



CODE

Sharing and Utilizing Health Data for AI Applications

Roundtable Report

Table of Contents

| | |
|--|-----------|
| Executive Summary | 3 |
| Introduction | 4 |
| Background and Key Concepts | 5 |
| Data Requirements for Artificial Intelligence | 5 |
| Supervised and Unsupervised Algorithms | 5 |
| Machine Learning, Deep Learning, and Natural Language Processing | 5 |
| Augmented Intelligence | 6 |
| Stages of AI Development | 6 |
| AI Applications in Healthcare | 7 |
| Reducing Costs and Administrative Burden | 7 |
| Connecting Patients to Resources and Care | 7 |
| Informing Population Health Management | 8 |
| Improving Diagnosis and Early Detection | 8 |
| Developing New Drugs and Therapeutics | 9 |
| Health Data for AI Applications | 10 |
| High-Value Health Data Types | 10 |
| Challenges with Sharing and Utilizing Health Data | 12 |
| Recommendations and Actionable Opportunities | 14 |
| Ensure Access to Data for AI While Protecting Privacy | 15 |
| Use Standards to Improve Data Quality and Interoperability | 16 |
| Remove Administrative Barriers to Data Sharing | 16 |
| Clarify Appropriate Use of Patient-Generated Data | 17 |
| Address Concerns About Accountability and Bias | 17 |
| Conclusion | 19 |
| Acknowledgements and Appendices | 20 |

Executive Summary

The independent nonprofit Center for Open Data Enterprise (CODE) and the Office of the Chief Technology Officer (CTO) at the U.S. Department of Health and Human Services (HHS) are co-hosting a series of three Roundtables to find ways to improve how health data is shared and utilized for the public good.¹

As part of this series, CODE and the HHS Office of the CTO convened a *Roundtable on Sharing and Utilizing Health Data for AI Applications* on April 16, 2019. This Roundtable brought together over 70 expert stakeholders from government, industry, clinical research institutions, nonprofit organizations, and academia to discuss opportunities to share and utilize health data for artificial intelligence (AI) in healthcare. The purpose of this Roundtable was to identify high-priority health applications of AI and key issues for an HHS AI strategy to address. Participants discussed high-value health data types, challenges associated with utilizing health data for AI, and strategic considerations that HHS and other stakeholders should consider as they explore AI development in healthcare.

This report summarizes the findings of the Roundtable in the following sections:

Introduction. This section provides a brief overview of the role of AI in healthcare and the goals of HHS in developing a department-wide AI strategy.

Background and Key Concepts. This section presents an overview of key concepts and terminology related to artificial intelligence.

AI Applications in Healthcare. Building on stakeholder input gathered during the Roundtable and after the event, this section presents a basic typology of AI applications in health.

Health Data for AI Applications. This section presents an overview of high-value data types and challenges associated with utilizing this data in the context of AI.

Recommendations and Actionable Opportunities. This section puts forth recommendations and actionable opportunities for the continued development of AI in healthcare.

Conclusion. Finally, the report concludes by summarizing key findings and presenting relevant updates since the April 2019 Roundtable.

¹ Patient-Centered Outcomes Research Institute, "Open Data Roundtable Series: Sharing and Utilizing Data to Enhance and Protect Health and Well-Being," Retrieved from <https://www.pcori.org/research-results/2018/open-data-roundtable-series-sharing-and-utilizing-data-enhance-and-protect>

Introduction

Artificial intelligence can help transform healthcare by improving diagnosis, treatment, and the delivery of patient care. Researchers in academia, the private sector, and government have gained increasing access to large amounts of health data and high-powered AI-ready computing systems. These powerful tools can greatly improve doctors' abilities to diagnose their patients' medical issues, classify risk at a patient level by drawing on the power of population data, and provide much-needed support to clinics and hospitals in under-resourced areas. AI can also expand the operational capacity of different organizations, identify potentially fraudulent health claims, and streamline manual tasks to boost productivity.

Much of this progress depends on sharing and utilizing large amounts of health data, which informs the development of algorithms and machine learning. While the private sector has driven much of the innovation in this field, the federal government and its partners can play a major role by both sharing their own data and addressing challenges across the sector. HHS, private sector stakeholders, and academic and clinical researchers can support this transformation by collaborating to apply AI both inside and outside of government.

Researchers and practitioners now face multiple challenges in using AI to improve healthcare. These challenges include limited access to data, poor data quality, concerns over data governance, and the ethical use of data, including accountability and liability for data applications. Multiple stakeholders will need to work together to address these challenges as new technical applications emerge.

The HHS Office of the CTO is now exploring the potential for a department-wide AI strategy to help realize the potential of AI, and to establish policies and practices for facilitating AI development. This strategy comes in tandem with the February 2019 "Executive Order on Maintaining American Leadership in Artificial Intelligence"² and *The State of Data Sharing at the U.S. Department of Health and Human Services* report, published by the HHS Office of the CTO in September 2018.³

The *Roundtable on Sharing and Utilizing Health Data for AI Applications* was designed to bring together HHS leaders, and experts in AI and health data from other federal and state government agencies, industry, academia, and patient-centered research organizations. The Roundtable began with calls to action by Mona Siddiqui, the HHS Chief Data Officer, and Ed Simcox, the HHS Chief Technology Officer. Following these keynotes, several speakers gave lightning talks on high-priority use cases for AI in healthcare, including representatives from Verily, Amazon Web Services, Health Catalyst, and the Michael J. Fox Foundation. The second part of the Roundtable transitioned into possibilities for AI strategies, where representatives from Pfizer, the Center for Medicare and Medicaid Innovation (CMMI), the Government Accountability Office (GAO), and the Assistant Secretary for Preparedness and Response at HHS outlined possible paths forward. The day also featured a keynote address from Eric Hargan, the Deputy Secretary of HHS, who spoke about how AI is being deployed within the department.

Throughout the day, Roundtable participants engaged in three in-depth breakout sessions. These sessions focused on the following topics: (1) Identifying high-priority AI applications, (2) Improving and using data for AI applications, and (3) Outlining key issues and objectives for an HHS AI strategy. The day concluded with a presentation of highlights and actionable recommendations for HHS to advance its own AI strategy across the department.

² The White House, "Executive Order on Maintaining American Leadership in Artificial Intelligence," February 11, 2019, Retrieved from

<https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>

³ U.S. Department of Health and Human Services Office of the Chief Technology Officer, "The State of Data Sharing at the U.S. Department of Health and Human Services," September 2018, Retrieved from

https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf

Background and Key Concepts

Artificial intelligence has gained significant attention in recent years, particularly in the context of improving health and well-being. The following section presents an overview of key concepts and terminology related to artificial intelligence. For more information on the use of AI in healthcare, please refer to the briefing paper that CODE developed in preparation for the *Roundtable on Sharing and Utilizing Health Data for AI Applications*.⁴

Data Requirements for Artificial Intelligence

At the core of artificial intelligence is the need for high-quality, clean and accurate data to fuel the development of algorithms. Researchers emphasize the need for large, multifaceted datasets that allow machine learning processes to incorporate as many factors as possible into analysis.⁵ Artificial intelligence also demands clear, accountable data governance with defined data elements and processes for ensuring data quality and access.⁶ Researchers are now attempting to tap large troves of health data - from electronic health records (EHRs) to data collected from wearable devices and sensors - to improve diagnostics and predictive analytics. More connected and interoperable data in the healthcare system will enable more transformative AI applications in the future.

Supervised and Unsupervised Algorithms

Most AI applications depend on algorithms, which describe a logical process that follows a set of rules. Computers can be taught a series of steps in order to process large amounts of data to produce a desired outcome. There are two forms of algorithm:

1. **Supervised algorithms** use 'training datasets' in which the input factors and output are known in advance. Supervised processes can produce highly accurate algorithms because the 'right answers' are already known. For example, scientists may feed a dataset of retina images into the algorithm in which board-certified physicians have already identified and agreed upon diagnoses for each image.
2. **Unsupervised algorithms** are developed through a process whereby data is fed into the algorithm and the computer has to 'learn' what to look for. Unlike the training datasets fed into supervised algorithms, the data fed into unsupervised algorithms does not necessarily include the 'right answers.' Unsupervised algorithms are adept at finding clusters of relationships between observations in the data, but may identify erroneous relationships because they are not instructed what to look for.

Machine Learning, Deep Learning, and Natural Language Processing

Machine learning is the process by which computers are trained to 'learn' by exposing them to data. Machine learning is a subset of AI, and deep learning is a further subset of machine learning. Deep learning is the process by which algorithms can learn to identify hierarchies within data that allow for truly complex understandings of data. Natural language processing (NLP) refers to the subfield of machine learning designed to allow computers to examine, extract, and interpret data that is structured within a language.

Augmented Intelligence

Augmented Intelligence is a form of AI that enhances human capabilities rather than replacing physicians and healthcare providers. Augmented Intelligence has been embraced as a concept by physician organizations to underscore that emerging AI systems are designed to aid humans in clinical decision-making, implementation,

⁴ The Center for Open Data Enterprise, *Briefing Paper: Roundtable on Sharing and Utilizing Health Data for AI Applications*, April 2019, Retrieved from <http://reports.opendataenterprise.org/RT1%20Briefing%20Paper.pdf>

⁵ Nick Ismail, "The Success of Artificial Intelligence Depends on Data," *Information Age*, April 23, 2018. Retrieved from <https://www.information-age.com/success-artificial-intelligence-data-123471607/>.

⁶ Rob Thomas, "The Road to AI Leads through Information Architecture." *VentureBeat*, January 12, 2018. Retrieved from <https://venturebeat.com/2018/01/12/the-road-to-ai-leads-through-information-architecture/>.

and administration to scale healthcare.⁷ In a 2019 white paper, Intel framed augmented intelligence as the AI tools that perform specific tasks and are designed to support users, rather than replacing human experts.⁸

Stages of AI Development

Existing and potential AI applications vary in their level of sophistication, ranging from simple augmentation of common tasks to full automation of systems and processes. Experts have begun categorizing these stages of AI development. Among them, venture capitalist and author Kai-Fu Lee has characterized four “waves” of AI applications:⁹

| | |
|---------------|---------------|
| Wave 1 | Internet AI |
| Wave 2 | Business AI |
| Wave 3 | Perception AI |
| Wave 4 | Autonomous AI |

Figure 1. “The Four Waves of AI”
Adapted from Kai-Fu Lee (2018)

According to Lee, the first wave of AI applications uses data generated on the Internet to better understand the habits, interests, and desires of an individual or population.¹⁰ The second wave of AI applications uses algorithms to inform and improve decision making. Clinical researchers, for example, can construct treatment plans by using algorithms “to digest enormous quantities of data on patient diagnoses, genomic profiles, resultant therapies, and subsequent health outcomes.”¹¹ The third wave of AI applications relates to the proliferation of sensors and devices that collect data about the physical world such as smart watches and virtual assistants.¹² The fourth wave of AI applications integrates all previous waves and gives machines the ability to make decisions without human intervention.¹³ This includes technologies such as automated vehicles.

⁷ACT | The App Association, “Appendix: Key Terminology for AI in Health,” *Connected Health*, Retrieved from <http://actonline.org/wp-content/uploads/Artificial-Intelligence-in-Health-Appendix.pdf>

⁸ Intel, “Intel’s Recommendations for the U.S. National Strategy on Artificial Intelligence,” Retrieved from <https://newsroom.intel.com/wp-content/uploads/sites/11/2019/03/intel-ai-white-paper.pdf>.

⁹ Kai-Fu Lee, “The Four Waves of AI.” *Fortune*, October 22, 2018, Retrieved from <http://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/>

¹⁰ Kai-Fu Lee, “The Four Waves of AI.” *Fortune*, October 22, 2018, Retrieved from <http://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/>

¹¹ Kai-Fu Lee, “The Four Waves of AI.” *Fortune*, October 22, 2018, Retrieved from <http://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/>

¹² Kai-Fu Lee, “The Four Waves of AI.” *Fortune*, October 22, 2018, Retrieved from <http://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/>

¹³ Kai-Fu Lee, “The Four Waves of AI.” *Fortune*, October 22, 2018, Retrieved from <http://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/>

AI Applications in Healthcare

The following section outlines five general uses of AI in healthcare, including examples of existing and near-term applications. These applications, while not mutually exclusive, are examples of cross-cutting themes that emerged from the Roundtable discussions.

Reducing Costs and Administrative Burden

Roundtable participants emphasized the value of AI in improving clinician and administrative workflows. Through NLP and other AI tools, machines can rapidly process EHRs and automatically transcribe medical notes. Automation can free up time and reduce costs by eliminating manual data entry. Moreover, participants noted that AI helps reduce administrative burden by correcting human errors in billing processes.

- **Interpreting handwritten medical records.** Amazon Web Services has used NLP to extract and interpret handwritten notes and text from medical records.¹⁴ NLP is particularly well-suited to deciphering physician input since EHRs do not follow a single, unified structure, yet contain important information for understanding diagnostic trends and risk profiles of individuals.¹⁵
- **Detecting fraud and improper payments.** The Centers for Medicare and Medicaid Services (CMS) uses statistical analysis to identify fraudulent and improper payments made to healthcare providers. In 2018, CMS determined that 8.12 percent of all Medicare payments were improper. In order to address this problem, CMS employs a testing methodology called Comprehensive Error Rate Testing (CERT) and uses AI to engage in predictive analysis of fraudulent and improper healthcare payments. This process has saved the government approximately \$42 billion, according to CMS.¹⁶

Connecting Patients to Resources and Care

Roundtable participants emphasized the value of using AI to connect patients with available resources and care, especially in rural areas. Examples include:

- **Providing patients with personalized healthcare recommendations.** Sage Bionetworks launched mPower as a study using surveys and phone sensors to track symptoms of Parkinson's Disease.¹⁷ The results can help patients, doctors, and caregivers better understand changes over time and the impact of exercise or medication. Using artificial intelligence, data from mPower could also be used to develop specific healthcare recommendations for patients.
- **Creating virtual care programs for patients with chronic conditions.** Verily Health's Onduo project, which combines a smart device and mobile application, offers virtual care for people with type 2 diabetes. Onduo can measure blood glucose levels as well as provide information on nutrition and medication management. The app also offers a coaching dimension that identifies lifestyle patterns and gives patients feedback to improve their health.

¹⁴ Rachel Arndt, "Amazon Technology Deciphers Text in Electronic Health Record," *Modern Healthcare*, November 27, 2018, Retrieved from

<https://www.modernhealthcare.com/article/20181127/TRANSFORMATION01/181129951/amazon-technology-deciphers-text-in-electronic-health-record>.

¹⁵ Mike Miliard, "EHR Natural Language Processing Isn't Perfect, but It's Really Useful," *Healthcare IT News*, May 18, 2017. Retrieved from <https://www.healthcareitnews.com/news/ehr-natural-language-processing-isnt-perfect-its-really-useful>.

¹⁶ Jack McCarthy, "CMS Snags \$42 Billion in Medicare and Medicaid Fraud with Predictive Analytics," *Healthcare IT News*, July 22, 2016, Retrieved from <https://www.healthcareitnews.com/news/cms-snags-42-billion-medicare-and-medicaid-fraud-predictive-analytics>.

¹⁷ Sage Bionetworks, "mPower Parkinson Study and Public Researcher Portal," Retrieved from <http://sagebionetworks.org/research-projects/mpower-researcher-portal/>

- **Expanding treatment access for rural populations.** Through voice assistants and chatbots, AI has the potential to improve and increase access to treatment in rural and other resource-constrained environments. There is increasing evidence that AI-driven chatbots can address routine patient questions and help doctors communicate with patients about their diagnosis and risk evaluations.¹⁸

Informing Population Health Management

Population health management (PHM) involves using population-level data to identify broad health risks and treatment opportunities for a group of individuals or community. AI can contribute to PHM by combining, synthesizing, and analyzing datasets from third parties with clinical or patient-generated data. For example, researchers and health providers can use AI to aggregate longitudinal patient-generated data into larger datasets that tell better stories about the incidence and prevalence of disease.

- **Identifying at-risk populations.** AI can be used to identify populations at risk for opioid abuse or overdose. One population health management company, for example, integrates data on social determinants of health and pharmacy claims to better understand the diverse “spectrum of opioid abuse cases.”¹⁹

Improving Diagnosis and Early Detection

Diagnostic errors are a major problem in the healthcare system, with most patients experiencing at least one diagnostic error in their lifetime.²⁰ AI promises to help physicians accurately diagnose medical conditions in their patients and treat disease at an early stage. AI algorithms draw upon large datasets on medical and social determinants of health to better identify patterns and assist physicians in making diagnoses and developing treatment plans.²¹

AI can deploy technologies like image recognition, NLP, and deep learning to quickly detect life-threatening conditions and assess risk for diseases like brain cancer or heart disease. Roundtable participants noted that it may be more accurate to think of these applications as “augmented intelligence” rather than artificial intelligence. The goal is not to replace the doctor’s clinical judgment, but to help physicians rapidly prioritize patient symptoms and assess a range of diagnostic possibilities rather than ask patients a standard slate of questions. Examples from the Roundtable include:

- **Diagnosing diabetic retinopathy through image recognition.** AI can help doctors diagnose diabetic retinopathy, the world’s leading cause of blindness, by using image recognition. Researchers at Google have trained algorithms to analyze images of retinas and diagnose this disease with over 90 percent accuracy.²²

¹⁸ Rachel Arndt, “Healthcare providers are teaming with chatbots to assist patients,” *Modern Healthcare*, December 8, 2018, Retrieved from <https://www.modernhealthcare.com/article/20181208/TRANSFORMATION01/181209977/healthcare-providers-are-teaming-with-chatbots-to-assist-patients>

¹⁹ ZeOmega, “ZeOmega Launches Jiva Opioid AI to Stem Drug Abuse Crisis,” May 28, 2019, Retrieved from <http://www.zeomega.com/zeomega-media-center/press-releases/2019-press-releases/zeomega-launches-jiva-opioid-ai-to-stem-drug-abuse-crisis/>

²⁰ The National Academies of Sciences, Engineering, and Medicine (2015) *Improving Diagnosis in Health Care*, Washington, DC: The National Academies Press, Retrieved from <https://doi.org/10.17226/21794>

²¹ Laura Landro, “The Key to Reducing Doctors’ Misdiagnoses,” *The Wall Street Journal*, September 13, 2017, Retrieved from <https://www.wsj.com/articles/the-key-to-reducing-doctors-misdiagnoses-1505226691>.

²² Rohit Varma, “Using Artificial Intelligence to Automate Screening for Diabetic Retinopathy,” *Ophthalmology Times*, October 22, 2018, Retrieved from <https://www.opthalmologytimes.com/article/using-artificial-intelligence-automate-screening-diabetic-retinopathy>.

- **Predicting brain deterioration using advanced machine learning.** AI can be used to analyze a diverse array of datasets and identify potential biomarkers that can indicate the onset of deterioration in cases that range from concussion to coma.²³

Developing New Drugs and Therapeutics

Drug development is a costly and time-consuming process. AI can help improve drug development through the entire development lifecycle, from identifying gaps in current therapeutics to bringing new products to market. Pharmaceutical researchers can use AI to sort through huge numbers of research papers and patents, as well as comprehensive lists of chemical compounds and their properties, to suggest opportunities for drug development. By analyzing the growing databases of biomarker data, they can then work to target different treatments to different types of patients. And when drugs or other treatments reach the clinical trial stage, AI can help match ideal patients to the right trials. Examples include:

- **Improving clinical trial participation.** HHS recently completed a “tech sprint” engaging external experts, such as TrialX and Intel, to develop AI applications to match patients to appropriate clinical trials. This kind of matching can help researchers find appropriate subjects for their studies and help patients find potentially valuable treatments at the same time.²⁴
- **Supporting precision medicine.** The National Institutes of Health (NIH) defines precision medicine as “an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment and lifestyle for each person.”²⁵ Researchers at startups like Lam Therapeutics and Lantern Pharma are using supervised machine learning strategies to generate new correlations between genomic biomarkers and drug activity to pilot individualized cancer treatments.

²³ Suzanne Leigh, “Artificial Intelligence Aids Scientists in Uncovering Hallmarks of Mystery Concussion,” *University of California San Francisco Blog*. Retrieved from <https://www.ucsf.edu/news/2017/03/405926/artificial-intelligence-aids-scientists-uncovering-hallmarks-mystery-concussion>.

²⁴ Gil Alterovitz and Kristen Honey, “TOP Health’ Tech Sprint Unleashes the Power of Open Data and AI,” *U.S. Department of Health and Human Services Office of the Chief Technology Officer Blog*, January 17, 2019, Retrieved from <https://www.hhs.gov/cto/blog/2019/1/17/top-health-tech-sprint-unleashes-the-power-of-open-data-and-ai.html>.

²⁵ National Institutes of Health, “All of Us Research Program,” Retrieved from <https://allofus.nih.gov/>

Health Data for AI Applications

Data is the foundation of all AI applications. During the Roundtable, participants identified a number of high-value health data types that can be used for AI development. Building on the expert feedback gathered at the Roundtable and subsequent research, this section provides a summary of six major health data types and the challenges associated with their use.

High-Value Health Data Types

| |
|---|
| Administrative and Claims Data |
| Clinical Data |
| Clinical Trials Data |
| EHR Data |
| Genomic Data |
| Patient-Generated Data |
| IoT Data |
| Social Media Data |
| Social Determinants of Health Data |
| Surveillance Data |
| Registry Data |
| Survey Data |
| Vitals Data |

Administrative and Claims Data generally comes from federal, state, and local government agencies as well as healthcare providers and insurers. This can range from hospital discharge summaries to payment records between insured patients and the healthcare system.²⁶

Clinical Data is a broad term that encompasses different kinds of data generated “in a clinical setting and controlled by a clinician, as opposed to a patient or caregiver.”²⁷

- **Clinical Trials Data** includes registries and results from publicly and privately funded clinical studies. Large amounts of data, including sensitive information about participants, are generated over the course of a clinical trial. Researchers must obtain regulatory approval to collect and use this data.
- **EHR Data** is focused on individual patients, and can include information on routine checkups, prescriptions, and medical procedures. Physicians can draw upon EHR data to develop individual treatment plans and diagnose conditions. This data can also be combined with social determinants of health to develop rich longitudinal profiles of individual patients and populations.

Genomic Data can include many different characteristics, ranging from full DNA sequences to individual DNA variants.²⁸ Recent advances have made it possible to analyze and store data on a person's entire genome sequence. According to the National Institutes of Health, “Genome-based research is already enabling medical researchers to develop improved diagnostics, more effective therapeutic strategies, evidence-based approaches for demonstrating clinical efficacy, and better decision-making tools for patients and providers.”²⁹ Genomic data is considered highly sensitive and must be shared and used under carefully controlled conditions.

Patient-Generated Data includes “health-related data created and recorded by or from patients outside of the clinical setting to help address a health concern.”³⁰ This data type is becoming increasingly prevalent through the creation of mobile health applications and wearable health devices.

- **IoT Data** includes data from mobile software applications, voice assistants, and wearable devices such as smart watches. These technologies are part of the “internet of things,” or IoT, which refers to the growing system of machines and devices connected to the internet. This data is generally collected under “terms of service” agreements and has the potential to provide important information on a variety of critical health indicators, such as heart rate, sleep cycles, and diet.
- **Social Media Data** includes interactions on social media platforms such as Facebook and Twitter. Researchers have noted that “Social media may offer insight into the relationship between an individual's health and their everyday life, as well as attitudes towards health and the perceived quality of healthcare services,” among other opportunities.³¹ Like IoT data, social media data is collected under “terms of service” agreements.

²⁶ University of Washington Health Sciences Library, “Data Resources in the Health Sciences,” Retrieved from . <http://guides.lib.uw.edu/hsl/data/findclin>

²⁷ Office of the National Coordinator for Health Information Technology, *Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024*, January 2018, Retrieved from https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf

²⁸ PHG Foundation at the University of Cambridge, *Identification and genomic data*, December 2017, Retrieved from <http://www.phgfoundation.org/documents/PHGF-Identification-and-genomic-data.pdf>

²⁹ National Institutes of Health National Human Genome Research Institute, “A Brief Guide to Genomics,” Retrieved from <https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics>

³⁰ Office of the National Coordinator for Health Information Technology, *Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024*, January 2018, Retrieved from https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf

³¹ Kevin Padrez *et al*, “Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department,” *BMJ Quality & Safety*, 2016, Retrieved from <http://dx.doi.org/10.1136/bmjqs-2015-004489>

Social Determinants of Health Data represent “conditions in the environments in which people are born, live, learn, [and] work...that affect a wide range of health, functioning, and quality-of-life outcomes and risks.”³² Examples of these social determinants include access to transportation, education, and job opportunities as well as the availability of food and housing options. Social determinants of health data can come from many sources inside and outside of government, and can be used to better understand population health.

Surveillance Data is a broad term that encompasses the “ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice.”³³

- **Registry Data** includes data shared voluntarily by individuals that is generally focused around a specific diagnosis or condition such as cancer or cystic fibrosis. This data can be used to track trends and better understand conditions over time. According to the NIH, this data “belongs to the sponsor of the registry and...may be shared with the participants and their families, and approved health care professionals and researchers. However, personal, identifying information is kept private.”³⁴
- **Survey Data** includes the results of surveys and studies conducted to assess population health. This data can help stakeholders monitor the spread of disease, track health insurance coverage across regions, and assess trends in nutrition and exercise, among other uses.³⁵
- **Vitals Data** is generally collected and exchanged between local jurisdictions and the federal government. This data represents “vital events,” such as births, deaths, marriages, divorces, and fetal deaths.³⁶

Challenges with Sharing and Utilizing Health Data

Roundtable participants identified numerous legal, cultural, and technical challenges associated with sharing and utilizing health data for AI applications. While some of these challenges are specific to AI development, others are general issues that impact all applications of health data.

Legal challenges

- **Inconsistent restrictions on data use.** Among the legal challenges, participants noted that health data types have different legal and regulatory constraints on their use. For example, administrative and claims data, clinical data, and certain types of surveillance data, such as survey data, can include sensitive, individual-level information.³⁷ The use of these data types is often restricted under existing privacy frameworks such as HIPAA. Patient-generated data, such as data collected from mobile applications and wearable devices, can also contain sensitive information about individuals ranging from fertility treatments to mental health conditions. However, there are relatively few legal guidelines that protect this emerging data type from misuse.

³² Office of Disease Prevention, and Health Promotion, “Social Determinants of Health,” Retrieved from <https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health>

³³ World Health Organization, “Public health surveillance,” Retrieved from https://www.who.int/topics/public_health_surveillance/en/

³⁴ National Institutes of Health, “List of Registries,” Retrieved from <https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries>

³⁵ National Center for Health Statistics, “Resources for Survey Participants,” Retrieved from https://www.cdc.gov/nchs/nchs_for_you/survey_participants.htm

³⁶ National Center for Health Statistics, “National Vital Statistics System,” Retrieved from <https://www.cdc.gov/nchs/nvss/index.htm>

³⁷ Amy Bernstein and Marie Haring Sweeney, Marie Haring, “Public Health Surveillance Data: Legal, Policy, Ethical, Regulatory, and Practical Issues,” *Morbidity and Mortality Weekly Report*. Retrieved from <https://www.cdc.gov/mmwr/preview/mmwrhtml/su6103a7.htm>

- **Concerns about intellectual property.** Roundtable participants also discussed the challenges of using and sharing proprietary data and algorithms. Data collected in drug development trials, through private-sector health surveys, or in other ways could benefit researchers and organizations in the health sector developing AI applications, and proprietary AI models could be developed for greater accuracy if the algorithms they use were shared. But while all parties stand to benefit from sharing data and algorithms, it is difficult to balance that benefit against companies' need to protect their intellectual property for competitive advantage.

Cultural challenges

- **Underlying bias in health data.** Some Roundtable participants highlighted concerns about bias and lack of diversity in health data, which can have serious consequences when utilized for AI development. As one expert notes, "If the data are flawed, missing pieces, or don't accurately represent a population of patients, then any algorithm relying on the data is at a higher risk of making a mistake."³⁸
- **Data silos and administrative hurdles.** While HHS is developing more efficient ways for its operating agencies to share data - for example, by developing common data use agreements (DUAs) - it is still difficult for HHS agencies to share data with each other, and can be difficult for organizations outside of government to obtain data from HHS. Roundtable participants said that it can take 12 to 18 months to get access to data from various agencies and offices within HHS. Culture changes are needed to reduce the administrative hurdles that prevent timely data sharing.
- **Overly restrictive interpretations of HIPAA.** Some Roundtable participants noted that fears about violating HIPAA have created a risk-averse environment for data sharing. While HIPAA is intended to protect patient privacy, it does allow data sharing and use under specific conditions.³⁹ Participants suggested that HHS could provide more guidance on what is and is not permissible under HIPAA in different contexts.

Technical challenges

- **Limited technical capacity for data management and analysis.** Roundtable participants inside and outside of government noted the need for more staff with data science training. In particular, both government and the private sector need more experts in AI and its application to health data and issues.
- **Inadequate IT infrastructure for hosting and analyzing large datasets.** AI applications require large quantities of data, and large computational capacity, to train and test algorithms. The increasing demand for real-time data adds to these technical requirements. Both HHS and the stakeholders that work with the department may need to upgrade their infrastructure to meet these challenges.
- **Poor data interoperability.** Roundtable participants flagged a number of challenges related to joining and combining health datasets. Across the healthcare system, large amounts of data are structured in different ways, preventing stakeholders from easily exchanging and integrating this information. Participants attributed these challenges to a lack of common data standards and issues with enforcement where standards do exist.

³⁸ Dave Gershorn, "If AI is going to be the world's doctor, it needs better textbooks," Quartz, September 6, 2018. Retrieved from <https://qz.com/author/dgershgornqz/>

³⁹ Office of the National Coordinator for Health Information Technology, "How HIPAA Supports Data Sharing," Retrieved from <https://www.healthit.gov/topic/interoperability/how-hipaa-supports-data-sharing>

Recommendations and Actionable Opportunities

The February 2019 “Executive Order on Maintaining American Leadership in Artificial Intelligence” outlined a number of strategic objectives for developing AI.⁴⁰ These include (text bolded for emphasis):

Executive Order Objectives

- A. “Promote sustained investment in AI R&D in collaboration with industry, academia, international partners and allies,** and other non-federal entities to generate technological breakthroughs in AI and related technologies and to rapidly transition those breakthroughs into capabilities that contribute to our economic and national security.
- B. Enhance access to high-quality and fully traceable federal data, models, and computing resources** to increase the value of such resources for AI R&D, **while maintaining safety, security, privacy, and confidentiality** protections consistent with applicable laws and policies.
- C. Reduce barriers to the use of AI technologies** to promote their innovative application while protecting American technology, economic and national security, civil liberties, privacy, and values.
- D. Ensure that technical standards minimize vulnerability** to attacks from malicious actors and reflect federal priorities for innovation, public trust, and public confidence in systems that use AI technologies; and develop international standards to promote and protect those priorities.
- E. Train the next generation of American AI researchers and users** through apprenticeships; skills programs; and education in science, technology, engineering, and mathematics (STEM), with an emphasis on computer science, to ensure that American workers, including federal workers, are capable of taking full advantage of the opportunities of AI.
- F. Develop and implement an action plan,** in accordance with the National Security Presidential Memorandum of February 11, 2019 (Protecting the United States Advantage in Artificial Intelligence and Related Critical Technologies) (the NSPM) to protect the advantage of the United States in AI and technology critical to United States economic and national security interests against strategic competitors and foreign adversaries.”

Under the Executive Order, agencies funding and deploying AI are expected to use these government-wide objectives to inform their work.

⁴⁰ The White House, “Executive Order on Maintaining American Leadership in Artificial Intelligence,” February 11, 2019, Retrieved from <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>

The HHS Office of the CTO is exploring the potential for a department-wide AI strategy to help realize the value of AI within government, and to establish policies and practices for facilitating its development across the health sector. Over the course of the Roundtable, participants outlined a number of recommendations for HHS and other stakeholders that align with the objectives of the Executive Order. These recommendations and related actionable opportunities are summarized below:

Invest in IT Infrastructure and Expertise to Support AI

AI demands a robust information technology (IT) infrastructure, including data infrastructure, and staff with the skills to apply it. Both infrastructure and expertise must be able to manage the large amounts of data needed to support AI as well as the development of advanced AI applications.

Actionable Opportunities:

- **Develop comprehensive technology investment plans** to support organizational AI strategies. Within HHS, this can include improving legacy systems for managing the data that will fuel AI applications as well as IT modernization overall. It may also include public-private collaboration to reduce the cost to government of technical improvements.
- **Build expertise in designing and implementing AI applications.** Most federal departments, including HHS, have limited organizational knowledge and experience needed to develop AI applications with their data. HHS can bridge this gap through hiring programs, public-private collaborations, or fellowship programs to bring AI experts into government on a temporary basis.
- **Create national testbeds for AI development.** Industry has led the development of AI applications, since commercial companies collect massive amounts of data and have the resources, expertise, and technical capacity to apply it. HHS and other agencies can help remove these barriers to entry by creating collaborative environments where data and code for AI applications can be tested, stored, and shared.

Ensure Access to Data for AI While Protecting Privacy

Concerns about privacy are paramount in the application of individual health data. While the use of health data in EHRs and other medical records is governed by federal and state legislation, other data types like IoT data are only regulated through “terms of service” agreements developed by the private sector. HHS and its partners will need to ensure that sensitive information is not disclosed or misused when these data sources are applied. At the same time, researchers need to be able to use sensitive data appropriately to develop new insights, diagnostic methods, and treatments. (The challenge of balancing privacy with health data access will be the subject of the next Roundtable in this series.)

Actionable Opportunities:

- **Provide guidance for de-identifying sensitive data.** In order to protect privacy, health data can be de-identified in different ways before researchers analyze it. Data scientists, for example, have utilized codes that make it possible to link data about an individual from different sources without revealing the person’s identity. HHS could provide additional guidance on de-identification methods to protect data privacy and security while encouraging its use for AI applications. This guidance may include updating HIPAA’s rules around de-identification to meet modern demands.

- **Develop credentialing systems for controlled access to sensitive health data.** Some sensitive data maintained by the federal government, such as collections of genomic data, are now available only to qualified researchers only, under agreements that prohibit them from sharing the data more widely. HHS could apply this model more broadly and develop credentialing systems to determine who should have access to what kinds of data and under what conditions.

Use Standards to Improve Data Quality and Interoperability

Data for AI applications should be clean, timely, accurate, and standardized. Roundtable participants identified numerous challenges related to integrating data and metadata from multiple sources. Common standards for data collection and management can ensure that data and metadata are accurate and consistent across healthcare applications, using a shared library of variables that are applied across datasets. Standardization also ensures that datasets will be interoperable between agencies within HHS or between HHS and its external partners. The current lack of interoperability is a major obstacle to applying data for AI development.

Actionable Opportunity:

- **Adopt and expand existing common data models.** Many participants noted the value of adopting existing common data models for data and metadata, wherever possible. Common data models standardize the way information is structured and make it easier to use in combination with other data. Examples of existing common data models include Patient Centered Outcome Research Network (PCORNet) model and the Observational Medical Outcomes Partnership (OMOP) model. Participants also mentioned ICD-10, which is a widely used coding system that could be expanded to improve health data quality and interoperability nationwide.

Remove Administrative Barriers to Data Sharing

AI applications are most effective when they can integrate large amounts of data about diverse facets of health. However, researchers inside and outside of HHS often have difficulty accessing the data they need. To share data from other sources, researchers must have DUAs that abide by HIPAA regulations, whether HHS is sharing data with outside researchers or whether different operating divisions within HHS are sharing data with each other. Drawing up and approving separate DUAs can take time and administrative resources that are a burden on researchers and slow down the research process.

Actionable Opportunity:

- **Update and standardize data use agreements.** A set of standard DUAs, using common terms and conditions, could accelerate and simplify data sharing between operating agencies within HHS. Revised DUAs could substantially reduce the time it takes for HHS researchers to request and receive important, time-sensitive data: Finalizing DUAs can now take up to 12 months. Standard DUAs for internal use within HHS could also become a model for agreements between HHS and outside partners.

Participants also highlighted two areas for action that go beyond the Executive Order:

Clarify Appropriate Use of Patient-Generated Data

Increasingly, patients are generating data about themselves that can complement research and clinical data.⁴¹ Patient-generated data includes data collected through sensors and wearables, and through social media and mobile applications. Large amounts of this data are collected under “terms of service” agreements and are being used by entities that are not covered by HIPAA.⁴² As interest in patient-generated data increases, there is a need for clearer rules around its appropriate use, particularly in the context of AI development.

Actionable Opportunities:

- **Develop specific guidelines for entities not covered by HIPAA.** HIPAA applies to traditional entities, such as health plans and healthcare providers, but does not apply to software development and social media companies that may be collecting patient-generated data with sensitive health information.⁴³ While the HHS Office for Civil Rights has developed several informational resources for health app developers,⁴⁴ entities that are not covered by HIPAA, and the individuals whose data they collect, would benefit from further guidance and best practices on appropriate uses of patient-generated data.

Address Concerns About Accountability and Bias

Many AI applications that use health data are being developed as a “black box” without clear information about the algorithms and data being used to make decisions. AI strategies should include steps to address concerns about accountability, bias, and oversight. This will require improved transparency for both AI algorithms and the data that supports them.

Actionable Opportunities:

- **Develop guidelines for mitigating bias in health-related AI applications.** Some Roundtable participants expressed interest in having HHS and its partners develop guidance for identifying and reducing bias in AI applications. Participants also suggested including an internal HHS review function to enforce such guidelines and help increase transparency.
- **Pilot and implement an FDA regulation for health-related AI applications.** Roundtable participants expressed similar concerns about a lack of quality assurance and oversight for AI development in healthcare. The Food and Drug Administration

⁴¹ U.S. Department of Health and Human Services, “Patient-Generated Health Data,” Retrieved from <https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data>

⁴² U.S. Department of Health and Human Services, *Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA*, June 2016, Retrieved from https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf

⁴³ U.S. Department of Health and Human Services, *Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA*, June 2016, Retrieved from https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf

⁴⁴ U.S. Department of Health and Human Services Office for Civil Rights, “Health app developers, what are your questions about HIPAA?” Retrieved from <https://hipaaqportal.hhs.gov/>

(FDA) has established a set of iterative, agile guidelines to precertify the rapid development of Software as a Medical Device (SaMD).⁴⁵ The FDA should continue its efforts to adopt a revised regulatory framework for AI applications in which proposed changes to algorithms must be disclosed to the FDA prior to market release.⁴⁶ This framework should take into account the ability of AI applications to adopt in real time and provide ways to assess any risks from those changes.

- **Publish metadata about data sources.** Metadata provides information about the structure of a dataset, the meaning of each variable within the data, the method of collection, and other important characteristics. Metadata can also provide information about the source of the data, the way it was collected, and other factors that may indicate potential causes of bias. Publishing metadata will make it easier to assess whether the data and the algorithms it supports are at risk of being biased in any way.

⁴⁵ U.S. Food and Drug Administration, “Developing Software Precertification Program: A Working Model,” June 2018, Retrieved from

<https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM611103.pdf>

⁴⁶ U.S. Food and Drug Administration, “Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device,” April 2019, Retrieved from <https://www.fda.gov/media/122535/download>

Conclusion

While the promise of AI in healthcare is significant, a number of challenges can impede its successful implementation. The *Roundtable on Sharing and Utilizing Health Data for AI Applications* was a first step to finding solutions by identifying innovative examples of AI applications, high-value data types, and ways that all stakeholders can contribute to the successful and appropriate use of AI.

In the two months since the Roundtable, HHS has demonstrated its commitment to exploring the use of AI inside and outside of government. For example, HHS is moving forward with its “Reimagined — Buy Smarter” program designed to use AI to conduct strategic comparative analysis of industry pricing to ensure that HHS is saving taxpayers as much money as possible.⁴⁷ HHS and CMS are also working to expand their cloud capacity to manage the growing data assets that are critical to their daily operations and future AI applications.⁴⁸

The Executive Office of the President has also advanced government-wide AI initiatives. In addition to the February 2019 “Executive Order on Maintaining American Leadership in Artificial Intelligence,” the National Science and Technology Council updated *The National Artificial Intelligence Research and Development Strategic Plan* in June 2019.⁴⁹ The Plan recommends that the federal government develop a coordinated approach to maximize the impact of AI technology as it grows in scope. The Plan also proposes eight different strategies to bolster AI development such as understanding the ethical, legal, and societal implications of AI, adopting effective strategies for AI-human collaboration, and supporting the safety and security of AI systems.

This summary report presents research and diverse stakeholder input from the *Roundtable on Sharing and Utilizing Health Data for AI Applications* that can inform the development of an HHS AI strategy. The report outlines a number of ways that HHS can take action that align with the Executive Order and other government-wide AI initiatives.

The same kinds of recommendations and actionable opportunities may be useful to the growing number of stakeholders outside of government who are working to develop applications based on health data. Private-sector companies, patients and their advocates, academic researchers, healthcare providers, and other stakeholders will all play critical roles in the development of health-related AI in the months and years ahead. CODE hopes that this report will provide context, perspective, and the beginnings of a framework for the important work to come.

⁴⁷ Amelia Brust, “HHS wants to use AI to help it ‘Buy Smarter,’” *Federal News Network*, May 17, 2019, Retrieved from <https://federalnewsnetwork.com/ai-machine-learning-month/2019/05/hhs-wants-to-use-ai-to-help-it-buy-smarter/>

⁴⁸ Joseph Goedert, “CMS works with MarkLogic to expand cloud platforms,” *Health Data Management*, April 30, 2019, Retrieved from <https://www.healthdatamanagement.com/news/cms-works-with-marklogic-to-expand-cloud-platforms>

⁴⁹ The National Science and Technology Council, *The National Artificial Intelligence Research and Development Strategic Plan*, June 21, 2019, Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2019/06/National-AI-Research-and-Development-Strategic-Plan-2019-Update-June-2019.pdf>

Acknowledgements and Appendices

The *Roundtable on Sharing and Utilizing Health Data for AI Applications* was funded through a Patient-Centered Outcomes Research Institute (PCORI) Engagement Award Initiative (12667-CODE). This Roundtable is part of the *Open Data Roundtable Series: Sharing and Utilizing Data to Enhance and Protect Health and Well-Being* funded through this award.



CODE would like to thank the HHS Office of the Chief Technology Officer for their partnership in co-hosting this Roundtable series. We also thank the Multi-Stakeholder Advisory Committee for this series:

Lisa Bari, CMS Innovation Center, Centers for Medicare and Medicaid Services
Sohini Chowdhury, Michael J. Fox Foundation
James Craver, National Center for Health Statistics, Centers for Disease Control and Prevention
Gwen Darien, National Patient Advocate Foundation
Stephanie Devaney, *All of Us* Research Program, National Institutes of Health
Natalie Evans-Harris, BrightHive
Jason Gerson, Patient-Centered Outcomes Research Institute
Joel Gurin, Center for Open Data Enterprise
William Hoffman, World Economic Forum
Charles Keckler, Associate Deputy Secretary, HHS
Lisa Khorey, Allscripts Healthcare Solutions
Michael Seres, 11 Health
Mona Siddiqui, Chief Data Officer, HHS
Paul Tarini, Robert Wood Johnson Foundation
John Wilbanks, Sage Bionetworks

The following resources are included as appendices to this report:

- List of Participating Organizations
- Roundtable Agenda

Roundtable on Sharing and Utilizing Health Data for AI Applications

PARTICIPATING ORGANIZATIONS

Academia

Dell Medical School, University of Texas at Austin is the graduate medical school of The University of Texas at Austin in Austin, Texas. The school opened to the inaugural class of 50 students in the summer of 2016 as the newest of 18 colleges and schools on the UT Austin campus.

Duke Margolis Center for Health Policy's mission is to improve health and the value of health care through practical, innovative, and evidence-based policy solutions. Duke-Margolis catalyzes Duke University's leading capabilities to inform policy making and implementation for better health and health care.

Harvard Law School, Petrie-Flom Center founding mission is to promote interdisciplinary analysis and legal scholarship in the fields of Health Law Policy, Biotechnology, and Bioethics.

University of Maryland School of Public Health offers an unusual breadth of expertise to explore complex issues through public health disciplines, and lenses as varied as engineering, arts and humanities, business and public policy.

Stanford University, one of the world's leading teaching and research universities, dedicated to finding solutions to big challenges and to preparing students for leadership in a complex world.

Hall Center for Law and Health, Indiana University Robert H. McKinney School of Law was established in 1987 to expand the curriculum and teaching of health law and provide opportunities for students.

Civil Society

ACT | The App Association represents more than 5,000 app companies and information technology firms across the mobile economy. ACT advocates for an environment that inspires and rewards innovation, while providing the necessary resources to help its members leverage their intellectual assets to raise capital and create jobs.

American Medical Association (AMC) promotes the art and science of medicine and the betterment of public health. AMC provides timely, essential resources to empower physicians, residents and medical students to succeed at every phase of their medical lives.

Center for Data Innovation educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things.

Center for Open Data Enterprise (CODE) is an independent nonprofit organization based in Washington, D.C. whose mission is to maximize the value of open government data for the public good.

Healthcare Leadership Council (HLC), a coalition of chief executives from all disciplines within American healthcare, is the exclusive forum for the nation's healthcare leaders to jointly develop policies, plans, and programs to achieve their vision of a 21st century system that makes affordable, high-quality care accessible to all Americans.

World Economic Forum is the International Organization for Public-Private Cooperation. The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

Government

The U.S. Department of Health and Human Services is a cabinet-level department of the U.S. federal government with the goal of protecting the health of all Americans and providing essential human services.

The Agency for Healthcare Research and Quality's (AHRQ) mission is to produce evidence to make health care safer, higher quality, more accessible, equitable, and affordable, and to work within the U.S. Department of Health and Human Services and with other partners to make sure that the evidence is understood and used.

The Centers for Disease Control and Prevention Center (CDC) works to protect America from health, safety and security threats, both foreign and in the U.S. Whether diseases start at home or abroad, are chronic or acute, curable or preventable, human error or deliberate attack, CDC fights disease and supports communities and citizens to do the same.

The National Center for Health Statistics (NCHS), part of the CDC, compiles statistical information to help guide policies to improve the health of Americans. Holds a biennial data user conference; consult the NCHS website for date and location. NCHS disseminates data and statistics online and in print.

Center for Medicare and Medicaid Services (CMS) administers the Medicare program and works in partnership with state governments to administer Medicaid, the Children's Health Insurance Program (CHIP), and health insurance portability standards.

The Innovation Center with CMS supports the development and testing of innovative health care payment and service delivery models.

Food and Drug Administration (FDA) is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices; and by ensuring the safety of our nation's food supply, cosmetics, and products that emit radiation.

Health Resources and Services Administration (HRSA) is the primary federal agency for improving health care to people who are geographically isolated, economically or medically vulnerable.

The Immediate Office of the Secretary (IOS) is responsible for operations and coordination of the work of the Secretary.

The National Institutes of Health (NIH) seeks fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability.

NIH Clinical Center mission is to provide hope through pioneering clinical research to improve human health. The center has an Individual and collective passion for high reliability in the safe delivery of patient-centric care in a clinical research environment.

The Office of the Assistant Secretary for Preparedness and Response leads the nation's medical and public health preparedness for, response to, and recovery from disasters and public health emergencies.

The Office of the Chief Technology Officer (CTO) provides leadership and direction on data, technology, innovation and strategy across the HHS. Areas of focus include promoting open data and its use to create value, driving more efficient operations through technology utilization, and

coordinating innovation strategy across the Department to improve the lives of the American people and the performance of the Department.

Office of Inspector General (OIG) mission is to protect the integrity of Department of Health & Human Services (HHS) programs as well as the health and welfare of program beneficiaries.

Office of the National Coordinator Improve the health and well-being of individuals and communities through the use of technology and health information that is accessible when and where it matters most.

U.S. Government Accountability Office (GAO) examines how taxpayer dollars are spent and provides Congress and federal agencies with objective, reliable information to help the government save money and work more efficiently.

Healthcare Insurers and Providers

HCA Healthcare is committed to the care and improvement of human life. HCA follows a vision of healthcare the way it should be: patient-centered, constantly evolving and constantly improving, practiced with integrity and compassion.

UnitedHealth has a mission to help people live healthier lives and make the health system work better for everyone and are working to create a system that delivers high quality care, responsive to the needs of each person and the communities in which they live.

New York Presbyterian Hospital is one of the nation's most comprehensive, integrated academic health care delivery systems, dedicated to providing the highest quality, most compassionate care and service to patients in the New York metropolitan area, nationally, and throughout the globe.

Nonprofit & Philanthropic Organizations

Michael J. Fox Foundation is dedicated to finding a cure for Parkinson's disease through an aggressively funded research agenda and to ensuring the development of improved therapies for those living with Parkinson's today.

Robert Wood Johnson Foundation (RWJF) is the nation's largest philanthropy dedicated solely to health. RWJF supports research and programs targeting some of America's most pressing health issues—from substance abuse to improving access to quality health care.

Patient Advocacy

National Patient Advocate Foundation (NPAF), the advocacy affiliate of the Patient Advocate Foundation, represents the patient voice, both the powerful stories of individuals and the collective needs of the community. The NPAF's primary objective is to prioritize the patient voice in health system delivery reform to achieve person-centered care.

Patient-Centered Outcomes Research Institute (PCORI) helps people make informed healthcare decisions, and improves healthcare delivery and outcomes, by producing and promoting high-integrity, evidence-based information that comes from research guided by patients, caregivers, and the broader healthcare community.

Patient Privacy Rights' purpose is to honor and empower the individual's right to privacy through personal control of health information wherever such information is collected and used. They educate, collaborate

and partner with people to ensure privacy in law, policy, technology, and maximize the benefits from the use of personal health information with consent.

Private Sector

Allscripts is a leader in healthcare information technology solutions that advance clinical, financial and operational results. Its innovative solutions connect people, places and data across an Open, Connected Community of Health.

Amazon Web Services is a subsidiary of Amazon that provides on-demand cloud computing platforms to individuals, companies and governments, on a metered pay-as-you-go basis.

Apple is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, and online services.

Blackfynn helps the neuroscience and neurology communities make optimal use of data by powering an innovative platform that integrates and puts complex data in context.

Booz Allen Hamilton provides management and technology consulting and engineering services to leading Fortune 500 corporations, governments, and not-for-profits across the globe. Booz Allen partners with public and private sector clients to solve their most difficult challenges through a combination of consulting, analytics, mission operations, technology, systems delivery, cybersecurity, engineering, and innovation expertise.

Epic hires smart and motivated people from all academic majors to code, test, and implement healthcare software that hundreds of millions of patients and doctors rely on to improve care and ultimately save lives around the globe.

FlatIron Health is a healthcare technology and services company focused on accelerating cancer research and improving patient care. Their mission is to improve lives by learning from the experience of every cancer patient.

Geisinger Healthcare is a coordinated intersection of services and providers – primary care and specialists, hospitals and trauma centers, insurance, medical education and research. Geisinger has expanded and evolved to meet regional needs and developed innovative, national programs in the process.

Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, search engine, cloud computing, software, and hardware.

Google AI conducts research that advances the state-of-the-art in the field, applying AI to products and to new domains, and developing tools to ensure that everyone can access AI.

Health Catalyst is dedicated to enabling health care organizations to fundamentally improve care by building the most comprehensive and fully integrated suite of healthcare data warehousing and process improvement solutions available.

IBM Research is a community of thinkers dedicated to addressing some of the world's most complex problems and challenging opportunities for the benefit of all. They are one of the world's largest and most influential corporate research labs, with more than 3,000 researchers in 12 labs located across six continents.

Intel Corporation invents at the boundaries of technology to make amazing experiences possible for business and society for everyone. Leading on policy, diversity, inclusion, education and sustainability, we create value for our stockholders, customers, and society.

KB Stack Consulting helps government and non-profit organizations develop creative strategies for using data, evidence, and innovation to improve the impact of government social programs.

Mathematica Policy Research is dedicated to improving public well-being by bringing the highest standards of quality, objectivity, and excellence to bear on public policy. It advances its mission through objective, evidence-based standards, superior data, and collaboration.

Microsoft is an American multinational technology company with headquarters in Redmond, Washington. It develops, manufactures, licenses, supports and sells computer software, consumer electronics, personal computers, and related services.

Mpirica Health is a digital health company that uses machine learning, backed by a robust methodology, to score hospitals and surgeons based on objective clinical outcomes. Our cloud-based platform and API helps patients and payers, especially self-insured employers, avoid surgery risks and costs.

Oncology Analytics, Inc. provides health plans with an evidence-based, technologically driven approach to utilization management, which is purpose-built for oncology. They provide a technology enabled service to ensure cancer patients get the right treatment at the right time, and at an affordable price.

Pfizer's purpose is grounded in their commitment to fund programs that provide public benefit, advance medical care and improve patient outcomes. Their belief is that all people deserve to live healthy lives. This drives their desire to provide access to medicines that are safe, effective, and affordable.

TrialX brings researchers and patients together, accelerating clinical research to find cures for millions. They put the patient at the center and build solutions that work together seamlessly to meet the patient where they are, helping them find the trials they are looking for and connect with investigators close by.

Verily Life Science's mission is to bring together technology and life sciences to uncover new truths about health and disease. As an independent company, they are focused on using technology to better understand health, as well as prevent, detect, and manage disease.

Viz.ai's world class team of technologists, doctors, executives, and advocates has developed an AI-driven approach to care that is fast, effective, collaborative, and sustainable.

11 Health and Technologies have a complete care management platform that is the new gold standard in healthcare for both patients and clinicians by combining Smart Technology with the world's first one to one Patient Coach Program that provides both technical and emotional support to patients.

Roundtable on Sharing and Utilizing Health Data for AI Applications

U.S. Department of Health and Human Services | April 16, 2019

Purpose: Identify high-priority health applications of AI and key issues for an HHS AI strategy to address.

| | |
|-------|--|
| 9:00 | Registration and Light Breakfast |
| 9:40 | Welcome Mona Siddiqui, Chief Data Officer, U.S. Department of Health and Human Services (HHS) |
| 9:45 | Opening Remarks Ed Simcox, Chief Technology Officer, HHS |
| 9:55 | Structure of the Day Joel Gurin, President, Center for Open Data Enterprise (CODE) |
| 10:00 | Lightning Talks: Emerging Applications of AI Sohini Chowdhury, Deputy Chief Executive Officer, Michael J. Fox Foundation Jason Jones, Chief Data Scientist, Health Catalyst Vivian Lee, President of Health Platforms, Verily Life Sciences James Wiggins, Senior Solutions Architect for Academic Medical Centers, Amazon Web Services |
| 10:20 | Breakout Session 1: Identifying High-Priority AI Applications |
| 11:15 | Keynote Eric Hargan, Deputy Secretary, HHS |
| 11:30 | Networking Break |
| 11:45 | Breakout Session 2: Improving and Using Data for AI Applications |
| 12:30 | Lunch Break |
| 1:30 | Lightning Talks: Use Cases for an HHS AI Strategy Lisa Bari, Health IT and Interoperability Lead, Center for Medicare and Medicaid Innovation Christopher Boone, Vice President of Real World Data and Analytics, Pfizer Tim Persons, Chief Scientist and Managing Director, Government Accountability Office Kristen Finne, Senior Program Analyst, Assistant Secretary for Preparedness and Response, HHS |
| 2:00 | Breakout Session 3: Outlining Key Issues and Objectives for an HHS AI Strategy |
| 3:00 | Presentation of Highlights |
| 3:45 | Closing Remarks & Next Steps Mona Siddiqui, Chief Data Officer, HHS Joel Gurin, President, CODE |
| 4:00 | Adjourn for Optional reception |

*To ensure openness of discussion, the Roundtable will be held under the Chatham House Rule:
Any participant is free to use information from the day but is not allowed to reveal who made any comment.*

This Roundtable is brought to you by

