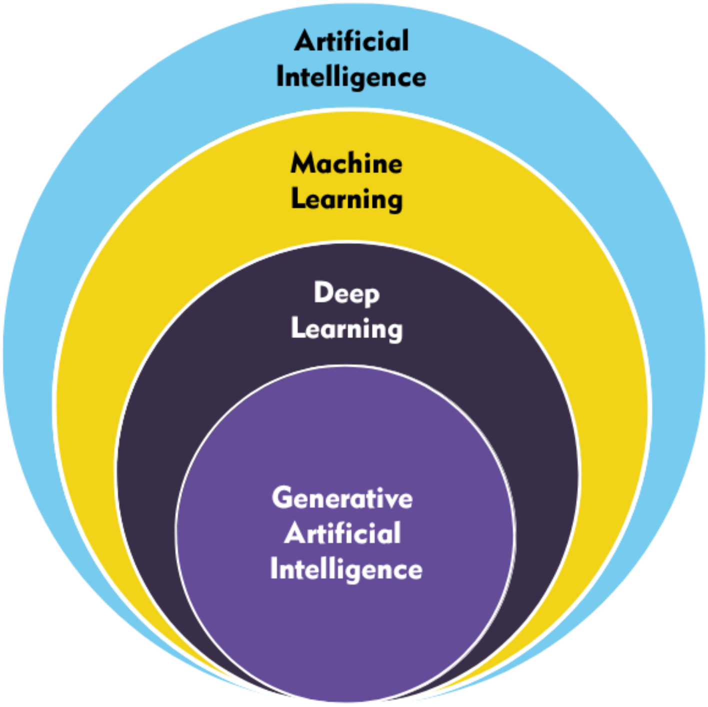
With this context in mind, the critical question regarding AI is not whether the technology will create health disparities. Rather, a major concern regarding the adoption of AI in health care is how these legacy and systemic forms of bias and discrimination can be integrated into the design and operational capabilities of clinical AI and machine learning algorithms—leading to practices that automate and scale these longstanding disparities. Throughout the white paper, we focus on the relationship between the development of Health AI systems and health equity. We offer recommendations and insights on designing and deploying AI to strategically mitigate health disparities and advance more equitable health outcomes.

Finally, the AI cultural discourse—the logic and language that is used to shape how we think about AI—is patterned primarily by the companies and interests building these systems. For example, the big AI labs and tech companies in the U.S. are defining the ways in which the public understands and talks about AI. This includes society's tendency to think of some AI technologies as human, defer to them due to their perceived intelligence and reasoning capabilities, and submit to the collection of a variety of personal, behavioral, and medical data with very few guardrails for how that data will be used. The AI companies are positioning AI in general, and their products, specifically, as inevitable facts of daily and economic life. There is, of course, a major financial interest in maintaining this view of AI, one that obscures many of the challenges related to the power asymmetries and governance issues related to AI that are discussed throughout this white paper. A major societal, educational, and political goal should be to expand who participates in the debates about AI to ensure that the transformative benefits of AI are evenly distributed.

Some Definitions

Next, we offer some basic definitions of AI and illustrations of how the technology works in general and in health care specifically. Figure 1 depicts four major types of AI and the relationship among them.

Figure 1. AI Terms and Types



The term **artificial intelligence** refers to computer systems designed to perform tasks that normally require human intelligence such as understanding language, recognizing images, or making decisions.

Machine learning is a type of AI that allows computers to learn patterns from data and make predictions or decisions without being explicitly programmed for each task. Instead of following fixed rules, the system improves its performance as it is exposed to more data. In health care, for example, machine learning is used to analyze medical images such as MRI scans to help doctors detect diseases like cancer earlier and more accurately. A machine learning model can be trained on thousands of labeled images to learn what cancerous tissues look like and then automatically flag suspicious areas on new scans for review by clinicians.

Deep learning is a specialized branch of machine learning that uses multi-layered artificial neural networks—modeled loosely after the human brain—to automatically learn complex patterns from large amounts of data. Deep learning algorithms excel at recognizing images, sounds, and language without the need for manual feature engineering, which means they can learn important patterns on their own instead of relying on humans to define which features in a dataset matter. In health care, a deep learning model can analyze patterns across time in a patient's electronic health record including lab results, medications, and vital signs to predict risks like hospital readmission, sepsis, or kidney failure. By learning how certain sequences of events lead to poor outcomes, these models can alert clinicians earlier and support preventative care before a health crisis occurs.

Generative AI is a type of AI that, not only learns patterns from large amounts of data, it uses that data to predict how to create content—text, images, audio, or video—that resembles the form of data used to train the model. In health care, generative AI models can learn from vast databases of chemical structures and biological data to generate new drug candidates that could interact effectively with specific disease targets. For example, a model might design new molecules that bind to a cancer-related protein, or an enzyme involved in diabetes. This approach can dramatically speed up the early stages of drug development and reduce the cost and time required to find promising new treatments.

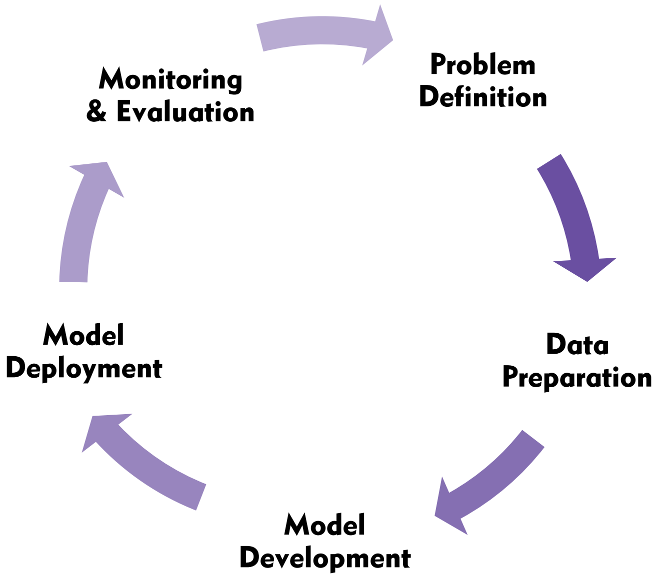
Agentic AI refers to AI systems that can act autonomously toward specific goals. In this instance, the AI system is capable of independent action and decision-making that does not require prompting or direct instruction from a human. These systems combine memory-like functions and planning to operate more like autonomous assistants rather than passive tools. Imagine an AI Care Coordinator for a patient with diabetes. An agentic system could autonomously track the patient's progress by monitoring data from glucose sensors, medication logs, and appointment records and then act based on some indication. If the diabetic patient's blood sugar trends show a rising risk, the AI could message the patient with personalized guidance, adjust their meal plan, or notify their clinician to review treatment. By acting proactively rather than reactively, agentic AI can support more real-time patient care independent of human direction.

Artificial general intelligence or AGI is a form of AI that can display markers of understanding, learning, and the application of knowledge across a wide range of tasks at a human scale. Unlike today's AI systems, which are narrower and trained for a specific task (e.g., predicting the risk of a kidney transplant patient being readmitted to the hospital less than 60-days after the procedure), AGI would have flexible reasoning-like and problem-solving-like skills that allow it to adapt to entirely new situations without additional training. In health care, an AGI system could simultaneously consider multi-modal data from a patient, including their lab results, medical images, electronic health record, and voice-based mood disorder data to predict disease risk and recommend preventative measures—working much like an experienced multidisciplinary team of clinicians. AGI only exists in theory and is not in any current real-world application.

The AI Development Cycle

One way to understand AI is to develop some basic awareness of how these systems are built through the AI development lifecycle (see Figure 2). Any AI system that is designed, developed, and deployed generally goes through a development cycle that consists, broadly speaking, of five phases: Problem Conceptualization and Definition, Data Preparation, Model Development, Model Deployment, and Monitoring and Evaluation. Each phase plays a very specific role in the development of an AI model and involves very specific requirements and levels of expertise. The **problem definition** phase involves defining the use case and purpose of an algorithm. It addresses the question: what task will the algorithm perform? During the **data preparation** phase, raw data is collected and processed to support the next phase, model development. This multi-level phase considers the question: which data inputs are most relevant to the use case and model training? In the **model development** phase, the algorithm is trained and optimized for

Figure 2. AI Development Cycle



the specific task it has been assigned. In this phase, developers ask: which features in the data are most predictive and clinically meaningful? The **model deployment** phase integrates the algorithm into a real-world workflow, thus indicating an application of some kind. This phase addresses the question: in what context will the AI model be used? Finally, the **monitoring** phase involves an evaluation of the algorithm's performance once deployed and activated. In this phase a primary question ask: How is the model performing over time?

Most equity-based discussions about AI generally focus on the problems associated with data. To be sure, inadequate data collection, processing, and labeling can diminish the performance and reliability of an AI system. This is euphemistically characterized as "garbage in (data inputs), garbage out (system outputs)." But equity issues can present across all five phases of the AI lifecycle, thus necessitating governance frameworks and processes to mitigate risks at any point during development. To illustrate this point, take for example the development of an AI algorithm designed to predict the onset of Type 2 diabetes within five years.

Developing a Predictive Algorithm for Type 2 Diabetes

Diabetes is one of the most prevalent chronic diseases in the world as an estimated 583 million people were living with the condition in 2024, according to the International Diabetes Federation (2025). The IDF projects that 1 in 8 adults, or 853 million people, will be living with diabetes by 2025. A staggeringly high number of people with diabetes, 4 in

10, are unaware that they have the disease. Undetected diabetes exposes individuals to long-term health complications and a greater likelihood of death. Some of the factors that contribute to Type 2 diabetes include an aging population, decreasing levels of physical activity, and increasing overweight and obesity prevalence. In the U.S., diabetes is most prevalent among Native Americans/Alaskan Natives (13.6%), Black Americans (12%), and Latinos (11.7%) comparted to Whites (6.9%) (American Diabetes Association, 2023). In 2021, diabetes was the eighth leading cause of death in the U.S. The prevalence of diabetes and the projected rise in cases demands innovations that can help manage the disease and, more importantly, identify the signals that predict the onset of the disease before it forms. Yet the social, environmental, behavioral, and geographical complexities of diabetes require careful consideration when developing predictive algorithms.

Next, we present an illustration of the AI development lifecycle for a Type 2 diabetes use case, highlighting the equity implications at each phase.

Problem Definition: Develop an AI algorithm to predict the onset of Type 2 diabetes within five years, based on lifestyle, clinical, and behavioral data. Some of the challenges during this phase include problem framing bias, which could occur if the algorithm does not account for lifestyle factors that are common in low-income or uninsured groups. Decisions like these could predispose the model to be unreliable in populations segments that were not accounted for in this phase of development, which could lead to inequitable applications of the algorithm.

Data Preparation: This phase may involve the composition of datasets that integrate EHR data, wearable data, and patient-reported outcomes while also including sociodemographic and lifestyle information. The potential for bias in this phase is the underrepresentation of minoritized populations, rural residents, and underinsured patients in the datasets used to train the Type 2 diabetes prediction model. By excluding diverse lifestyles, geographic areas, and economic factors that impact patient behaviors and health outcomes, the dataset and subsequent training models will be limited in their ability to detect relevant patterns that may be specific to diverse populations.

Model Development: Developers will train, optimize, and validate the model to predict diabetes onset by fine tuning or adjusting the model's setting—how fast it learns or how complex it is—to effectively predict who will and who will not develop the disease. A model may be fine-tuned to associate doctor visits, lab tests, and prescription data—high health care utilization—with high risk for chronic disease thus inadvertently treating populations who cannot access or afford health care as low risk but whose social, economic and health status may, in fact, put them at higher risk for chronic diseases like Type 2 diabetes. Once again, the prospects for differential harm—in this case underdiagnosis for populations who cannot afford health care—emerge.

Model Deployment: After being developed, the Type 2 diabetes predictive algorithm is integrated into the clinical workflow. An equity challenge in this phase includes the inadequate training of the medical staff to interpret and exercise decision-making agency in the face of algorithmic authority, which leads to automation bias or deference to AI systems even when a prediction error is evident.

Monitoring: At this stage, the bias and limitations noted above can degrade model performance leading to high rates of prediction errors for populations that are underrepresented in the training data and model fine tuning process. Without continuous monitoring and evaluation, the model could lead to misdiagnosis and, as a result, diminish the model's ability to accurately predict the onset of Type 2 diabetes in certain populations.

The development of a clinical algorithm is a detailed and finely tuned process that is susceptible to bias and equity issues along the entire cycle such as problem framing bias (Problem Definition phase), unrepresentative datasets (Data Preparation phase), inadequate staff training (Model Deployment phase), and automation bias (Monitoring phase). This example illustrates the need to be attentive to all phases of AI development in order to identify biases and, also, reduce the likelihood of AI leading to disparate health outcomes that put some populations at greater risk than others.

The ACE Your Health AI Task Force

Historically, discourses about AI have treated the technology as a technical problem that required technical solutions. Questions guided by this approach ask, for example, “What are the technical specifications for an algorithmic model?” These types of questions are generally regarded as the purview of experts from fields like engineering and computer science. But the deployment of AI in high-stakes contexts like health care means that AI is not simply a technical problem that demands technical solutions. The use of AI to predict the onset of disease or develop a new drug for high blood pressure is also an inherently social and ethical matter that requires sociological and ethical solutions, too. Questions guided by this approach ask, for example, “How do we ensure that the models for drug development work reliably across diverse patient populations?” These types of questions expand the fields of expertise in AI to include social scientists, ethicists, and domain experts, just to name a few. The ACE Your Health AI Task Force members bring diverse disciplinary training and expertise—spanning life science, sociology, decision science, and psychology—and represent sectors such as pharmaceuticals, academia, medicine, and public policy, all of which are critical for building more robust AI governance frameworks.

You can find more on the Task Force at the end of the white paper, in the Appendix.

PART TWO

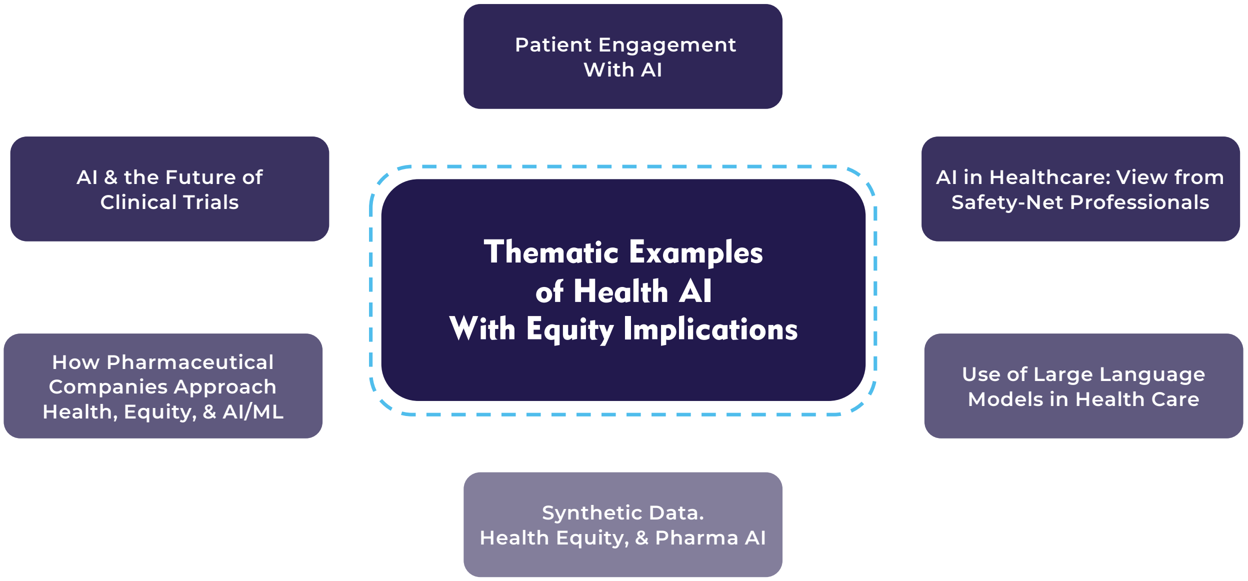
AI for Health Equity: Sample Illustrations, Risks, & Opportunities

Building a Healthier Future: Designing AI for Health Equity
ACE Your Health AI Task Force, Powered by NAACP and Sanofi

AI for Health Equity: Sample Illustrations, Risks, & Opportunities

This section of the white paper is organized around six thematic examples that illustrate the dynamic interactions between health, equity, and AI (see Figure 3). Two criteria determined our selection of these thematic examples: (1) relevance to current and anticipated applications of AI in health and (2) how each example underscores certain nuances regarding the equity stakes in clinical applications of AI. These examples are not meant to be exhaustive. Rather, they function primarily to illustrate some of the ways in which the adoption of AI in health-related settings among a diversity of stakeholders—clinicians, patients, academic medical centers, life science, and technology companies—provokes important social and ethical questions about the future of health care.

Figure 3. Thematic Examples of Health, AI, and Equity



Theme 1. Patient Engagement with AI

A 2024 report by the Kaiser Family Foundation suggests that the transition from Dr. Google to Dr. ChatGPT is underway. Indeed, a growing number of people are turning to AI-based services like ChatGPT as a source of health information and medical guidance. The Kaiser Family Foundation report finds that about one in six (17%) of adults they surveyed say they use AI chatbots at least once a month to find health information and advice. Twenty-five percent (25%) of those under the age of 30 used AI chatbots at least once a month for health-related purposes. The internet is frequently used to access information about symptoms, offer feedback and second opinions on clinical diagnoses, and generate recommendations for managing routine illnesses and minor injuries.

In research with clinician-researchers who are experimenting with the use of generative AI in their clinical decision-making process conducted by members of the ACE Your Health AI Task Force, a recurring theme is the use of chatbots like ChatGPT among patients as a source of health information and guidance. In a typical scenario a patient might input their symptoms or lab results in a chatbot to generate a diagnosis, a second opinion, or ideas for a treatment regimen. While some doctors note that this suggests a level of patient engagement in the management of their own health—an activity they welcome—they also explain how AI chatbots can promote confusion when the generated outputs are inaccurate or not applicable to the patient. Increasingly, doctors are contending with systems like ChatGPT when meeting with and treating patients, which can occasionally provoke frustration and burdens on their time and cognitive resources. In addition to finding that more people are using AI chatbots for health information and guidance, the 2024 Kaiser Family Foundation report notes that 63% of the adults surveyed responded that they were either “not too confident” or “not confident at all” in the accuracy of the information generated by AI chatbots.

Users’ engagement with AI chatbots are quite different than their engagement with more conventional informational tools like the Google search engine. A surging number of people are turning to AI chatbots for companionship and therapeutic relief. Behaviors like these are risky for several reasons. First, it could lead to an overreliance on AI chatbots that lack sufficient training in evidence-based medicine. This could expose an individual to inaccurate medical information and dangerous health outcomes. Second, it could diminish use of and trust in human professionals who are trained to deliver clinical care based on extensive training, clinical ethics, and emotional intelligence. Finally, deeply personal engagement with AI chatbots exposes highly personal and sensitive data to companies that, to date, have not been held accountable for maintaining the data privacy and rights of their users.

There is, however, an opportunity to create frameworks and approaches that strengthen people's ability to use AI expertly and safely. Expert use of AI chatbots includes, for example, developing prompt engineering skills or the ability to pose effective queries to a chatbot to enhance the generation of health information that is reliable and relevant. Skillful prompt engineering can also empower patients to refine their ability to ask questions, which can improve their communication and engagement with human clinicians. Safe use of AI includes, for example, prompting and engagement strategies that prioritize patient data and privacy.

Finally, the ability to evaluate and assess the quality of an output an AI model generates is also a component of AI literacy and agency. The "black box" nature or lack of transparency about AI-generated outputs raises concerns about their validity and safety. Even when outputs generated by a chatbot may be easily accessible, comprehensible, and appear authoritative, matters related to accuracy, clinical validity, and transparency must always be considered. Helping users develop greater proficiency in using these tools should be a priority for building a more AI-literate society.

Theme 2. AI in Health Care: The View from Safety-Net Health Care Professionals

The innovative applications of AI in health systems take place primarily in academic medical centers and well-financed technology companies. Unfortunately, this is consistent with a long history of medical innovations forming with little-to-no engagement with marginalized populations or the institutions and practitioners who provide them with health services. Take safety-net populations as an example. These are populations and health care settings made up largely of patients who are either uninsured or underinsured. Excluding safety-net practitioners and patients from the development and deployment of Health AI has several consequences. First, the problem definitions and datasets used to develop use cases for AI and to train models typically lack input and representation from safety-net constituents, who may require different approaches to framing and building AI solutions. Moreover, any advances in predictive, precision, or preventative medicine are less likely to benefit the safety-net context, thus exacerbating longstanding inequities in health care.

Any effort to integrate Health AI systems into safety-net settings requires careful thought and attention to the social, clinical, and cultural contexts that make these environments distinct but not monolithic in their needs and characteristics. Researchers from The University of Texas at Austin surveyed 230 safety-net health care providers to get their perspectives on AI (Kammer-Kerwick et al., 2025). Safety-net service providers represent an important but often overlooked perspective related to the use of AI in health care. First, because safety-net providers work closely with uninsured and underserved patients, they are keenly aware of these patients' health conditions and health care needs. Second, lived experience with safety-net patients equip these health care professionals with a depth of knowledge about the social and economic factors—social determinants of health—that impact health outcomes in safety-net settings. As a result, safety-net health care providers develop deep knowledge and expertise that could greatly inform the design of Health AI systems that can attend to the unique attributes of life in the social and economic margins.

Safety-net providers surveyed identified a number of benefits and barriers regarding the use of AI (See Table 1). In the survey, for example, they noted that AI could help reduce some of the administrative burden that health care professionals routinely contend with. Safety-net providers also note that AI, when deployed ethically and responsibly, can support patient outcomes. From their perspective, the ability to identify key patterns in behavior, environment, and social determinants could lead to deeper patient insights and, as a result, better patient outcomes. Safety-net providers also believe that the ability to streamline diagnostic accuracy could support more effective patient management for improved outcomes.

Table 1. Top Benefits and Barriers to Health AI, Findings from Safety-Net Health Care Providers

Top Benefit of Health AI		Top Barrier to Health AI Integration	
Streamlined administrative tasks	38.8%	Concerns about data privacy/security	31.9%
Enhanced patient outcomes	24.2%	Insufficient staff training/knowledge	19.5%
Improved diagnostic accuracy	23.3%	Lack of funding for AI implementation	17.3%

Importantly, the safety-net service providers identified some of the critical barriers to integrating Health AI solutions into their clinical environments. A top concern was related to data privacy and security. This perspective highlights the degree to which the larger safety-net population—care providers and community members—place a high priority on ensuring that health services are safe, non-discriminatory, and trustworthy. Another key barrier to adopting AI cited by safety-net providers surveyed is insufficient staff training and knowledge. One emerging gap in the Health AI space is the growing divide between those who have access to these systems and those who do not—a gap we refer to as the AI divide. In addition, safety-net providers noted that inadequate funding severely restricts their ability to integrate AI into their workflow. This economic barrier reflects a legacy of underdevelopment and under-resourcing that defines clinical environments serving safety-net populations.

These and other factors suggest that, even as safety-net providers understand the potential value of AI in health care for the populations that they serve, they face significant barriers to adopting the technology. Still, these providers identified some additional areas where AI could add value to their delivery of health care services. One area is patient education. As one safety-net practitioner explains:

> *"There are also a lot of educational elements to what we do with clients. You know, basic kinds of education about certain life skills, or health and aspects of health and well-being that they can learn. I'm hopeful that AI can fill that void, the void being the discrepancy between the providers that we need and the number that we actually have in order to be able to provide this service to people in our communities. I'm hopeful that the AI can help fill some of those spaces to make our work more efficient."*
>
> – Senior Health Care Leader (in Kammer-Kerwick et al., 2025)

Safety-net providers can inform the design of AI systems that help educate patients about the role of medication, behavior, and environment in their health outcomes, thus significantly enhancing a patient's health literacy and, by extension, the patient's ability to make choices that support good health outcomes.

Another area of potential impact identified by safety-net providers was the ability of AI systems to promote remote patient monitoring. One of the longstanding aspects of inequities in health care is the lack of physical proximity to critical and specialized forms of health care among marginalized groups. Safety-net providers identified AI as a potential remedy, noting that the technology's ability to help collect and analyze digital and medical biomarkers while the patient is away from the clinic could be transformational.

We believe that applications of AI in health care have much to gain by including safety-net practitioners and their patients in decisions related to the development and deployment of Health AI systems. In addition to providing unique clinical insights, the safety-net context supports building systems that are more adaptive and generalizable to larger segments of the population, including those segments that have been historically underserved by the healthcare system.

Theme 3. The Use of Large Language Models in Health Care

Since the breakthrough success of ChatGPT in 2022, large language models or LLMs have been adopted in a wide range of clinical contexts including orthopedics (Jin & Zhang, 2024), drug discovery (Lu et al., 2025), protein modeling and design (Madani, Krause, & Greene, 2023), and the management of chronic conditions like cardiovascular disease (Mathur et al., 2020), just to name a few. The use of LLMs in the delivery of health care is still largely experimental and episodic. There are no formal policies or guidelines for how clinicians should use LLMs. Most clinicians understand the potential value of LLMs in health care but express concern about the ability of these systems to generate outputs that are based on clinically validated evidence.

A 2024 study demonstrated what will be a common application of LLMs in health care in the future—diagnostic reasoning (Goh et al., 2024). The study had three conditions: (1) clinicians who performed a series of diagnostic reasoning tasks using conventional resources; (2) clinicians who executed diagnostic reasoning using ChatGPT-4 and conventional resources; and (3) ChatGPT-4 conducting diagnostic reasoning alone. To the surprise of the researchers, the LLM alone condition was the most accurate in the diagnostic reasoning experiments. This finding was surprising insofar as it contradicts the widely held view that humans + AI is generally the most effective application of AI in high-stakes decision-making contexts. The researchers did offer an important caveat, which stated that many of the clinicians who used the LLM in the study often neglected the recommendations made by the system, choosing to follow their own instincts and reasoning process. Thus, the human + AI dynamic may still be most effective in high-stakes decision-making tasks if humans know how to work more effectively with AI.

LLMs are rapidly being adopted in the mental health domain, with mixed results. This could be especially useful in a clinical space that is witnessing unprecedented demand for services but facing limited capacity in terms of the number of mental health care professionals available to provide clinical and other forms of service. LLMs in the form of chatbots, for example, are gaining some degree of acceptance among end-users as a tool for mental health therapy, emotional support, and companionship (Skjuve, et al., 2021). Unlike human therapists, for example, AI therapists are available anytime of the day, never get tired, and can display seemingly endless empathy. But unlike human therapists, most AI therapists are not clinically trained and may or may not provide evidenced-based mental health support.

Humans gravitate toward AI chatbots because they are convenient and available, non-judgmental, and inclined to flatter their users (Fogg & Naas, 1997). Furthermore, these tools have become more user-friendly, marked by the ability to engage in human-like conversations. Researchers note that LLMs may be a useful source of education, functioning as a viable tool for answering mental health-related questions that a patient might (Lawrence et al., 2024). LLMs may also be useful for assessing mental health symptoms or identifying behavioral signals that may predict the onset of a mental health condition (Lawrence et al., 2024). The rising number of AI-based mental health-based solutions suggests that LLMs, for better or worse, will be a featured aspect of future mental health care services.

There are several risks associated with the use of LLMs in health care (Lee et al., 2023). We briefly highlight two. LLMs tend to “hallucinate,” or make up information that may appear credible but is a complete fabrication. This can be especially dangerous in the context of health care, where an imprecise diagnosis or treatment plan can lead to substantial harm and even death. It is important to note that despite being labeled “intelligent,” LLMs do not technically “know” anything in the same way that humans know things (Harper, 2025). As a result, any use of LLMs in health care should be mindful that these systems generate probable answers to user-initiated prompts and not factual answers. A second risk regarding the use of LLMs in health care is the propensity for bias by these systems. LLMs are trained on specific types of data, hence their machine learning functions and outputs will be influenced by the biases inherent in any given dataset (Singhal et al., 2022). For example, if a clinical LLM is trained on data from academic medical journals, it will likely reflect studies with limited diversity given the selection bias present in many studies—reducing how well the model’s results apply to broader patient populations. Few people question that LLMs will be a significant part of future health care systems and practices. Instead, the questions are: How do we build governance frameworks that ensure that LLMs are

There are many risks associated with the use of LLMs in healthcare, including hallucinations or made-up content and bias, that is, content that is based on non-representative and, therefore, unreliable data.

trained, adopted, and evaluated with high ethical standards for safe and responsible deployment?

Theme 4. Synthetic Data, Health Equity, and Pharma's AI Responsibilities

One of the more common equity challenges involves the data preparation process that is central to AI development. A recurring theme in critiques of AI is the lack of data representativeness in the development of algorithmic solutions. Data is the primary currency in AI systems. Data is to the AI economy what fossil fuels have been to the industrial and post-industrial economies.

An emergent solution regarding the lack of data representativeness in AI systems—an inherent form of bias—is synthetic data. More precisely, synthetic data is artificially generated data that mimics real-world data but is not collected from actual people or real environments. Synthetic data is created using computer simulations, algorithms, or generative models to reproduce patterns in real data. It helps fill in gaps in existing datasets, balance underrepresented groups, and protect privacy by avoiding the use of personal information. In health care, synthetic patient records can help train diagnostic models without exposing real patient data. Researchers may use synthetic data when there is not enough data to adequately train an AI system.

Synthetic data represents a paradigm shift in addressing health care AI bias through privacy-preserving, artificially generated records that reflect real-world population distributions without exposing individual identities (Giuffrè & Shung, 2023; Synthetic Data Generation Working Group, 2024). These engineered datasets address three critical barriers to equitable AI: representation gaps, privacy constraints, and lack of realistic validation scenarios across diverse populations and care settings (Synthetic Data Generation Working Group, 2024).

The core problem synthetic data solves is non-generalizability. AI models trained on narrow or biased datasets consistently underperform across demographic subgroups, care settings, and devices, systematically amplifying existing health disparities (Gallon,

Synthetic data addresses three critical barriers to equitable AI: representation gaps, privacy constraints, and lack of realistic validation scenarios across diverse populations and care settings.

2024; Daneshjou et al., 2022). In dermatology, for instance, training datasets historically underrepresent darker skin tones—less than 15% representation despite comprising approximately 40% of the U.S. population—driving measurable performance gaps of 20–50% for certain conditions (Phillips et al., 2020; Groh et al., 2021; Adamson & Smith, 2018). Geographic bias compounds this challenge, with over 80% of training data originating from specialty clinic settings rather than primary care environments where most patients receive initial screening (Synthetic Data Generation Working Group, 2024).

Modern tools can now use AI models with simulation techniques to create realistic, computer-generated patient data (Giuffrè & Shung, 2023; Synthetic Data Generation Working Group, 2024). These systems can mimic rare medical conditions, differences between medical devices, and variations in how care is delivered—factors that AI systems often face in real clinical settings. This makes it possible to test for bias and safety before using the technology with real patients. Reliable methods also use simulated patient groups and clear checks for accuracy and confidence to ensure that results are safe, trustworthy, and fair (Dell Medical School Research Team, 2025; Encord AI Research, 2024).

Recent validation studies show that synthetic data can be useful in real medical settings. For example, expert dermatologists judged computer-generated skin images to be about 90% as realistic as actual photos, especially in how they captured different lighting and image quality—like what is often seen in primary care (Dell Medical School Research Team, 2025). Because of this high level of realism, synthetic data can help reduce bias in AI systems while protecting patient privacy by automatically deleting personal information and using only anonymous features.

The broader implication is democratization of AI development. To date, safety-net, rural, and resource-constrained health care sites have virtually no capacity for building datasets that align with the populations that they serve. One consequence of this is insufficient datasets to train AI models that are technically specific and responsive to their patient populations. When a population is not sufficiently represented in the datasets that train clinical AI models, those models are much less reliable when used to support clinical decision-making within that population. Synthetic data has emerged as a viable, though not perfect, solution. Data quality and fidelity pose challenges to the development and clinical use of synthetic data. For example, because resource-constrained health care facilities may not have modern equipment, the images that they capture may lack resolution quality. As a result, the synthetic generation of images may not capture the complex features in a clinical image. AI models trained on low-fidelity synthetic data are more likely to make incorrect predictions or miss subtle clinicals signals. Still, synthetic data can validate and tune models without transferring sensitive patient data. Multi-site sponsors can conduct privacy-aligned, region-specific validation, while regulators can evaluate bias mitigations in controlled environments (Giuffrè & Shung, 2023, 2023; Synthetic Data Generation Working Group, 2024).

Addressing Bias Risks in Synthetic Data

While synthetic data offers powerful tools for enhancing equity and privacy, it is not immune to reproducing or even amplifying biases that exist in source datasets or in the design and deployment processes. If generated without the input and oversight of impacted community stakeholders, synthetic data may perpetuate inaccurate representations, cultural blind spots, or clinical gaps that result in new disparities. For example, simulation methods that do not reflect real-world patient experiences or local care environments may introduce artifacts that skew model performance for marginalized groups.

To avoid these pitfalls, it is essential that synthetic data initiatives are co-designed and continuously guided by multi-stakeholder engagement, including patients, clinicians, and local advocacy organizations. Community representation must shape the choice of variables to simulate, validation protocols, and the benchmarks for fairness and accuracy.

This participatory approach ensures that synthetic data supports intended equity goals and is sensitive to the lived realities of those most affected by bias in Health AI.

Theme 5. How Pharmaceutical Companies Should Approach Health, Equity, and AI/ML

Pharmaceutical companies face mounting expectations to accelerate research and development timelines, diversify clinical trials, and ensure AI-enabled tools perform equitably across populations and geographies. Evidence-based practices can align innovation with equity imperatives across the development lifecycle.

Discovery and Development Equity

AI applications in target identification, hit discovery, and ADMET prediction can compress early timelines and expand chemical space exploration (Synthetic Data Generation Working Group, 2024). However, realizing equity potential requires intentional design to ensure AI-optimized compounds perform effectively across diverse populations. Integration of mechanistic modeling with AI approaches generates explainable, patient-specific predictions accounting for genetic and environmental factors affecting drug response across populations (Synthetic Data Generation Working Group, 2024). This hybrid approach addresses AI's limitations in causal reasoning while leveraging pattern recognition capabilities to identify novel therapeutic opportunities.

Post-market surveillance enhanced by AI enables continuous monitoring of real-world effectiveness across diverse patient populations, facilitating rapid identification of differential outcomes and appropriate responses (Synthetic Data Generation Working Group 2024). This capability becomes increasingly critical as regulatory agencies emphasize post-market commitments for diversity and equity (U.S. Food and Drug Administration, 2024; Health Affairs Research Consortium, 2023).

Clinical Trials and Bias Mitigation

Industry-sponsored trials frequently lag in achieving representative enrollment, with persistent disparities across race, ethnicity, and geography (BMJ Open Research Group, 2025; American Urological Association, 2024; American Society of Clinical Oncology, 2025). Diversity-by-design approaches leverage multimodal data (EHR, social determinants of health, insurance claims), outreach optimization, and barrier prediction to design inclusive recruitment while anticipating site-level operational constraints (IBM Research, 2023; Cascini et al., 2022).

Synthetic data makes it possible to test research protocols before involving real patients. It helps researchers see how well their systems perform across different devices, languages, image or audio qualities, and clinical settings. This early testing also helps identify safety limits and performance differences among patient groups in a transparent way (Synthetic Data Generation Working Group, 2024). Ongoing fairness checks of trial data—such as how responses are predicted or when a trial is stopped early—help ensure that all groups are treated fairly and that no unintended disparities arise during the study. (IBM Research, 2023; Lakhan, 2024).

Mechanistic-AI Hybrids for Generalizability

By combining pattern-recognition AI with models based on biology and cause-and-effect relationships, researchers can make better decisions about drug dose, timing, and which patients are more likely to benefit. This approach also helps predict how different groups of people might respond to treatment, reducing uncertainty after a drug is approved (Synthetic Data Generation Working Group, 2024). Using "digital twins"—computer models that simulate how a real patient might react—allows researchers to make safer go/no-go decisions in early trials, design studies that are better grounded in biology, and feed what they learn back into AI systems to make them smarter and more reliable (Synthetic Data Generation Working Group, 2024).

Transparency and Governance as Quality Attributes

Community members have repeatedly called for clear, standardized AI "labels" that explain what data an AI system was trained on, how well it performs for different groups, where it tends to make mistakes, how uncertain its results can be, and what situations it should or should not be used in (Dell Medical School Research Team, 2025; Cary et.al., 2024). Multi-stakeholders also stress that decision about AI use should not be left to experts alone. Governance needs to include community voices, voices given real authority to raise concerns or even shut down a system if necessary. Community members want clear answers to basic questions like, "Who is responsible if something goes wrong in an AI system?” or “Who has the power to deactivate the system?" (Dell Medical School Research Team, 2025).

Regulatory Alignment and Policy Framework

Regulators are now asking companies to show how well their AI systems and medical products work for different groups of people, to plan for diversity in their studies, and to keep monitoring performance over time (U.S. Food and Drug Administration, 2024; Hasanzadeh et.al., 2025). Pharmaceutical companies can gain competitive advantage and set themselves apart by making equity a core requirement before a product is released. This means clearly documenting how the system performs across subgroups, what steps are taken to mitigate bias, and how oversight is built from the start. Companies can also strengthen trust by joining privacy-protected data partnerships, using shared learning methods that do not expose patient data, and contributing to common testing tools that allow others to independently check their results (Synthetic Data Generation Working Group 2024).

Community-Centered Trust Building

Trust requires visible accountability, iterative improvement, and patient-centered transparency. Community feedback underscores the need for plain-language disclosures about where AI helps, where it fails, and what recourse exists (Dell Medical School Research Team, 2025; PLOS Digital Health, 2024). Transparent, tiered communications—general disclosure plus point-of-care details on model scope, uncertainty, and next steps—reduce overreliance and empower shared decision-making (PLOS Digital Health, 2024).

Institutionalizing community oversight through quarterly AI report cards shared with community advisory boards creates continuous improvement loops where community priorities guide technical refinements (Dell Medical School Research Team, 2025). AI report cards, also known as model cards, provide standardized documentation that strengthens transparency, accountability, and risk management protocols. This process ensures AI systems evolve in response to community needs rather than purely technical optimization metrics.

Theme 6. AI and the Future of Clinical Trials

Clinical trials are a key cornerstone in the advancement of medicine. One of the most enduring problems in the execution of clinical trials is the lack of diversity in participant populations (Oertelt-Prigione & Turner, 2024). Historically, most clinical trials participants have been White, male, urban, and higher income. Consequently, the discoveries and treatments that have been developed from trials have been optimized for subsets of the population, thus reducing the generalizability of trial results beyond the population, setting, and conditions of the trial itself.

Some of the factors that contribute to limited participation in clinical trials include recruitment barriers. Recruitment for clinical trials often relies on clinicians' referrals or site-based advertising that may overlook populations that lack access to health care services. In addition, recruitment protocols that assume high rates of literacy and require patients to fill out extensive forms and surveys may limit access to trial participation. Socioeconomic barriers such as time, transportation, and lack of access to research centers, which are often located in cities and academic hospitals, restrict access for rural residents and underserved populations. Historically, clinical trials were largely conducted in North America and other higher-income countries, thus excluding residents from more economically underdeveloped countries like those in Asia and across the continent of Africa. Another barrier is low levels of trust in the medical establishment due to historical and contemporary abuses like the U.S. Public Health Service (USPHS) Untreated Syphilis Study at Tuskegee.

Limited participation in clinical trials has real impacts and consequences for the delivery of health care, medicine, and health outcomes. For decades, women were underrepresented in cardiovascular drug and device trials as most of the research on heart disease relied heavily on middle-aged White men. Researchers incorrectly assumed that the findings would generalize. Mass and Appelman (2010) found that women were 50% more likely to be misdiagnosed after a heart attack than men, largely because trial data did not inform physician education or diagnostic algorithms for women. Early clinical trials for blood pressure medications primarily included White participants, even though hypertension prevalence is higher among Black adults. The widely used beta-blocker drugs were found to be less effective in lowering blood pressure among Black patients. These differences, which could have life and death consequences, were not recognized until after widespread use of beta-blockers, delaying the development of race-sensitive treatment guidelines.

Strategic Applications of AI in Clinical Trials

Inadequate race and ethnic representation in clinical trials limits the ability to detect notable differences between demographic groups with regards to an intervention's safety or efficacy. As Brandon Turner, a resident physician at Massachusetts General Hospital, notes, "This risk is especially important in the precision medicine era where we are increasingly reliant on exploiting small molecular differences to achieve therapeutic or diagnostic benefits" (Oertelt-Prigione & Turner, 2024). AI-based solutions, ethically designed and carefully deployed, can mitigate this and other risks by transforming the clinical trial lifecycle from design and recruitment to monitoring and analysis.

For example, AI can be deployed to predict optimal trial locations based on disease prevalence, demographic features, and health care access patterns. This particular use of AI can support the expansion of trials to underrepresented geographic locations, including rural areas. Natural Language Processing can be used to simplify informed consent forms, translate them into multiple languages, and personalize explanations. Such AI use could improve trial communication and comprehension for non-English speaking and low-literacy

The strategic application of AI could expand access to clinical trials, improve trust in trials, and pave the way for enhancing the generalizability of new medical discoveries.

populations. By using smartphones and wearables, researchers can remotely capture behavioral and physiological data, making it easier for people who cannot travel to sites to take part in clinical trials. Furthermore, real-time remote data capture and participant monitoring in clinical trials provide opportunities to measure drug-related effects with greater precision. Predictive recruitment tools could leverage EHR data to identify eligible participants across diverse systems, thus reducing the inherent biases from clinician referrals. The strategic application of AI could expand access to clinical trials, improve trust in trials, and pave the way for enhancing the generalizability of new medical discoveries.

PART THREE

Health AI Governance: Health Systems

Building a Healthier Future: Designing AI for Health Equity
ACE Your Health AI Task Force, Powered by NAACP and Sanofi

Health AI Governance: Health Systems

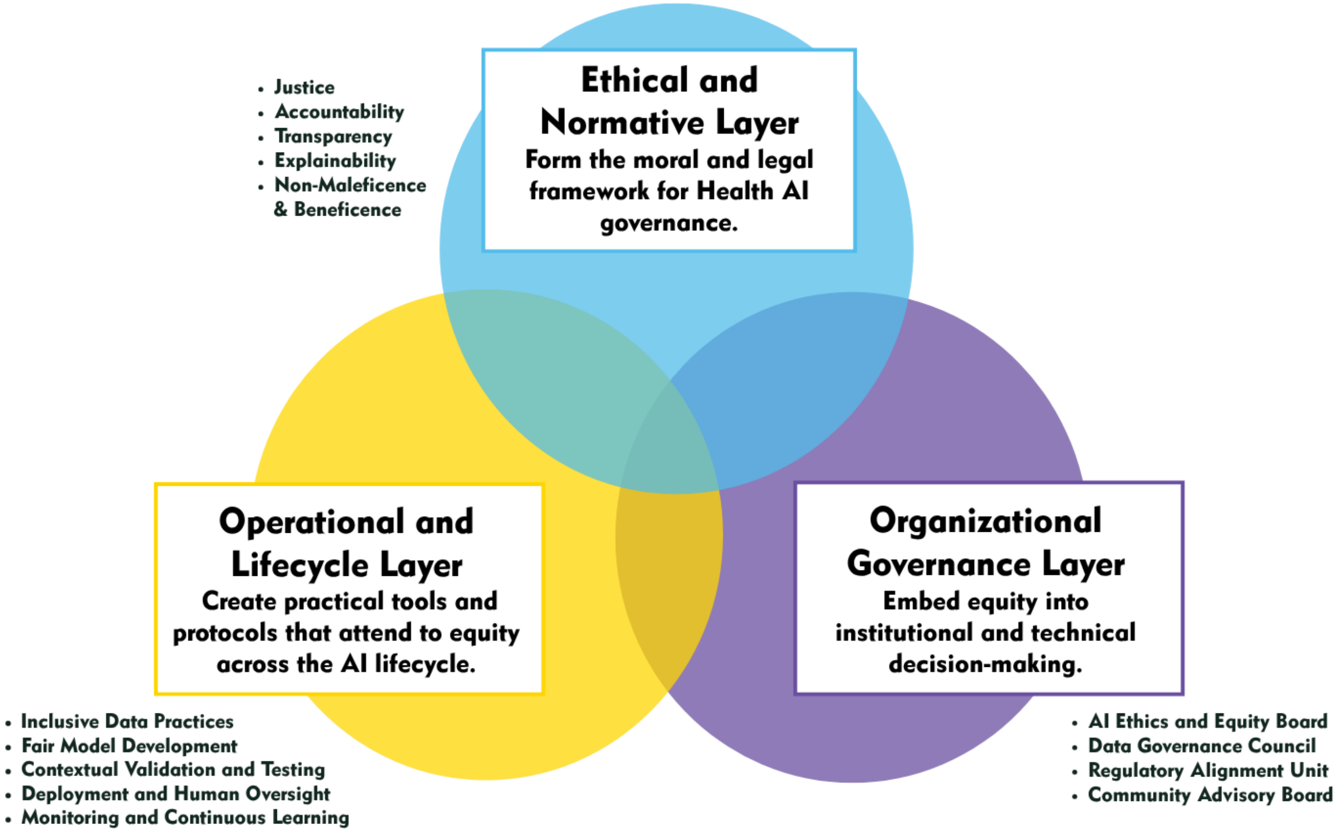
The future of safe, ethical, and responsible AI will depend largely on the development of robust AI governance mechanisms that make equitable Health AI a primary rather than a secondary consideration. In this section, we present an AI governance framework that aligns with the ACE Your Health initiative. Specifically, AI governance refers to the framework of policies, processes, and technologies that ensure AI systems are developed, deployed, and managed responsibly. The main goal of the framework that we propose is to identify actionable pathways that strengthen the capacity of distinct stakeholders to advance equitable outcome in a steadily evolving Health AI landscape. As we note throughout this white paper, the application of AI in health is rapidly evolving and, eventually, will drive much of the innovations and day-to-day operations of health care. In this paper, we focus on governance that aims to ensure that AI systems are transparent, accountable, safe, and aligned specifically with societal values such as fairness, justice, and equity when used within direct healthcare systems. In future papers, ACE Your Health will explore the application of this governance framework for other stakeholders, including patients, industry, policymakers, and local communities.

Any approach to building an AI governance framework should be adaptive, which is to say responsive to an ever-changing AI environment. These changes may be technical, social, or ethical. Broadly speaking, our AI governance framework keeps two things in mind. First, because the technical capabilities of AI systems are constantly evolving, approaches to AI governance must be agile and flexible. Not only is the technology changing, but the adoption of AI systems and subsequent use cases are evolving too, which has implications for how we think about AI governance. Second, AI governance is not a one-size-fits-all solution. More specifically, AI governance has very distinct considerations, applications, and implications for the many stakeholders that make up the health care ecosystem. This includes health care organizations, health-based industry, health care practitioners, patients and patient advocacy groups, insurance companies, and community-based entities. So, while we present a three-layer AI governance framework, we also acknowledge the nuance and complexities of the broader AI landscape, the different interests of multiple stakeholders, and the challenges associated with building more safe, ethical, and responsible AI systems.

Our proposed AI governance framework (Figure 4) is composed of three components:

1. The **Ethics and Normative Layer** forms the moral and legal foundation of AI governance in health care.
2. The **Organizational Governance Layer** focuses on the structures that integrate equity into institutional and technical decision-making.
3. The **Operational and Lifecycle Layer** focuses on creating practical tools and routines that turn equity into action throughout the AI development process.

Figure 4. AI and Health Equity Governance Framework



Next, we briefly elaborate on each layer and offer sample recommendations for action within direct healthcare systems.

Ethical and Normative Layer

What actions can be taken?

1. Develop Ethical Principles and Mission Alignment

Goal: Establish a shared ethical framework for all AI-related activities.

Stakeholders: Health System Leadership, Chiefs Ethics Officer, AI Governance Board

Recommended Actions:

- Require all AI initiatives to undergo an "Ethical Impact Review" before launch.
- Integrate equity goals into key performance indicators (e.g., reduce disparity in algorithmic performance by 25%).

Impact: Builds organizational culture centered on ethical stewardship.

2. Transparency and Explainability

Goal: Ensure all AI systems are interpretable and understandable to clinicians, patients, and oversight bodies.

Stakeholders: Data Scientists, Product Managers, Clinicians, Patients

Recommended Actions:

- Mandate Model Cards and Data Sheets for Datasets for every deployed model.
- Require documentation of performance across subgroups.
- Make transparency reports publicly accessible.

Impact:

- Enables external auditing and community review.
- Reduces "black box" risk.

3. Accountability and Governance Structure

Goal: Ensure clear responsibility for ethical and fairness outcomes in all AI-related decisions.

Stakeholders: C-Suite Executives, Compliance Teams, Legal Counsel

Recommended Actions:

- Designate a Chief AI Accountability Officer responsible for oversight of all AI models.
- Create reporting and redress mechanisms for algorithmic harms.

Impact:

- Establish ownership and responsibility.
- Creates measurable performance incentives.

Table 2. Policy and Actions for Ethical and Normative Layer

Dimension	Stakeholders	Actions & Policies	Anticipated Impacts
Ethical Principles & Mission Alignment	Health System Leadership, Chief Ethics Officer, AI Governance Board	Require all AI initiatives to undergo a pre-launch “Ethical Impact Review.” Integrate equity goals into KPIs (e.g., reduce disparity in algorithmic performance by 25%).	Builds organizational culture centered on ethical stewardship. Creates clear accountability for fairness outcomes
Transparency and Explainability	Data Scientists, Clinicians, Regulatory Compliance Officers, Patients	Mandate Model Cards and Data Sheets for Datasets for every deployed model. Require documentation of performance across subgroups. Make transparency reports publicly accessible.	Enables external auditing and community review. Reduces “black box” risk or lack of algorithmic transparency.
Accountability and Governance Structure	C-Suite Executives, Compliance Teams, Legal Counsel	Designate a Chief AI Accountability Officer responsible for oversight of all AI models. Create reporting and redress mechanisms for algorithmic harms.	Establishes ownership and responsibility. Creates measurable performance incentives.

Organizational Governance Layer

What actions can be taken?

1. Equity Impact Assessments (EIAs)

Goal: Identify, measure, and mitigate potential inequities arising from AI design or deployment; embed equity into institutional and technical decision-making processes.

Stakeholder: AI Project Mangers

Recommended Action: Standardize EIAs as part of project initiation and post-deployment reviews.

Impact:

- Identifies structural inequities early.
- Embeds continuous accountability.

2. Form an AI Ethics and Equity Review Board

Goal: Provide oversight, review, and approval of all AI systems before deployment.

Stakeholders: Clinicians, Data Scientists, Community Representatives

Recommended Actions:

- Form a multidisciplinary Equity Review Board to evaluate datasets and models.
- Include community members in review panels.

Impact:

- Ensure models reflect community needs.
- Improves legitimacy and ethical oversight.

3. Data Governance Council

Goal: Ensure data integrity, representativeness, and lawful use across all AI systems.

Stakeholders: Data Stewards, IT Leadership

Actions:

- Create a data provenance registry to document data set sources and demographics.
- Require bias detection and rebalancing protocols for underrepresented groups.

Impact:

- Improve data representativeness and integrity.
- Reduces hidden biases in datasets.

Table 3. Policy and Actions for Organizational Governance Layer

Dimension	Stakeholders	Actions & Policies	Anticipated Impacts
Equity Impact Assessments (EIA)	AI Project Managers	Standardize EIAs as part of project initiation and post-deployment reviews.	Identifies structural inequities early. Embeds continuous accountability.
AI Ethics and Equity Board	Clinicians, Data Scientists, Community Representatives	Form a multidisciplinary Equity Review Board to evaluate datasets and models. Include community members in review panels.	Ensure models reflect community needs. Improves legitimacy and ethical oversight.
Data Governance Council	Data Stewards, IT Leadership	Create a Data Provenance Registry to document data set sources and demographics. Require bias detection and rebalancing protocols for underrepresented groups.	Improves data representativeness and integrity. Reduces hidden biases in datasets.

Operational and Lifecycle Layer

This layer considers the practical tools and protocols that attend to equity across the AI development lifecycle.

What actions can be taken?

1. Inclusive Data Practices

Goal: Ensure that datasets used to train and validate AI systems are representative, ethical, and demographically inclusive.

Stakeholders: Data Engineers, Clinical Researchers

Recommended Actions:

- Create and maintain Datasheets for Datasets detailing provenance, consent, demographic composition, and known limitations.
- Require minimum thresholds for subgroup representation.

Impact:

- Improves models generalizability.
- Reduces data bias and blind spots.

2. Fair Model Development and Validation

Goal: Design, train, and test models with measurable fairness standards and equitable performance across population subgroups.

Stakeholders: AI Engineers, Fairness Auditors

Recommended Actions:

- Apply fairness metrics to ensure parity in algorithmic outputs.
- Test models in diverse clinical setting (i.e., urban, rural).
- Require external validation before clinical integration.

Impact:

- Improves cross-groups predictive performance generalizability.
- Enhances clinical relevance across diverse environments.

3. Monitoring and Auditing

Goal: Continuously track AI performance and fairness after deployment to detect bias, drift, or harm in real-world use.

Stakeholders: Compliance Teams, Data Scientists

Recommended Actions:

- Implement continuous post-deployment audits for bias and differential harm.
- Require public annual AI Equity Reports.

Impact:

- Ensures sustained fairness over time.
- Builds regulatory readiness.

Table 4. Policy and Actions for Operational and Lifecycle Layer

Dimension	Stakeholders	Actions & Policies	Anticipated Impacts
Inclusive Data Practices	Data engineers, clinical researchers	Mandate Datasheets for Datasets. Require minimum thresholds for subgroup representation.	Improves models' generalizability. Reduces data bias and blind spots.
Fair Model Development and Validation	AI Engineers, Fairness Auditors	Apply fairness metrics to ensure parity in algorithmic outputs. Test models in diverse clinical settings (i.e., urban, rural). Require external validation before clinical integration.	Improves cross-groups predictive performance. Enhances clinical relevance across diverse environments.
Monitoring and Auditing	Compliance Teams, Data Scientists	Implement continuous post-deployment audits for bias, drift, and differential harm. Require public annual AI Equity Audit Reports	Ensures sustained fairness over time. Builds regulatory readiness.

Additional Actions for AI Governance

Next, we present three additional actions to demonstrate the kinds of real-world initiatives that can translate AI governance into organizational action, cross-sector engagement, and real-world impact.

1. **Grassroots Partnerships**. Facilitate collaborations with various groups invested in health equity, AI and technology equity, and community empowerment. A recurring theme in this white paper is the recognition that most AI systems neglect to integrate data that is responsive to the lived experiences of marginalized populations including, for example, people of color, rural populations, and the uninsured. There are community-based organizations that are expanding conventional notions of what kinds of information, values, and

experiences can make datasets more culturally responsive. We call this value-based data. Additionally, partnering with equity-oriented health-based groups like the Association of Black Cardiologists and the National Medical Association could foster opportunities to leverage combined resources to advance AI governance. Working with cardiologists, for example, to develop datasets and AI algorithms that consider culturally specific social determinants of health could address the Operational and Lifecycle Layer of AI governance. Engagement with health care professionals and patient advocacy groups that work specifically with populations that disproportionately suffer from cardiovascular disease could address the Organizational Governance Layer.

2. **Community and Stakeholder Engagement.** Another specific action focuses on community and stakeholder engagement. Research cited in this white paper from the Pew Research Center, The University of Texas at Austin, and the Kaiser Family Foundation finds that there is a significant public trust gap regarding AI. This is driven by many factors, including a lack of understanding about what AI is and how AI systems function. As a result, there is a significant need to develop educational programs designed to build greater AI literacy at the community level.
3. **Establish an AI Governance Committee.** As organizations integrate AI into operations related to research and product development, they should create an AI Governance Committee. The committee might include members from legal, compliance, data science, AI governance, and user advocacy teams. The committee could meet on a regular cadence to review projects, monitor risks, and ensure accountability. The committee could support the development of an AI governance process that has clear tasks and goals that define and orient its mission, specifically, and the AI-related activities of the organization more broadly, thus meeting some of the goals of the Ethical and Normative Layer.

PART FOUR

Maternal Health & AI: A Case Illustration

Maternal Health & AI: A Case Illustration

Next, we apply many of the issues we have discussed in this white paper to a distinct health care issue in the U.S.—maternal health. Specifically, we map some of the risks and opportunities of deploying wearables and algorithmic modeling in the context of maternal health and, also, identify concrete actions for deploying AI in a more culturally responsive way. Maternal mortality is a growing public health crisis in the United States, underscored by the fact that Black women are four times more likely than White women to die from pregnancy-related causes (Petersen et al., 2019). In recent years, researchers have experimented with applying AI and wearable technologies to maternal health care, a development that highlights many of the opportunities and risks noted throughout this white paper.

Maternal Health and AI: Automating Inequalities

In the digital health and AI landscape, the inequities in maternal health are mirrored and sometimes even magnified. Many machine learning models used in clinical risk prediction, including those for maternal complications, are trained on EHRs that reflect past biases in diagnosis and treatment. Without careful subgroup validation, these models risk reproducing historical disparities and misclassifying risk for already marginalized patients (Obermeyer et al., 2019). Moreover, racial and ethnic minorities are consistently underrepresented in clinical trials, digital health studies, and datasets used to develop maternal health tools (Vargas et al., 2021). This underrepresentation limits the generalizability and safety of AI applications, including wearable technologies that rely on biometric inputs like vocal tone, activity, or language use. When communities of color are excluded from the development and validation of new technologies, those tools are unlikely to serve their needs and risk reinforcing the same inequities they aim to address.

Real-Time Care: The Promise and Pitfalls of Wearable Technologies

Wearable and ambient sensing technologies are poised to transform maternal health by offering new ways to capture and analyze behavioral and physiological processes in real time. These tools—ranging from heart rate monitors to smart clothing, voice-activated audio devices, and respiratory sensors—have the potential to fill critical gaps in perinatal care. Unlike periodic clinic visits or self-report questionnaires, wearable devices can passively and continuously monitor markers such as sleep patterns, heart rate variability, physical activity, and stress physiology across the prenatal and postpartum period (Chan et al., 2021; Liu et al., 2023). This approach provides a high-resolution window into the lived experiences of mothers and infants during a time when early intervention can have lifelong impact.

In maternal health specifically, real-time data collection has enabled researchers to detect early signs of maternal depression and anxiety, and monitor caregiving behaviors such as holding, vocal soothing, or sleep–wake alignment. Further, machine learning approaches are increasingly being used to model stress responses, classify caregiving behaviors, and predict outcomes related to bonding, distress, or regulatory capacity (Goodday et al., 2021; Chien et al., 2020). These developments signal a shift toward personalized, preventive maternal health models that extend beyond the clinic and into the home.

However, the integration of wearables and AI into maternal health is not without risk. The proliferation of commercial devices raises urgent questions about equity, access, and algorithmic bias. Wearables remain underused among lower-income families and rural populations due to cost and digital literacy barriers (Torous et al., 2021). Even when tools are available, most commercial algorithms have been trained and validated on specific populations (e.g., predominantly White, English-speaking, and middle-class users) leading to performance issues in other populations. Studies have documented sensor inaccuracies in individuals with darker skin tones and misclassification of behavioral cues based on language, dialect, or cultural caregiving norms (Buolamwini & Gebru, 2018; Fawzy et al., 2022).

Privacy and autonomy also emerge as key concerns, especially in communities with justified mistrust of medical and surveillance systems. Many wearable devices collect passive data, including audio, location, and biometric signals, raising questions about who owns the data, how it is stored, and if it could be repurposed for non-clinical use.

Culturally Responsive and Ethical AI Design

When grounded in culturally responsive design, inclusive validation, and ethical safeguards, these technologies can support early identification of risk, reduce barriers to accessing care, and empower mothers through personalized feedback and real-time support. As wearables continue to enter maternal health research and practice, they must be accompanied by equity-centered frameworks that prioritize usability, trust, and representation across every stage of development and deployment.

Ethical AI design in maternal health must reflect the realities of those most at risk. Rather than optimizing models solely for predictive accuracy, researchers and developers must co-create tools with the communities they aim to serve. This includes building AI systems that incorporate linguistically and culturally responsive interfaces, allow for participant data control and opt-out mechanisms, and are validated across diverse caregiving environments and family structures. Community health workers, physicians, doulas, nurses, and behavioral health counselors must be integrated into both the development and deployment of AI-enhanced sensing systems, not only as users, but as experts. Doing so ensures that these technologies augment rather than replace relational care and that their integration into maternal health systems reflects real-world needs, values, and constraints.

Rather than optimizing models solely for predictive accuracy, researchers and developers must co-create tools with the communities they aim to serve. This includes building AI systems that incorporate linguistically and culturally responsive interfaces, allow for participant data control and opt-out mechanisms, and are validated across diverse caregiving environments and family structures.

Health AI Governance for Maternal Health

Here we apply selected AI governance principles to the domain of maternal health to demonstrate a sample set of specific actions that could support the development and deployment of safe and ethical AI in this specific context.

Ethical use of wearables to develop algorithms that capture mother-infant data should:

- Prioritize accessible data use agreements **(Operational and Lifecyle Layer).**
- Mandate routine audits of algorithmic performance across demographic groups **(Operational and Lifecyle Layer).**
- Establish clear documentation of model training datasets to prevent bias and misclassification **(Operational and Lifecyle Layer).**
- Support the caregiving workforce by ensuring that community-based providers receive adequate training and infrastructure to interpret AI outputs and integrate them into practice. **(Ethical and Normative Layer).**

PART FIVE

Conclusion

Conclusion

Rights, Privacy, and Trust

During the roll-out of the ACE Your Health Initiative for the Emerging Technologies Involving Artificial Intelligence and Machine Learning Track, the NAACP and Sanofi held three convenings: (1) a webinar to introduce this Task Force; (2) a panel during Congressional Black Caucus Annual Legislative Conference; and (3) a panel that was featured at AfroTech, a technology-centered conference focused on innovation, inclusion, and impact. Ahead of these events, attendees were invited to submit questions regarding the role of AI in health care. These questions, along with other conversations and considerations, serve as a proxy for the Health and AI issues resonating with the broader constituency engaged by ACE Your Health. One set of questions cluster around the theme of *Rights, Privacy, and Trust.* Sample questions from this cluster include:

- What recourse do I have if I believe an AI system made a mistake in my diagnosis or treatment? How can I ensure my doctor overrides the technology if necessary?
- How do I know if the AI tools being used by my health care provider are contributing to or reducing existing racial or ethnic biases in care?
- Is it possible to request that my doctor not use AI in my care, and what are the potential pros and cons of making that decision?

Questions like these remind us that patients should not only have a voice in the development and deployment of AI in contexts that affect their health and well-being; they should be included in the decision-making process regarding how these systems are used in the delivery of the care that they receive.

Forthcoming debates about the application of AI in direct health care should emphasize the rights of patients, which fall along several dimensions including:

- The right to know if and how AI is being used in their care.

- The right to know to what extent the AI being used has been validated and trained on representative data.
- The right to know what risks may be posed by the AI tool being used.

Furthermore, patients should have the right to exercise some degree of control regarding how their data is being used. As we note earlier in this paper, data is the primary source of currency in the AI economy. The behavioral, medical, and psychosocial data that is generated from patients is extremely valuable. It is the resource that builds the datasets that train the AI and machine learning models that power Health AI systems. Some of the questions submitted to the ACE Your Health initiative underscore a growing awareness of the clinical and legal value of patient data. The Task Force believes that any future effort to build a more equity-focused approach to Health AI should consider data rights, data ownership, and data sharing agreements that afford patients greater power and influence in the AI economy.

Using AI to Build a More Diverse Health Care Workforce

According to the U.S. Bureau of Labor Statistics (BLS) (2025) the U.S. economy is projected to add 5.2 million jobs between 2024–2034. Total employment during this period is expected to grow 3.1%, which is low compared to the 13% growth between 2014-2024. Jobs in the health care and social sector are projected to grow the most, 8.4%. An aging population will drive labor demand in this sector as will the growing prevalence of chronic disease conditions like heart disease, cancer, and diabetes.

A deeper probe of the employment projections for the next decade indicates that specific occupations in health care will grow at an especially sharp pace. These include, health care support occupations and health care practitioners, which are projected to grow 12.4% compared to the overall employment growth rate of 3.1%. Specific occupations such as nurse practitioners, physical therapist assistants, and physician assistants will grow the most. In addition, the BLS projects that community and social service occupations will grow by 6.4% during this period in response to increasing demand for counseling and support services.

Three things stand out in these projections. First, many occupations projected to grow in health care—nurse practitioners, social services, physician assistants, and counseling—will play critical roles in safety-net health settings. Second, these occupations will grow, in part, due to increases in chronic diseases—conditions that disproportionately effect safety-net populations. Third, these occupations are gaining greater importance as health care systems and practitioners attempt to be more holistic and attentive to the non-medical drivers of health, i.e., the social determinants of health.

In this white paper, we have referred to the AI divide. One dimension of the AI divide is the sharp division between clinicians who are learning how to integrate AI into their clinical decision-making and those who are not. The clinicians who are developing greater AI competency will gain a significant advantage over other clinicians in areas such as diagnostic reasoning, patient management, and their ability to identify, sooner rather than later, the medical and behavioral signals that may predict the onset of a serious health problem for a patient (IC^2 Institute, 2025). By contrast, vast segments of the global health care workforce lack sufficient opportunities to develop the clinical, social, and ethical competencies that will be required to use AI effectively in clinical settings. At stake is the underdevelopment of an AI literate and empowered health care workforce.

There is widespread concern that as AI advances and performs several of the skills long associated with human intelligence such as decision-making and data analysis, the technology will lead to massive unemployment or what some refer to as the “job apocalypse (VandeHei & Allen, 2025). Current research suggests that AI will not necessarily take over whole jobs. Rather, one potential scenario is “task encroachment,” or the idea that AI will take over specific tasks within any given occupation rather than whole occupations (Susskind, 2020). For example, a machine learning algorithm might be used to analyze a CT or MRI scan for medical irregularities, but the radiologist will be responsible for synthesizing those results with patient history, lab data, and physical findings to reach nuanced judgments regarding patient care.

Whereas AI is reducing the need for some types of occupations and workers, it is also creating new occupations and requiring new skills in already established occupations (IC^2 Institute, 2025). The challenge ahead is, first, to identify the human skills that will retain value in an AI-augmented society and, second, to ensure equal opportunity to develop those skills. This will require innovative policies, education, and workforce development programs to ensure that more people can develop AI and future-ready-skills.

The future of AI does not have to be dystopian. MIT economist David Autor (2024) argues that AI could be used to build up rather than tear down the middle segments of the paid labor force. More specifically, researchers like Autor note that, if designed strategically, AI could help workers in the middle segments of the labor market develop deeper expertise and skills that maintain—and even increase—their value within their industries, specifically, and the broader economy, generally. We believe this is especially the case with regards to health care professionals such as nurse practitioners, physical therapist assistants, and physician assistants.

Educators, policy makers, and health care managers should prioritize the development of educational and training programs that prepare personnel in these occupations to use AI to sharpen their expertise, modernize their skills, fortify their value, and advance the delivery of more precision and personalized care in health care ecosystems.

Shaping the Future of Health AI

AI is much more than a technical solution; it is a socio-technical system that has the potential to transform health care across populations and communities. When viewed through this lens, there are opportunities to engage diverse stakeholders and expertise to support thoughtful and thorough consideration of the social and ethical dilemmas AI poses. Our aim is for the collective parts of the health ecosystem, the civil rights community, industry, communities of practice, and the public to equitably understand and apply the digital and technological tools transforming the world around us. This white paper serves as a primer to ignite the socio-technical transformation necessary to ensure that AI/ML are developed, funded, and deployed within an ethically-centered, equity-first model. From here, we are poised to go deeper: cultivating shared knowledge that fuels AI and digital fluency; convening policy leaders and regulators to embed equity into rulemaking; engaging venture and capital markets to invest in technologies that serve diverse communities; and developing pathways for our members and institutions to participate directly in the design and oversight of health innovation. As AI defines the future of health care, we can define the future of Health AI.

A Final Note

The Task Force recognizes the significant toll that the development of AI is having on the environment in general, and for certain populations more specifically. The New York Times reports that the big AI companies plan to invest $375 billions in data centers in 2025 and $500 billion in 2026. Media reports (Mozur et al., 2025) are beginning to document the social, environmental, and health impacts of these data centers on communities around the world. This includes extreme demands on local water resources and electricity grids that threaten quality of life and health outcomes. The Task Force supports a position adopted by the NAACP that, among other things, maintains the need to build the capacity to ensure that international, federal, state, and local government agencies mandate comprehensive environmental impact assessments for all data centers to mitigate the environmental and health risks posed by this massive new infrastructure.

APPENDIX

About the Task Force

References

About the Task Force

Matt Kammer-Kerwick is a PhD trained research scientist and the head of the Bureau of Business Research at the IC^2 Institute at The University of Texas at Austin, where he leads research in the science of decision support, computational methods, and AI governance. Dr. Kammer-Kerwick has worked with large datasets from the National Institutes of Health to develop computational models that explore the relationship between social determinants of health and a diagnosis of major depressive disorders. **Nicol Turner Lee** holds a PhD in sociology and leads the Center for Technology Innovation and the AI Equity lab at the Brookings Institution. Dr. Lee brings substantial experience and authority in studying the social, ethical, and equity dimensions of AI.

Ruben Rathnasingham, PhD, is the Head of Health Innovation at the Dell Medical School at The University of Texas at Austin, where he focuses on community health and health innovation. Dr. Rathnasingham also leads a team experimenting with the development and use of synthetic data to boost more robust datasets in the development of clinical algorithms that are trained to address skin cancer. **Taylor Wilds** is a Ph.D. candidate in Psychology at The University of Texas at Austin. Her work focuses on maternal health with a particular emphasis on deploying and understanding the application of smart technologies like wearables to develop algorithms that advance the analysis of how maternal-infant interactions impact the health and well-being of mothers and their infants. **Cris Woolston** is the Head of Science Policy at Sanofi, where he is responsible for formulating and coordinating Sanofi's position on, and response to, public policy issues of strategic and operational significance to biomedical innovation, research, and development. He has a PhD in molecular virology. The ACE Your Health AI Task Force Chair, **S. Craig Watkins** is a PhD trained social scientist. As Executive Director of the IC^2 Institute at The University of Texas at Austin, he leads several multidisciplinary research teams in Health AI research that probe how clinical decision-making is evolving in AI-augmented health care environments and the complex relationship between human and artificial intelligence. Dr. Watkins is also working with clinicians and researchers to develop a digital health platform that collects and analyzes clinically relevant and real-time digital biomarkers.

References

Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology, 154*(11), 1247–1248.

American Diabetes Association. (2023). *Statistics about diabetes*. https://diabetes.org/about-diabetes/statistics/about-diabetes

American Society of Clinical Oncology. (2025). Maintaining momentum in cancer health equity: Examining representation in oncology clinical trials. *Journal of Clinical Oncology, 43*(16).

American Urological Association. (2024). Diversity, equity, and inclusion in urolithiasis clinical trials: Representative enrollment by race, ethnicity, and sex. *Urology Practice Journal, 11*(3), 345–352.

Autor, D. (2024, February). *Applying AI to rebuild middle class jobs.* National Bureau of Economic Research Working Paper No. 32140. https://www.nber.org/papers/w32140

Ayanian J.Z., Udvarhelyi I.S., Gatsonis C.A., Pashos C.L., and Epstein A.M. (1993). Racial differences in the use of revascularization procedures after coronary angiography. *Journal of the American Medical Association, 269,* 2642-6.

BMJ Open Research Group. (2025). Embedding equity, diversity and inclusion processes within clinical trials and health and social care research. *BMJ Open, 15*(2).

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*, 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Cary, M. P. Jr, Bessias, S., McCall, J., Pencina, M. J., Grady, S. D., Lytle, K., & Economou-Zavlanos, N. J. (2025). Empowering nurses to champion Health equity & BE FAIR: Bias elimination for fair and responsible AI in healthcare. *Journal of Nursing Scholarship*, 56(4), 445–456.

Cascini, F., Beccia, F., Causio, F. A., Melnyk, A., Zaino, A., & Ricciardi, W. (2022). Scoping review of the current landscape of AI-based applications in clinical trials. *Frontiers in Public Health*, *10*, 949377. https://doi.org/10.3389/fpubh.2022.949377

Centers for Disease Control and Prevention (CDC) (2015). *CDC vital signs: Hispanic health*. Retrieved October 30, 2025, from https://archive.cdc.gov/www_cdc_gov/vitalsigns/hispanic-health/index.html.

Chan, S. R., Wong, J. A., Gribben, V., & Stewart, D. E. (2021). Applications of wearable technology in perinatal mental health: A scoping review. *Archives of Women's Mental Health, 24*(5), 823–835.

Chien, I., Enrique, A., Palacios, J., Regan, T., Keegan, D., Carter, D., Tschiatschek, S., Nori, A., Thieme, A., Richards, D., Doherty, G., & Belgrave, D. (2020). A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Network Open, 3*(7), e2010791.

Daneshjou, R., Vodrahalli, K., Novoa, R. A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., Mukherjee, P., Phung, M., Yekrang, K., Fong, B., Sahasrabudhe, R., Allerup, J. A. C., Okata-Karigane, U., Zou, J., & Chiou, A. S. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances, 8*(32), eabq6147. https://doi.org/10.1126/sciadv.abq6147

Dell Medical School Research Team. (2025). *Advancing ethical, equitable, and trustworthy AI in skin cancer detection: Final report*. Internal research report, The University of Texas at Austin.

Encord AI Research. (2024). Classification metrics in machine learning: Accuracy, precision, and recall optimization for clinical applications. *AI Evaluation Frameworks Quarterly, 3*(2), 45–62.

Fawzy, A., Wu, T. D., Wang, K., & Leaf, D. E. (2022). Racial bias and pulse oximetry accuracy in patients with COVID-19. *JAMA Internal Medicine, 182*(7), 803–805.

Fogg, B.J., Nass, Clifford. 1997. Silicon sycophants: The effects of computers that flatter. *International Journal of Human-Computer Studies, 46*, 551—561.

Gallon, K. (2024, 4 September). *Synthetic data and health equity*. Just Tech. Social Science Research Council. DOI: doi.org/10.35650/JT.3073.d.2024

Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large Language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open*, 7(10), e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969

Goodday, S. M., Atkinson, L., Goodwin, L., Saunders, K. E. A., South, M., & Mackinnon, A. (2021). Understanding perinatal digital health interventions: A systematic review and classification. *Journal of Affective Disorders, 293*, 247–262.

Giuffrè, M., & Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digital Medicine*, *6*(1), 186. https://doi.org/10.1038/s41746-023-00927-3

Groh, M., Harris, C., Soenksen, L., Lau, F., Han, R., Kim, A., Koochek, A., & Badri, O. (2021). Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1820–1828.

Harper, T.A.. (2025, June 6). What happens when people don't understand how AI works. *The Atlantic*. https://www.theatlantic.com/culture/archive/2025/06/artificial-intelligence-illiteracy/683021/

Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsewo, D., Azizi, Z., & White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *NPJ Digital Medicine*, *8*(1), 154. https://doi.org/10.1038/s41746-025-01503-7

Health Affairs Research Consortium. (2023). Mitigating racial and ethnic bias and advancing health equity in clinical AI systems. *Health Affairs, 42*(10), 1445–1454.

IBM Research. (2023). Leveraging machine learning and AI to improve diversity in clinical trials: A comprehensive framework. *IBM Healthcare AI Review, 8*(4), 123–145.

IC^2 Institute, The University of Texas at Austin. (2025). *Exploring the role of human and artificial intelligence in clinical practice.* https://ic2.utexas.edu/research/exploring-the-role-of-human-and-artificial-intelligence-in-clinical-practice/

International Diabetes Federation. (2025). *Facts and figures*. https://idf.org/about-diabetes/diabetes-facts-figures/.

Jin, Y. & Zhang, Y., (2024). OrthoDoc. Multimodal large language model for assisting diagnosis in computed tomography [Internet]. *arXiv:2409.09052* [eess.IV]. https://doi.org/10.48550/arXiv.2409.09052

Kaiser Family Foundation. (2022). State health facts: Mental health care professional shortage areas. https://www.kff.org/other-health/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas

Kaiser Family Foundation. (2024, August 14). Poll: *Most who use artificial intelligence doubt AI chatbots provide accurate health information.* [News Release]. https://www.kff.org/health-information-trust/poll-most-who-use-artificial-intelligence-doubt-ai-chatbots-provide-accurate-health-information/

Kammer-Kerwick, M., Pogue, G., Purohit, I., Spandikow, E., De Jesus, R., Lifton, J., & Watckins, S.C. (2025). Integrating Health ai into safety-net and rural health care settings: Research, insights and recommendations. IC^2 Institute, The University of Texas at Austin. https://doi.org/10.26153/tsw/60608

Kaplan, G.A., Everson, S.A., & Lynch, J.W. (2000). The contribution of social and behavioral research to an understanding of the distribution of disease: A multilevel approach. In B.D. Smedley and S.L. Syme (Eds.), *Promoting health: Intervention strategies from social and behavioral research*. Washington, DC: National Academy Press.

Lakhan, S. (2024, December 23). How AI will transform health equity in clinical trials in 2025. *Clinical Leader, 15*(12), 34–39.

Lawrence, H. R., Schneider, R. A., Rubin, S. B., Matarić, M. J., McDuff, D. J., & Jones Bell, M. (2024). The opportunities and risks of large language models in mental health. *JMIR Mental Health*, *11*, e59479. https://doi.org/10.2196/59479

Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *The New England Journal of Medicine, 388*(13), 1233–1239.

Liu, J., Wang, Y., Tang, X., Li, Z., Li, Y., & Wang, T. (2023). Smart wearable systems for pregnancy and postpartum monitoring: A review. *Biosensors, 13*(1), 88.

Lu, J., Choi, K., Eremeev, M., Gobburu J., Goswami, S., Liu, Q., Mo, G., Musante, C.J., Shahin, M.H. (2025). Large language models and their applications in drug discovery and development: A primer. *Clinical and Translational Science, 18*(4):e70205. https://doi.org/10.1111/cts.70205

Maas, A. and Appelman, Y. (2010). Gender differences in coronary heart disease. *Netherlands Heart Journal*, 18, 598–603.

Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos Jr., J.L., Xiong, C., Sun, Z.Z., Socher, R., Fraser, J.S., & Naik, N. (2023). Large language models generate functional protein sequences across diverse families. *Nature Biotechnology, 41,* 1099–1106. https://doi.org/10.1038/s41587-022-01618-2

Mathur, P., Srivastava. S., Xu, X., & Mehta, J.L.. (2020, Sept. 9). Artificial intelligence, machine learning, and cardiovascular disease. *Clinical Medicine Insights. Cardiology*, *14*, 1179546820927404. https://doi.org/10.1177/1179546820927404

Mozur, P., Satariano, A., Rodriguez Mega, E.. (2025, October 20). From Mexico to Ireland, fury mounts over a global A.I. frenzy. *The New York Times.* https://www.nytimes.com/2025/10/20/technology/ai-data-center-backlash-mexico-ireland.html

Ndugga, N., Hill, L., & Artiga, S. (2024, June 11). *Key data on health and health care by race and ethnicity*. Kaiser Family Foundation. https://www.kff.org/racial-equity-and-health-policy/key-data-on-health-and-health-care-by-race-and-ethnicity/?entry=executive-summary-introduction

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453.

Oertelt-Prigione, S. & Turner, B. (2024). Tackling biases in clinical trials to ensure diverse representation and effective outcomes. *Nature Communications, 15*(1), 1407.

Petersen, E. E., Davis, N. L., Goodman, D., Cox, S., Mayes, N., Johnston, E., Syverson, C., Seed, K., Shapiro-Mendoza, C. K., Callaghan, W. M., & Barfield, W. D. (2019). Racial/ethnic disparities in pregnancy-related deaths—United States, 2007–2016. *MMWR. Morbidity and Mortality Weekly Report, 68*(35), 762–765.

Pew Research Center. (2023, February). *Public awareness of artificial intelligence in everyday activities.* https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/

Phillips, M., Marsden, H., Jaffe, W., Matin, R. N., Wali, G. N., Greenhalgh, J., McGrath, E., James, R., Ladoyanni, E., Bewley, A., Verma, A., Craven, M. P., & Taghipour, K. (2020). Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Network Open*, 3(10).

PLOS Digital Health. (2024). Achieving health equity through conversational AI: A roadmap for design and implementation of inclusive chatbots in healthcare. *PLOS Digital Health*, 3(4). https://doi.org/10.1371/journal.pdig.0000492

PLOS Digital Health Consortium. (2024). Empowering healthcare delivery organizations: Cultivating a community of practice to harness AI and advance health equity. *PLOS Digital Health*, 3(6), e0000513. https://doi.org/10.1371/journal.pdig.0000513

Sarikhani, Y., Najibi, S.M., & Razavi Z. (2024, June 5). Key barriers to the provision and utilization of maternal health services in low-and lower-middle-income countries; a scoping review. *BMC Women's Health,* 24(1), 325.

Shuval, K., Hébert, E. T., Siddiqi, Z., Leonard, T., Lee, S. C., Tiro, J. A., McCallister, K., & Skinner, C. S. (2013). Impediments and facilitators to physical activity and perceptions of sedentary behavior among urban community residents: The Fair Park Study. *Preventing Chronic Disease, 10*, E191.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agüera Y Arcas, B., … Natarajan, V. (2022, December 6). Large language models encode clinical knowledge. arXiv, Retrieved, October 10, 2025, from https://arxiv.org/abs/2212.13138

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion - A study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149,* Article 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Smedley B.D., Stith A.Y., Nelson A.R., (Eds.). (2003.). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. Washington (DC): National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK220358/

Susskind, D. (2020). *A world without work: Technology, automation, and how to respond*. Allen Lane: London.

Synthetic Data Generation Working Group. (2024). Comprehensive framework for synthetic data in health AI: Applications, privacy, and equity implications. *Journal of Medical Internet Research*, 26(8).

Todd, K.H., Deaton C., D'Adamo, A.P., & Goe L. (2000). Ethnicity and analgesic practice. *Annals of Emergency Medicine, 35*(1), 11-16.

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, *20*(3), 318–335. https://doi.org/10.1002/wps.20883

U.S. Bureau of Labor Statistics. (2025, August 28). *Employment projections: 2024-2034 summary*. https://www.bls.gov/news.release/ecopro.nr0.htm

U.S. Food and Drug Administration. (2024). *Guidance for industry: Diversity in clinical trials and AI-enabled medical devices*. FDA Center for Devices and Radiological Health.

VandeHei, J. & Allen, M. (2025, October 29*). Behind the curtain: How an AI job apocalypse unfolds*. Axios. https://www.axios.com/2025/10/29/ai-jobs-apocalypse-navigate

Vargas, D., Dillon, F. R., Gonzalez-Guarda, R., De Santis, J. P., & Gaston, G. B. (2021). A review and discussion of algorithmic bias in maternal health data. *Health Equity, 5*(1), 503–510.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need: Advances in neural information processing systems. *NeurIPS, 30*, 5998–6008.

Building a Healthier Future: Designing AI for Health Equity

ACE Your Health AI Task Force

Powered by

NAACP | sanofi