

Landscape Assessment of Public Opinion Work on Use of AI in Public Health

Commissioned Report

For the AAAS Center for Public Engagement with Science and Technology

This research was funded by AAAS through the Artificial Intelligence: Applications and Implications (AI2) Initiative, which is supported by Microsoft. The interpretations and conclusions contained in this document are those of the authors and do not necessarily represent the views of AAAS.

August 21, 2020

Prepared by

Todd P. Newman

Assistant Professor

Department of Life Sciences Communication
University of Wisconsin-Madison

Emily L. Howell

Civic Science Postdoctoral Fellow

Department of Life Sciences Communication
University of Wisconsin-Madison

Luye Bao

Doctoral Student

Department of Life Sciences Communication
University of Wisconsin-Madison

Becca Beets

Doctoral Student

Department of Life Sciences Communication
University of Wisconsin-Madison

Shiyu Yang

Doctoral Student

Department of Life Sciences Communication
University of Wisconsin-Madison

Recommended Citation: Newman, Todd P., Emily L. Howell, Luye Bao, Becca Beets, Shiyu Yang (2020). *Landscape Assessment of Public Opinion Work on Use of AI in Public Health*. Washington, DC: AAAS Center for Public Engagement with Science and Technology.

Executive Summary

Even prior to the COVID-19 pandemic, artificial intelligence (AI) and its potential uses across healthcare were receiving growing attention for the potential promise and challenges such applications could, and have already started, to bring. The pandemic, however, has further spread both advances in research and applications of AI and attention to the area of AI, particularly as it relates to healthcare. Yet, as with many of the applications of AI, codes of ethics and practice of AI in response to the pandemic and healthcare are now just being contemplated, often at a slower pace than the development and deployment of the technologies themselves. Similarly, despite public engagement being vital to this contemplation and to shaping decisions about how we design and use AI in society, policymakers often acknowledge the importance of engagement but have not yet seriously engaged the public in discourse and decision-making on AI.

This report summarizes existing work on public attitudes concerning AI, particularly in areas related to addressing the pandemic, such as AI and surveillance (e.g. contact tracing), balancing safety and privacy in use of AI, using AI in triage and resource allocation in hospitals and healthcare, and other uses related to the pandemic, such as advancing research on potential treatments of the COVID-19 virus. We first introduce AI and the limited existing policy in the U.S., summarizing the few recent committee recommendations and ongoing work at the federal level on AI as it relates to the current pandemic and economic crisis. We then describe some of the ongoing work and concerns related to ethics and human rights associated with AI in general and in the context of healthcare and COVID-19 in particular. Finally, we synthesize existing work on public attitudes related to AI in general and in the context of healthcare and, where studies exist, related to the current pandemic.

Very little of the small but growing body of public opinion work on public views of AI addresses the bigger ethical and equity questions connected to AI development and use, nor how views differ across diverse demographic and geographic experiences. To begin with, few studies exist that are focused on public opinion of AI in the U.S., and those that do tend to concentrate on Americans more broadly, without breakdowns, focus, or oversampling of different publics based on demographic and geographic distinctions. Polls also tend to focus on specific, sometimes very hypothetical or obscure-seeming potential applications of AI, making it difficult to piece together an overall picture of public opinion regarding AI in general and in healthcare contexts in particular.

When it comes to AI in healthcare contexts, there is also a lack of survey data on how diverse publics perceive the potential risks and benefits that come with AI applications in various healthcare domains, with especially little work focused on the key action areas of concern. Americans in general, however, seem to perceive healthcare as an area in which AI applications could be particularly beneficial. We also see concerns, however, regarding specific applications, especially ones in which AI is involved in decision-making, and regarding data and privacy of health information. Despite low overall levels of familiarity with AI as a concept or in specific applications, many Americans do have awareness and concern about lack of control over how

their data is collected and used, and by whom. Polls also indicate that most Americans, regardless of their gender or race, are hesitant to think that AI with “true” intelligence or decision-making power in healthcare contexts would be a good idea. They also have very little trust in companies when it comes to companies using personal health data.

Although questions about responsibility and rights involved in data collection and usage are becoming increasingly important, these are reflected in only a few existing public opinion surveys. Though surveys of experts found that they frequently mentioned potential inequalities caused by AI in healthcare (Pew Research Center, 2018), existing surveys rarely tapped into the public’s views on the inequity issues worsened by AI. Furthermore, the general public are likely to rely on their trust in different institutions and social groups and their perceptions of their own ability to control how AI and data use affects their lives. Little of the survey research that we found, however, included items measuring the American public’s trust levels in institutions and social groups involved in AI development (although a few of them assessed trust in different health data-receiving institutions).

To fill in these gaps, future research should tap into the public’s views regarding how, why, and for and by whom personal health data should be collected and used. In addition, current laws do not apply in contexts where much of the data used to inform AI applications are collected (such as through cellphone applications and consumer-buying behavior), posing legal gaps in responsible development of AI. Going forward, polls should also collect public awareness and opinion across the stages of the AI system lifecycle: 1) design, data collection, and model building, 2) verification and validation, 3) deployment, and 4) operation and monitoring of AI systems. Each of these stages can have unique implications and need for public engagement. In particular, we also need more research focused on understanding the variation in views across demographic differences and experiences with AI and with science and institutions more broadly, for facilitating representative and equitable engagement and decision-making on AI.

Future public opinion surveys can help us understand the unique opportunities and challenges underlying each stage of the AI lifecycle. Future surveys should also pay special attention to minority groups and groups that are traditionally digitally marginalized, as these groups are more susceptible to negative outcomes of AI usage and applications (for example, racial bias embodied in AI systems). Survey items gauging the AI digital divide among the U.S. public-- issues related to algorithmic literacy, AI awareness, understanding, access, affordability, availability, and adaptability--would be especially helpful for filling the gaps in AI fairness and equity and how we want to use AI, or not, to address the COVID-19 pandemic and other issues we face going forward.

Introduction

The breadth of potential applications of artificial intelligence (AI) and the lack of a clear definition to encompass all of these existing and potential tools and their applications makes communication and decision-making concerning AI difficult (Calo, 2017), as do the scientific and ethical complexity and uncertainties inherent in the technologies (e.g. Al-Almoudi & Latsis, 2019). While AI systems have been developing for decades, advances have accelerated in recent years, with many applications existing and in development across different sectors, such as healthcare, agriculture, national security, and transportation (Zhang & Dafoe, 2019). This expansion and the potential scope of AI and its implications for individuals and society mean that questions of ethics, policy, and public attitudes and engagement related to AI are increasingly important.

Further, even prior to the COVID-19 pandemic, AI and its potential uses across healthcare – such as hospital logistics and triage decision-making, population health tracking, diagnostics support, robotic-supported surgeries and remote care, and vaccine and treatment development – were receiving growing attention for the potential promise and challenges such applications already have and could easily continue to bring.

The pandemic, however, has further spread both research and applications advances and attention to the area of AI in healthcare. For example, the use of machine learning and AI in radiological imaging continues to grow given its promise to accelerate radiology workflow and patient care for COVID-19 (Kundu et al., 2020). Yet, as with many of the applications of AI, codes of ethics and practice are just now being contemplated. Similarly, despite public engagement being vital to this contemplation and for shaping decisions in how we design and use AI in society, policymakers often acknowledge the importance of engagement but have not yet seriously moved to engaging citizens in discourse and decision-making on AI.

Small but growing bodies of public opinion work capture public views of different aspects, applications, and potential impacts of AI. AI can reinforce or worsen existing disparities in society in general, and in healthcare in particular, with algorithms relying on systematically biased datasets, designed with biases in their code, and providing outputs interpreted and acted on by biased humans (Yu, 2020). As new technologies, there is also the likelihood that beneficial applications will not be accessible to those who are already underserved by existing healthcare and other systems, which we see with the existence of the digital divide. Of the public opinion research that does exist, however, less common is research that captures differences in views and experiences across groups that will likely be most affected by these disparities and systematic biases.

Here we provide a brief overview of the AI landscape regarding U.S. policy, discussions of ethics, and public attitudes across applications of AI, particularly in the areas of healthcare and addressing the COVID-19 pandemic. As mentioned above, AI's scope in terms of technologies and applications across sectors is so large that defining AI is challenging, and numerous definitions exist. Here we use a broad definition of AI, provided by the Organization for

Economic Cooperation and Development (OECD), reflecting the likelihood that many publics will similarly have a broad range of mental pictures of AI in mind. The OECD 2019 *Recommendation of the Council on Artificial Intelligence* defined an AI system as

a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (OECD, 2019).

This definition also informed the OECD “Principles for AI” (OECD, 2020) that member countries, including the U.S., signed in 2019 and that later informed the G20 agreement on AI. Helpfully, the OECD 2019 *Recommendation* also defined the stages of the AI system lifecycle. These are the typically iterative phases that make up development and use of AI and which each can have unique implications and need for discussions regarding policy, ethics, and public engagement. They are: 1) design, data collection, and model building; 2) verification and validation; 3) deployment; and 4) operation and monitoring.

In the rest of this report, we describe existing U.S. policy related to AI, particularly in healthcare, synthesize the many ethical questions tied up with AI research and applications, and highlight existing work and, more so, the many remaining gaps regarding research and polling on public views of and engagement with AI research and applications, policy, and ethical questions.

Legal implications and policies related to key action areas

Despite national concerns about AI, particularly related to the potential impacts of automation and robots, since at least the 1960s, the U.S. federal government has only recently started dedicating attention to AI technology development and usage and its potential impacts (Calo, 2017). The U.S. federal government has been much more active in recent years, however, in starting to assess the AI landscape and make policy concerning AI research and applications broadly (Castro & Johnson, 2020). This started with the Obama White House and Congress in 2016 (Calo, 2017) and has picked up markedly since then under the Trump White House (Castro & Johnson, 2020).

Most of the policy advances and recommendations have occurred since 2019, when Trump signed Executive Order 13859: Maintaining American Leadership in Artificial Intelligence (Castro & Johnson, 2020; Executive Order 13859, 2019). Reflecting the strong national security focus in AI policy in the U.S. in recent years, Congress also created the National Security Commission on Artificial Intelligence (NSCAI) in early 2019 (NSCAI, 2019). The NSCAI is providing many of the recommendations on national policy concerning AI, particularly in the context of COVID-19 and healthcare applications. The rest of this section gives an overview of executive and legislative branch policy related to AI more broadly and to health-related applications and sectors in particular.

Executive Branch

The Executive Order 13859 launched the official U.S. strategy on AI, called the American AI Initiative (The White House & Office of Science and Technology Policy, 2020). That strategy is based on five pillars:

- 1) investing in research and development in AI;
- 2) making more federal AI resources available;
- 3) removing barriers to AI innovation;
- 4) improving training for developers and users of AI in the workforce; and
- 5) promoting an international environment that is supportive for American AI innovation and responsible use of AI (The White House & Office of Science and Technology Policy, 2020).

Much of the White House focus is on AI and its implications for technological innovations and for American workers, particularly in the context of ensuring U.S. security and technological and economic competitiveness internationally and especially relative to China (The White House, 2020; NSCAI, 2020a). It focuses on programs to provide education and (re)training for a STEM workforce, both on the research and application sides. It also created a Select Committee on AI, under the National Science and Technology Council in the Office of Science and Technology Policy, to advise the White House on interagency research and development (R&D) priorities concerning AI, creating partnerships with industry and academia, and identifying structures and opportunities for improving planning, coordination, and support of AI R&D (The White House & Office of Science and Technology Policy, 2020).

The White House AI Initiative also stresses that it prioritizes AI that meets American values, which is a less filled-in section than the others focused on technological innovation and AI for American workers. It states that ensuring American values means ensuring AI technologies are “understandable, trustworthy, robust, and safe” (White House website, 2020). The Trump Administration also signed the OECD 2019 *Recommendation on AI*, however, which laid out much more detailed principles for human-centered AI design, based on inclusivity, sustainability, fairness, transparency and explainability, safety and security, and accountability (Castro & Johnson, 2020; OECD, 2019). These principles are to ensure that signing countries “proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet,” which include outcomes such as improving human capabilities, enhancing creativity, enhancing inclusion, reducing inequity, and protecting the environment (OECD, 2020). The principles also include recommendations for doing so, primarily through ensuring holistic and transparent public discourse, engagement, and assessment regarding AI systems and applications (OECD, 2020).

In addition to the Select Committee on AI under the National Science and Technology Council, part of the Office of Science and Technology Policy, a dozen executive branch agencies and organizations play a role in AI research, development, and deployment in the U.S. (The White

House & Office of Science and Technology Policy, 2020). These include the National Science Foundation (NSF) as funders of AI research, the Defense Advanced Research Projects Agency and Department of Defense (DARPA and DoD, respectively) in military applications, the Intelligence Advanced Research Projects Activity (IARPA) in intelligence-related challenges, the National Institute of Standards and Technology in informing technical standards for “reliable, robust, and trustworthy AI”, and the Departments of Energy, Veterans Affairs, Transportation, Labor, Commerce, and Health and Human Services, as well as the National Institutes of Health (NIH) and the National Oceanic and Atmospheric Administration (NOAA) across different applications and sectors. The Food and Drug Administration (FDA) and U.S. Department of Agriculture (USDA) also act in their regulatory roles concerning new AI applications in the areas that fall under each agency’s purview (The White House, 2020).

Legislative Branch: The National Security Commission on Artificial Intelligence

Following its creation by Congress in 2019 through the *John S. McCain National Defense Authorization Act*, the NSCAI developed four working groups that meet monthly to define national objectives, assess challenges, and create recommendations concerning AI (NSCAI, 2019). These working groups focus on 1) maintaining U.S. global leadership in AI research, 2) maintaining global leadership in AI national security applications, 3) preparing citizens for a future with AI, and 4) international competitiveness and cooperation regarding AI research and deployment (NSCAI, 2019). The NSCAI also decided to pursue three special projects on cross-cutting issues concerning 1) developing AI through public-private partnerships, 2) pursuing responsible and ethical use of AI for national security, and 3) managing data to support AI applications (NSCAI, 2019).

The Commission is made up of 15 Commissioners appointed by Congress and by the Secretaries of Defense and Commerce, is bipartisan, and represents experts from across the private sector, academia, and government. They collaborate with the White House, the National Security Council, the Office of Science and Technology Policy, and DoD, as well as engage with experts from academic and private sector organizations. Since the pandemic started, the NSCAI has treated the pandemic as a national security issue as well, and produced some of the most detailed recommendations regarding use of AI to address economic, health, and security implications of the COVID-19 pandemic, described below.

In the context of healthcare and COVID-19

As the pandemic hit, the NSCAI has released several reports recommending policies and best practices across potential and current uses of AI related to addressing the COVID-19 pandemic. Announcing that “health security is national security,” each of the reports tackles one aspect of the potential of AI to address COVID-19 and future pandemics: 1) *The Role of AI Technology in Pandemic Preparedness*; 2) *Mitigating Economic Impacts of the COVID-19 Pandemic and Preserving U.S. Strategic Competitiveness in AI*; and 3) *Privacy and Ethics Recommendations for Computing Applications Developed to Mitigate COVID-19*. These reports cleanly lay out their

recommendations in each respective area of focus, so we will not reproduce the specific recommendations here.

Broadly, however, all see AI technologies as offering “powerful potential for detecting and containing the virus, driving biomedical innovation . . . and improving response and recovery” (NSCAI, 2020c). The first report listed above, on pandemic preparedness, highlights four areas where the NSCAI sees promising existing and potential future uses of AI in pandemic response:

- 1) disease surveillance through: disease modeling (*already in use*); early outbreak detection (*already in use*); monitoring and preventing disease jumps from other species (*already in use and in development*); advancing bioforensics to identify whether future viruses are natural, intentional, or accidentally released (*in development*); and detecting pathogens with advanced sensing, including tracing droplet spread, detecting the early onset of infections in individuals, and using mobile proximity detected to encourage social distances (*in development*).
- 2) improved diagnostics and triage: in test design and informing expert decisions, detecting COVID infections, and assessing severity and care needs of patients (*already in use*).
- 3) developing vaccines, treatment, and equipment, such as through creating simulations of how a virus might interact with different treatments (*in use*), managing supply chains (*in use*), and creating self-correcting supply chains (*in development*).
- 4) ongoing healthcare, including: managing patients’ risk (*in use in some capacity*); reducing transmission rates through using robots (*in use*); and remotely monitoring patients (*in development*).

The report recommends several avenues for increasing and/or making permanent funding and resources to further research and development in these areas.

In the second report, on mitigating economic impacts, the NSCAI recommends using AI to inform reopening the economy and for understanding and expanding the capacity of the national security supply chain, among other recommendations related to expanding U.S. national and economic security as it relates to tech-related innovation, financing, and production (NSCAI, 2020a). Finally, in the report focused on privacy and ethics, the NSCAI makes two recommendations for applications designed to mitigate the effects of COVID-19. The two recommendations are 1) leverage technologies, policies, and laws to “put civil liberties considerations at the center” of contact tracing methods and applications, and 2) ensure that federally funded tools are created with sensitivity to and corrections for potential bias, to a level that ensures the tools at least do not introduce additional disparities and unfairness into healthcare delivery and outcomes (NSCAI, 2020c). For the first recommendation, the Commission outlines a dozen detailed best practices for actors to inform their contact tracing methods in ways that protect civil liberties and privacy, including ensuring that all data collection is voluntary and that the minimum amount of data is collected and is stored, securely, only as long as is needed to complete the tracing and only for the purposes in which individuals gave explicit consent.

In addition to the NSCAI recommendations, prior to the outbreak of COVID-19 in the U.S., in January 2020 the White House proposed regulatory principles for using AI in the private sector, which were open for public comment through March 2020 and are currently under consideration at the Office of Management and Budget (OMB) (The White House, 2020). The regulatory principles are meant to ensure public engagement in AI development and use, limit regulatory overreach in AI in the private sector, and promote trustworthy technology. The proposed principles would apply across many different sectors, including healthcare, and the White House specifically highlighted three areas of ongoing and future government-supported AI research and deployment related to COVID-19.

Most of these areas focus on how the administration is working to make data and resources available for researchers studying different aspects of the virus and pandemic. The first is a call to action to develop publicly available AI tools and methods to analyze the COVID-19 Open Research Dataset, which is a collection of machine-readable scholarly articles on the virus. The second is Oak Ridge National Laboratory's research using AI to speed COVID-19 treatment and vaccine research. The last is the White House partnering with the Department of Energy, National Science Foundation, and members of industry and academia to establish the COVID-19 High Performance Computing Consortium, where volunteers provide compute time and resources to COVID-19 researchers (The White House, 2020).

More broadly, much of the current debate about AI in healthcare focuses on issues of privacy, perhaps because those are areas where already issues are arising concerning how, why, and for and by whom personal health data is collected and used. Existing laws shaping policy concerning healthcare and privacy in the U.S. are the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and the Genetic Information Nondiscrimination Act (GINA) of 2008 (Rosenblat et al., 2014). The former protects patient privacy and health coverage in health, employment, and insurance contexts. The latter prohibits employers and insurance companies from discriminating against employees and insuring on the basis of their genetic profiles.

These laws are limited in how they have so far applied to many of the contexts in which AI and health privacy concerns appear. Much of the data that informs AI applications, however, comes from contexts outside of those these laws apply in, such as through cellphone applications and consumer buying behavior. As AI continues to improve, it also is easier and easier through pattern recognition and big data to infer individuals' health from seemingly non-health related information (Rosenblat et al., 2014) and from information that is collected in situations considered public (and therefore limited in the legal interpretations of what privacy people can expect) and/or in which individuals cannot reasonably consent, such as purchases in a grocery store (Calo, 2017).

Despite the prominence of the ethical issues raised by privacy concerns, relative to the many other ethical questions we synthesize in the following section, U.S. policy (and public opinion research) has largely not addressed the topic, let alone the many other ethical considerations

we turn to next (Calo, 2017; Al-Amoudi & Lastis, 2019). Further, despite the emphasis on the importance of public engagement in many of the NCSAI reports described above, there is similarly apparently little discussion in the existing federal policy materials of what publics, in what aspects of an AI system’s life cycle, and how this engagement will shape policy and AI development and use. This lack of in-depth attention to processes for ensuring consideration of ethics, human rights, and public engagement does not appear to be unique in national strategies related to AI from federal governments around the world (Bradley, Wingfield, & Metzger, 2020).

Although public engagement is vital to democratic and risk-related decision-making in general, it will likely be particularly important and difficult to accomplish in the case of AI applications, especially in healthcare contexts. AI, as a self-learning technology, makes decisions through black-boxed processes. In many cases, the algorithms and processes will not be accessible because they are proprietary, and even when they are accessible and transparent, many will likely be too complex for most people including experts to quickly interpret, making auditing difficult. But because AI is self-learning, another substantial barrier and concern is that we do not (yet, if ever) have the ability to understand the AI’s justifications for each decision it makes in an application. This makes it difficult to evaluate AI decisions in general, but particularly important for democratic public policy, as others have highlighted, is this black-boxing raises serious challenges to public discourse on the normative aspects of AI-based decision-making (Al-Amoudi & Lastis, 2019). These challenges make it so that being proactive in generating diverse and inclusive public discussion on the ethics and human rights issues related to AI, described below, is vital.

Ethics and human rights related to key action areas

Many of the ethical and human rights-based considerations for artificial intelligence can be thought of as tensions or tradeoffs between the potential benefits and risks of AI and the principles guiding its use (High-Level Expert Group on Artificial Intelligence, 2019; Whittlestone, Nyrup, Alexandrova, Dihal, & Cave, 2019). Although these tensions can be applied to a variety of applications of AI, they are particularly salient in the context of public health. For example, increased quality in AI-based health systems could have direct benefits for both individual and community health through improved access and medical diagnostics. However, these benefits must be weighed against risks to individual privacy and autonomy, or other risks that may be unevenly distributed. In the event of a global health crisis, AI could provide opportunities to predict and track a virus, identify “optimal” treatment plans, and provide improved monitoring, communications, and containment. Using AI in this context also raises questions about data quality and bias as well as concerns about the misuses of AI for containment purposes (Davies, 2019).

Ethics and AI

Organizations around the globe—including governments and intragovernmental organizations, private companies, academic institutions, and non-profit organizations, among others—have developed guidelines for the ethical use of AI. An analysis of international ethics guidelines for AI identified 84 unique sources, including 21 from the United States (Jobin, Ienca, & Vayena, 2019). Of the eleven recurring ethical principles categorized by the authors, five were the most prevalent: transparency (87%), justice and fairness (81%), non-maleficence (71%), responsibility (71%), and privacy (56%). The authors observed variation in how the principles were defined and applied across the guidelines. While some principles, such as responsibility, often lacked clear definitions, others were conceptualized in a variety of ways. Justice, for example, was “mainly expressed in terms of fairness, and of prevention, monitoring or mitigation of unwanted bias and discrimination” (p. 394). However, some guidelines referenced justice in terms of diversity, equality, and access.

A similar set of ethical principles were delineated by Floridi & Cowls (2019) in a review of six global initiatives aimed at “socially beneficial” AI. The authors identified 47 unique ethical issues related to responsible AI practices, which were categorized into the five overarching principles defined below:

- *Beneficence*: promoting well-being, preserving dignity, and sustaining the planet
- *Non-maleficence*: privacy, security and ‘capability caution’
- *Autonomy*: the power to decide (to decide)
- *Justice*: promoting prosperity, preserving solidarity, avoiding unfairness
- *Explicability*: enabling the other principles through intelligibility and accountability

While there is clear overlap in many of the ethical principles outlined in global AI initiatives, Jobin et al. (2019) note that some regions such as “Africa, South and Central America and Central Asia” are underrepresented in the existing guidelines (p. 396). Floridi & Cowls (2019) raise a similar limitation of their ethical framework, which is drawn from documents that are “global in scope, or from within western liberal democracies” (p. 9).

Although the guidance documents are typically concerned with AI broadly, many of the ethical concerns can be applied directly to public health applications of AI. A review by Morley et al. (2019) explores the ethical issues that are unique to AI in a healthcare context. The authors highlight concerns related to evidence, outcomes, effects, and traceability, including:

- *Inconclusive evidence*: infallibility and causality in diagnosis
- *Inscrutable evidence*: lack of transparency or clarity in the data
- *Misguided evidence*: reliability of data and concerns about bias
- *Unfair outcomes*: justice and fairness in who receives treatment
- *Transformative effects*: privacy and loss of autonomy in decision-making
- *Traceability*: accountability and responsibility for AI decisions

The ethical considerations for AI in public health may also vary at the individual and group level. For example, at the individual level there may be concerns about surveillance or misdiagnosis whereas at the group level there may be concerns about discrimination. Other considerations such as liability and allocation of resources may be most relevant at the institutional level, while broader questions of the national health system pertain more directly to the societal level (Morley et al., 2019).

Human Rights and AI

As an alternative to the ethics-based guidelines outlined above, some scholars have argued for the adoption of a human rights-centered approach to the governance of AI (Beduschi, 2020; Bradley, Wingfield, & Metzger, 2020; Donahoe & Metzger, 2019; Latonero, 2018; Yeung, Howes, & Pogrebna, 2020). The European Commission’s guidelines for trustworthy AI, for example, are grounded in human rights. As noted in their report, “respect for fundamental rights, within a framework of democracy and the rule of law, provides the most promising foundations for identifying abstract ethical principles and values, which can be operationalised in the context of AI” (High-Level Expert Group on Artificial Intelligence, 2019, p. 9). According to Donahoe and Metzger (2019), a human rights-based approach to the governance of AI has four distinct benefits:

- Human rights are focused on human impacts
- Human rights principles comprise many of the ethical issues raised by AI
- Human rights frameworks apply to both governments and to the private sector
- International human rights laws are established and recognized

Unlike AI ethics, which lacks a centrally defined framework (Yeung et al., 2020), there are three established international human rights instruments that could serve as a foundation for human rights considerations for AI: the Universal Declaration of Human Rights (UDHR), which laid the groundwork for international human rights law (United Nations, 2020), the International Covenant on Civil and Political Rights (ICCPR), and the International Covenant on Economic, Social, and Cultural Rights (ICESCR) (Raso, Hilligoss, Krishnamurthy, Bavitz, & Kim, 2018). The United States has signed the ICESCR and is a party to (signed and ratified) the ICCPR. Ratification represents a binding agreement to the treaty where governments are expected to “put into place domestic measures and legislation compatible with their treaty obligations and duties” (United Nations, 2020).

A number of scholars have identified specific human rights outlined in the UDHR, ICCPR, and ICESCR that are relevant to AI (Beduschi, 2020; Bradley et al., 2020; Donahoe & Metzger, 2019; Latonero, 2018; Raso et al., 2018). Examples include:

- Right to equal protection and nondiscrimination
- Right to freedom of expression
- Right to privacy
- Right to work

- Right to an adequate standard of living
- Right to life, liberty, and security of person
- Right to political participation
- Right to education

Many of these rights translate directly to public health applications of AI, such as the use of AI for medical diagnostics. Improvements in diagnostics could have benefits for individual health outcomes (right to life) and quality of life (right to standard of living) (Raso et al., 2018). However, improvements in certain diagnostic fields such as radiology may negatively impact individual job security (right to work) or raise concerns about individual data protections (right to privacy).

Some governments have included human rights principles and frameworks in their National Artificial Intelligence Strategies (NAS) (Bradley et al., 2020). In the U.S., the NAS—Executive Order 13859: Maintaining American Leadership in Artificial Intelligence—does not directly mention human rights, although it does refer to the protection of civil liberties and privacy (Executive Order 13859, 2019). Other organizations have incorporated both human rights and broader ethical principles into their guidelines for AI. For example, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017) report on *Ethically Aligned Design* includes both ethics and human rights considerations for the responsible use of autonomous and intelligent systems (A/IS). The report has five guiding principles for the development and implementation of A/IS:

- *Human Rights*: Ensure they do not infringe on internationally recognized human rights
- *Well-being*: Prioritize metrics of human well-being in their design and use
- *Accountability*: Ensure that designers and operators are responsible and accountable
- *Transparency*: Ensure they operate in a transparent manner
- *Awareness of misuse*: Extend benefits and minimize the risks of misuse

The IEEE report is “intended to be a working reference tool created in an inclusive process by those in the relevant scientific and engineering communities prioritizing ethical considerations in their work” (p.17).

Several other reports provide overviews of issues and frameworks related to ethics and human rights concerns connected to AI. These are synthesized in *Further Reading* at the end of this report.

Marginalized Groups and AI

The tradeoffs of AI may not be evenly distributed, particularly among marginalized groups or vulnerable populations in the United States. As Raso et al. (2018) note, “some individuals and groups experience positive impacts from the very same applications that adversely impact other rights-holders” (p. 17). For example, AI has the potential to widen existing divides in the U.S. healthcare system. Women, children, people with disabilities, and ethnic or racial minorities

have been traditionally considered to be vulnerable populations that may require additional protections with the rise of AI and other data-driven technologies (Malgieri & Niklas, 2020).

One example is members of Indigenous communities. The Indigenous Protocol and Artificial Intelligence Working Group recently outlined seven core considerations for the design and application of AI systems in Indigenous contexts (Lewis, 2020). The guidelines are designed to be considered broadly, keeping in mind that “a single ‘Indigenous perspective’ does not exist, as epistemologies are motivated and shaped by the grounding of specific communities in particular territories” (p. 4). The working group’s recommendations are summarized below:

- AI systems should be designed in partnership with local Indigenous communities
- AI systems should be designed to understand how humans and non-humans are related
- AI systems developed by, with, or for Indigenous communities should be responsible to those communities
- Governance for AI systems should be grounded in Indigenous protocol
- AI system designers should be aware of their own cultural frameworks and bias
- Every component of the AI system should be considered in the ethical evaluation of the system, which is grounded in doing no harm
- AI systems should be designed to respect and support data sovereignty

As Yu (2020) describes, many of the inequities currently gaining attention surrounding algorithmic bias are not new. Rather, prior research on the ‘digital divide’ over the last two decades provides a context to systematically think about the different attributes of the algorithmic divide. In other words, just as those without access to the Internet, information technology, and digital content are marginalized from fully benefiting from the affordances of these technologies, so too has it marginalized those from fully benefiting from AI. Yet, as the previous sections highlight, AI technologies have the potential to contribute to a number of key areas affecting those who are the most vulnerable, including in the areas of public health, food production, education, and policy analysis. In order to conceptualize the ways in which AI may lead to unequal benefits in these areas, five key attributes contributing to the algorithmic divide include (1) awareness, (2) access, (3) affordability, (4) availability, and (5) adaptability.

- **Awareness:** Many individuals remain unaware of the impacts of AI and may not appreciate how the increased use of machine-learning algorithms and intelligent machines can impact their lives - both positively and negatively. Moreover, even among those who are aware of the impacts of AI on both themselves and society, most have a limited understanding of how AI algorithms actually operate. This is not surprising: programmers are the only ones who can know for sure what an algorithm does, and at times they may not be clear about what is going on (Rainie & Anderson, 2017). There is a need, therefore, for individuals to become at least more aware of the strengths and drawbacks of AI technologies and services. To the extent AI impacts go unnoticed and unchallenged, these technologies can further steer individuals away from active or equal participation in new technologies.

- **Access:** While the use of algorithms can provide important individual and societal benefits, not everybody has access to AI products and services. Individuals may be shut out because they cannot afford these products and services, cannot find them on the local market, or do not have the needed skills to use them effectively. People who lack access are also overwhelmingly from groups that are already marginalized, including the elderly, those with disabilities, indigenous groups, and those living in poor, remote, or rural areas. One avenue that may provide lessons for addressing access related challenges is the prior research on information and communication technology for development (ICT4D) (De-Artega et al., 2018). While focused primarily on the developing world, many of the infrastructure limitations, such as handling small and messy datasets, may be applied to other contexts.
- **Affordability:** Related to access, but emphasizes the types of products or services that an individual can access, frequency of use, and the ability to upgrade services in the future to enjoy the full benefits. The less access one can afford, the more limited the benefits of AI products and services, and the less likely to fully realize the promise of the technology.
- **Availability:** Individuals have differing needs for AI products and services, and more often than not, those products or services feature AI designed by those who do not fully grasp the users' specific needs, interests, or conditions. As algorithms learn when new data is fed back into the algorithm, those algorithms that are improperly designed or that rely on problematic data could amplify these biases through feedback loops.
- **Adaptability:** If individuals are to succeed in the age of AI, they will need to take advantage of the different AI products and services. Adaptability, however, requires both knowledge and understanding (in addition to awareness). Therefore, just as important as awareness, individuals also need to develop algorithmic literacy. While algorithmic literacy may help individuals learn the full potential of AI technologies, it will also assist in knowing which products or services fail to protect privacy or other individual rights. Likewise, while many individuals do not fully comprehend how AI technologies operate, individuals are capable of "tricking" algorithms to provide desirable results (Chander, 2017). For example, when users were made aware of Facebook's algorithm for curating news feeds, they modified their behavior to take advantage of tailoring their news feed.

Public attitudes related to key action areas

In this section, we review public opinion surveys, reports, and journal articles on public and expert attitudes toward AI. We also perform secondary analyses on publicly available datasets and provide findings on attitudes toward AI and its application in healthcare, broken down by demographic variables, with a primary focus on marginalized groups. We draw on two publicly available datasets from Pew Research Center, one Monmouth University Poll, and a recent dataset from the SCIMEP research group at UW-Madison. Table 1 presents an overview of

existing nationally representative surveys reaching U.S. adults aged over 18 on their opinion of AI related to health. In total, we review 11 surveys and reports on public opinion of AI and healthcare. Only one of these surveys and reports contained questions about AI and COVID-19 (Pew Research Center, 2020).

Table 1. Overview of existing U.S. nationally representative surveys of public opinion of AI related to health

Survey	Year	Sampling	Weighted On	Topic
Pew Research Center*	2014, Feb	Non-probability sample; telephone interview, cellphone and landline (conducted by Princeton Survey Research Associates International; N= 1,001)	Age, education, gender, race, Hispanic origin, population density, region (U.S. Census definitions), telephone usage	Future; Health; Medicine; Science
Monmouth University Poll*	2015, Mar–Apr	Non-probability sample; telephone interview, cellphone and landline (conducted by Monmouth University Polling Institute; N= 1,005)	Age, race, education	Information; Science
Vanity Fair/60 Minutes	2016, Mar	Non-probability sample; telephone interview (conducted by CBS News; N = 1,021)		Communications Technology; Mood; Religion; Technology
Pew Research Center	2017, May–Jun	Non-probability sample; web-based survey (conducted by GfK KnowledgePanel; N = 4,024)	Age, education, gender, race/Hispanic origin, Hispanic nativity, home internet access	Media; Science
Ghafur <i>et al.</i>	2018	Non-probability sample; web-based survey (conducted by YouGov; N= 1,114)	Age, education, gender, race	Public perceptions on data sharing
Zhang & Dafoe	2018, Jun	Non-probability sample; web-based survey (conducted by YouGov; N= 2,000)	Age, education, gender, race	American public’s attitudes toward AI and AI governance
Pew Research Center	2019, Jan	Non-probability sample, from probability-based panel; web-based survey (conducted by Ipsos; N= 4,464)	Age, education, gender, race/Hispanic origin, Hispanic nativity, home internet access	Science
Pew Research Center*	2019, Jun	Non-probability sample, from probability-based panel; web-based survey (conducted by Ipsos; N= 4,272)	Age, education, gender, race/Hispanic origin,	Privacy and surveillance

			Hispanic nativity, home internet access	
Rock Health & Stanford Center for Digital Health	2019, Jul – Aug	Non-probability sample; web-based survey (conducted by Toluna; N= 4,000)	Age, gender, geographic region, income, race	Annual digital health consumer adoption
SCIMEP*	2020, Feb – Mar	Non-probability sample; web-based survey (conducted by YouGov; N= 2,700)	Age, education, gender, race	Public attitudes on artificial intelligence
Pew Research Center*	2020, Apr	Non-probability sample, from probability-based panel; web-based survey (conducted by Ipsos; N= 4,917)	Age, education, gender, race/Hispanic origin, Hispanic nativity, home internet access	Responses to the new coronavirus outbreak

Notes: 1. Non-probability sample: Post-stratification weights for each sample are applied to the analyses to ensure the representativeness.
2. * Used for secondary analyses.

Public familiarity with AI

Most Americans have at least some (although generally very little) exposure to AI broadly. By 2015, over two-thirds (70%) of Americans had heard of the term “artificial intelligence” or “AI” (Monmouth University Polling Institute, 2015). Broken down by race, 73% of whites indicated they had heard of the term, followed by 67% of Asians, 65% of Hispanics, and 55% of Blacks. When this was further broken down by educational attainment, we observed a general trend across racial groups that as the level of educational attainment went up, the percentage of respondents having heard of the terms also went up within each racial group. For example, while only 33% of Black Americans who did not complete high school had heard of AI, 71% of Black Americans who completed a graduate degree reported having heard of the terms before. Similarly, 55% of whites who did not complete high school had heard of AI, but this number jumped to 96% for whites who had a graduate degree. Overall, a notably greater proportion of respondents with a four-year college degree (83%) or above (90%) had heard of the term compared with respondents who completed only 8th grade or less (40%). Gender-wise, more men (76%) than women (64%) indicated they had heard of the term. No demographic questions regarding family income and urban/rural areas were asked in the 2015 survey.

According to the same Monmouth University survey, the majority of Americans have also had direct experience interacting with AI applications, mainly through their cellphones and smart personal assistants, such as Alexa, Siri, Cortana, and Google Assistant. In 2015, 61% of Americans reported owning a smartphone or other device with voice recognition technology such as Apple’s Siri or Ok Google (Monmouth University Polling Institute, 2015).

When examining differences across race, despite people in minority racial groups being less likely to report familiarity with AI, the same 2015 poll found that minorities were more likely to own smartphones. In 2015, 74% of Hispanics reported owning a smartphone or other device

with voice recognition technology, topping the list. This was followed by Asians (72%) and Blacks (68%), while just 56% of whites said they owned such a device (Monmouth University Polling Institute, 2015).

In 2015, notably more respondents having a four-year college degree (73%) or above (80%) reported owning such a device than respondents who completed only 8th grade or less (29%), which could also reflect income differences related to education level, although breakdowns by income are not available through this poll (Monmouth, 2015).

Small partisan differences seem like they could exist in familiarity and use, as well. In 2015 fewer respondents who were “somewhat conservative” reported having heard of the term AI (60%), whereas for each of the other ideological groups (including those who were “very liberal,” “very conservative,” “somewhat liberal,” and “moderate”) about 72-78% of respondents said they had heard of the term. By partisanship, 66% of those who self-identified as “Republicans” and 63% of those who described themselves as “Democrats” reported having heard of the terms before, compared with 76% of “Independents” and 76% of those who identified with none of those partisan categories. Similarly, 67% of Independents said they owned a device with voice recognition technology in 2015, but only 55% of Republicans and 61% of Democrats reported owning such a device (Monmouth University Polling Institute, 2015). Because no items were asked about income and geographic location in that survey, we can’t know if those differences overlap with differences based on socio-economic status and rural/urban experiences.

Despite this prevalent, basic familiarity with AI technology among Americans, poll data suggest that there is still ample room for Americans to learn more about the issue of AI. Only 12% of Americans said they had read or heard “a lot” about recent developments in the area of AI, while the majority of Americans read or heard either “a little” (53%) or “nothing at all” (35%) about advancements in AI technology (Monmouth University Polling Institute, 2015). Broken down by race, 54% of Hispanics said they had read or heard “nothing at all” about recent developments in AI, and 34% of whites, 30% of Blacks, and 18 % of Asians indicated this. The majority of whites (53%), Blacks (55%), and Asians (71%), as well as about one-third of Hispanics (35%), reported having read or heard “a little” about recent AI developments. In another survey, 2% of Americans reported having made an everyday life decision based on a recent news story on advances in technology (such as AI and nanotechnology) they read or heard about (Pew Research Center, 2017).

Public familiarity with AI in healthcare

In the public health area, when it comes to using technological developments to enhance human features and abilities, Americans are similarly generally not very aware of using AI technologies for such purposes. The existing surveys, however, often ask only about particular applications of AI, which, combined with the breadth of AI applications, makes it difficult to make a judgement about general awareness of AI in healthcare contexts. Asked in 2017 what types of human enhancement technologies they had read or heard about, only 2% of

Americans mentioned “artificial intelligence/AI,” 3% of Americans mentioned “computers/computer technologies,” and 4% of Americans mentioned “robots/robotic body parts” (AARP, 2017). Likewise, in a 2010 convenience sample survey, when asked to name a recent and influential event about AI, only three out of 419 respondents identified projects related to healthcare AI (Grant, Seruwagi, & Dodd, 2011).

A 2018 online survey asked 3,194 adults (1,114 in the U.S. and 2,080 in the UK) about their awareness and use of digital healthcare (Ghafur et al., 2020). We only review findings from the U.S. sample here. Most surveyed Americans were aware of electronic health records, and only about 20% of respondents indicated they did not know whether they had access to their electronic health records. Older respondents and women were more likely to report having access to their electronic health records: 50% of respondents aged 55 years and older reported having such access, compared with only 32% of respondents aged 18-24 reported as such; 48% of women indicated having access to their electronic health records whereas 41% of men did so.

Another 2019 survey of 4,000 U.S. adults on their adoption of digital health technologies found that 44% of Americans had tracked a health indicator or activity using digital tools, either with a wearable device, via a connective device or an app that did not connect to a wearable, or with a digital journal or log. Weight and blood sugar were two of the most digitally tracked measures across health conditions (Day et al., 2019).

Aside from public opinion polls, media coverage provides insights on the salience of AI issues in the public discourse. In a content analysis of coverage of AI in the *New York Times* from 1986 to 2016, researchers found that there has been a sharp increase in the number of articles since 2009 and that discussions have been more positive than negative (Fast & Horvitz, 2017). Among topics, “AI for healthcare” gained attention in the media over the last 10 years, especially on enhancement of health and well-being (Fast & Horvitz, 2017).

Public attitudes toward AI

Overall, a little over half of Americans believe it is important to advance the field of AI in general. In a 2016 poll, 53% of respondents said they felt the quest to advance AI was important, while 20% said it was not necessary, 15% said it was dangerous, 4% said it would “make God angry,” and a remaining 8% reported no answer or not knowing (Vanity Fair/60 Minutes, 2016). Of course, respondents could realistically hold more than one of these beliefs at once, but the forced-choice format of the question didn’t allow for capturing such overlap.

When it comes to general support for AI, surveys indicate consistent demographic differences regarding support, including across gender, education, and income (Zhang & Dafoe, 2019). In a 2018 nationally representative survey, about half of males in the U.S. strongly support (17%) or support (30%) the development of AI, while only about a third of women showed strong support (8%) or support (27%) for developing AI (Zhang & Dafoe, 2019). Education and income are positively associated with support for developing AI (Zhang & Dafoe, 2019). Different racial

groups showed similar support for the development of AI with 43% support among white Americans (12% strong support; 31% support) and 38% among non-white Americans (14% strong support; 24% support) but the survey did not include more nuanced breakdowns by race or ethnicity (Zhang & Dafoe, 2019).

When it comes to attitudes regarding the potential impact of AI, a 2019 Pew survey included open-ended questions that asked respondents what kinds of developments in science they thought would make people's lives better or worse. The survey found that only 8% of surveyed Americans expressed concern over people's lives becoming worse due to developments in AI/robotics/automation (7%) and in self-driving cars (1%), and about 2% of respondents mentioned loss of privacy/freedom being a potential cause for concern. In contrast, 4% of respondents believed that developments in AI/robotics/automation (3%) and in self-driving cars (1%) would make people's lives better (Pew Research Center, 2019a). The data seem to suggest that AI may not be a salient topic on Americans' minds.

For example, when respondents who had a favorable view of the impact of science were asked to list the ways they thought science had had a positive effect on society, only 1% mentioned AI and robotics in their answers to the open-ended question. For those who had negative or mixed views of the impact of science, 1% of them listed AI and robotics as a main way science is having a negative effect on society (Pew Research Center, 2019a). Another example is the perceived low likelihood of harmful consequences of AI, compared to other potential global risks within the next 10 years (Zhang & Dafoe, 2019). Only 16% of Americans thought harmful consequences of AI were very likely or likely to happen. When asked if harmful consequences of AI were to happen, over a third of Americans (36%) perceived the size of negative impacts for several countries or industries as "catastrophic" or "severe," which was relatively moderate compared to perceived negative impacts of other global risks (Zhang & Dafoe, 2019).

In contrast, Americans were much more aware of how little control they had over data privacy (Pew Research Center, 2019b). A large number of Black (60%) and Hispanic (56%) Americans believed that the government tracked their online and cellphone activity, and about four-in-ten white Americans (43%) held the same viewpoint (Pew Research Center, 2019b). Black (39%) and Hispanic Americans (30%) were also more concerned about the extent to which law enforcement may know information about them, which was higher than the concern of their white counterparts (21%; Pew Research Center, 2019b). All racial groups showed relatively deep concerns about the extent to which advertisers or the companies they bought things from knew about their personal information, with about one-in-five not at all concerned (Advertisers: 17.3% white, 20.9% Black, 10.8% Hispanic; Companies that they brought things from: 19.2% white, 22.2% Black, 21.6% Hispanic; Pew Research Center, 2019b).

Public attitudes toward AI development in healthcare

Healthcare seems to be the most prominent area for which people anticipate benefits from AI development and application. In a 2016 survey of 1,021 American adults, respondents were asked which fields were most likely to benefit from advances in AI. Slightly fewer than half

(44%) of all respondents answered “medicine,” topping the list. For comparison, 23% of respondents answered “military science,” 13% answered “automobile manufacturing,” 7% answered “hacking,” 3% answered “psychiatry,” and 3% answered “film making,” with the remaining 6% not knowing or providing no answer (Vanity Fair/60 Minutes, 2016).

While Americans generally have positive views of using AI technology to advance the public health and medical fields, their attitudes become more cautious when it comes to letting AI make important personal health decisions. According to a 2016 survey of 1,021 American adults, only 8% of surveyed respondents felt comfortable letting a computer with AI decide their end of life care (Vanity Fair/60 Minutes, 2016). In another telephone survey, the majority (65%) of respondents thought it would be a change for the worse if lifelike robots become the primary caregivers for the elderly and people in poor health, whereas 28% of respondents thought it would be a change for the better. Furthermore, whites and Blacks seemed to hold equally unfavorable views of letting lifelike robots become the primary caregivers, as 68% of white and 65% of Blacks said it would be a change for the worse. Only 30% of men and 26% of women believed it would be a change for the better. About 32% of Democrats and 30% of Independents thought it would be a change for the better, whereas only 20% of Republicans said so. Respondents who finished high school or less had a relatively more favorable view of the idea, as 32% of them said it would be a change for the better, whereas only 23% of college graduates and 25% of individuals who had some college experience agreed. Fewer people living in the Midwest (24%) and in the South (25%) agreed it would be a change for the better, as compared with people from the West (33%) and the Northeast (32%). The differences in attitudes across age groups and income levels were not noteworthy (Pew Research Center, 2014).

Relative to using AI in hospital-level decision-making and resource allocation, in a 2015 survey of 1,005 Americans, 65% of surveyed Americans said it was a “bad idea” to use robotic nurses with artificial intelligence to diagnose situations and decide when to administer medicine for bed-ridden patients. Only about three in ten thought it was a “good idea,” and the rest either did not know (2%) or thought the answer “depends” (2%). Broken down by demographics, Hispanics (49%) and Asians (44%) had the most favorable view of this healthcare application of AI, followed by Blacks (33%) and whites (27%). A gender gap exists here as well: while 38% of men thought it was a good idea, only 24% of women agreed. Respondents who described themselves as “somewhat liberal” had the most favorable view, as 43% of them approved of this idea, followed by those who were “moderate” (31%), “somewhat conservative” (29%), “very liberal” (27%) and “very conservative” (26%). Although 34% of Democrats and 34% of Independents agreed the application was a good idea, only 25% of Republicans thought so. Finally, individuals who completed 8th grade or less had an especially unfavorable view of this application, with only 7% of them saying it was a good idea, whereas around 23-35% of respondents from each of the other educational attainment groups said it was a good idea (Monmouth University Polling Institute, 2015).

A 2018 survey of 1,114 American adults found that the majority of respondents (52%) were happy to receive advice from their doctors and nurses without assistance from AI, whereas only

38% of respondents were happy to receive AI-assisted health advice. Moreover, respondents aged 44 years and younger were more willing to receive health advice from independent AI than their older counterparts. However, the survey did not ask about ethnic demographic questions (Ghafur, Van Dael, Leis, Darzi, & Sheikh, 2020).

Nearly half of Americans (44%) agreed that accuracy and transparency in AI used for disease diagnosis would likely be an AI governance challenge that impacts large numbers of people in the U.S. in the next ten years, with 10% thinking this would be “very likely; or > 95% of occurring” 13% “likely 80-95%,” and 21% “somewhat likely 60-80%” (Zhang & Dafoe, 2019). Majorities, however, or more than three-quarters of Americans, perceived AI applications in disease diagnosis as a very important (56%) or somewhat important (22%) issue for tech companies and governments to manage carefully. The perceived importance for carefully managing the challenge of AI disease diagnosis is comparable to surveillance (59% “very important” and 19% “important”) and a bit lower than data privacy (64% “very important” and 18% “important”). In terms of each of the three AI governance challenges, white Americans revealed slightly higher perceived importance than did their non-white counterparts (Zhang & Dafoe, 2019).

Public attitudes toward AI applications relevant to COVID-19

Myriad applications of AI have been developed amid the COVID-19 pandemic, such as diagnosis systems based on radiology scans (Ross, 2020), prediction of global spread using airline ticketing data (Niiler, 2020), data-driven decisions about testing locations (McCullom, 2020), and phone-tracking of movement (Marson, Stupp, & Hinshaw, 2020).

The conflict between data use and civil liberty is especially prominent during the pandemic, as the U.S. and other countries across the world deploy phone-tracking applications to try to trace and combat the virus’s spread (Marson et al., 2020). Despite debates over privacy concerns and strategies to combat coronavirus, few recent surveys we found so far have specifically focused on the case of COVID-19. In a 2020 Pew Survey, when asked about governments tracking people’s locations through their cellphones during the coronavirus outbreak, a majority of U.S. adults (60%) thought such actions would not make much of a difference in limiting the spread of the virus (Pew Research Center, 2020). In contrast, only 16% of Americans thought location tracking would “help a lot,” and 20% thought it would “help a little (Pew Research Center, 2020).” Hispanic Americans (50% “help a lot” or “a little”) were relatively more optimistic about the efficacy of location tracking strategies than their white (35% “help a lot” or “a little”) and Black (34% “help a lot” or “a little”) counterparts. Half of Americans with a postgraduate degree thought location tracking would help a lot or a little, which was the highest among education groups (College graduates: 41% “help a lot” or “a little”; Some college: 36%; High school or less 34%). About a third of Republicans (31%) thought the location tracking would help a lot or a little to limit the spread of the virus, holding more negative attitudes than Democrats (46% “help a lot” or “a little”) and Independents (42% “help a lot” or “a little”).

Interestingly, people seem to have nuanced views depending on who is being tracked. When it comes to public acceptance of the government using people's cellphones to track their location, Americans were relatively tolerant of the use for those who have *tested positive for the coronavirus* in order to understand how the virus may be spreading (52% "very" or "somewhat acceptable"), the use for those who may have had *contact with someone who tested positive* for the coronavirus (45% "very" or "somewhat acceptable"), or the use to ensure they are complying with experts' advice on *limiting social contact* during the coronavirus outbreak (37%). Young Americans aged 18 to 29 were less likely to accept government's location tracking on COVID-19 contacts (38%) and for limiting social contact (28%), however, than were Americans more than 30-years old ("very" or "somewhat acceptable" for tracking COVID-19 contacts: 46% - 49%; "very" or "somewhat acceptable" for tracking social contact: 38% - 40%).

The survey also showed partisan differences in acceptance for all three applications of location tracking. Six-in-ten Democrats (61%) said it was very or somewhat acceptable for the government to use people's cellphones to track people who tested positive, while less than half of Republicans (45%) accepted the tracking. The partisan difference persisted regarding acceptance for tracking people who had been in contact with positive COVID-19 cases (Democrats 52%; Republicans 47%) or for limiting social contact (Democrats 44%; Republicans 31%). Additionally, white Americans had the lowest acceptance for all three location tracking strategies among racial groups. Nearly half of white Americans (47%) said it was very or somewhat acceptable for the government to track people who tested positive for coronavirus, compared to the acceptance among their Black (56%) and Hispanic (66%) counterparts. Likewise, less than a third (31%) of white Americans thought the use of location tracking was very or somewhat acceptable to ensure limiting social contact, much lower than Black (45%) Americans and Hispanic Americans (55%).

Many scientists have raised concerns about potential racial discrimination worsened by AI applications related to COVID-19, but here as well we have limited data on Americans' views more broadly. One common issue of inaccurate and biased algorithms is from bias in the training data, with data suffering from demographic unrepresentativeness (McCullom, 2020). For example, the allocation of testing resources was based on information of early cases that skewed to white communities, which further reduced the resources for underserved populations and communities and delayed treatment for those groups, especially for Black Americans (McCullom, 2020).

This problem is not new. In a study published in *Science*, Obermeyer and colleagues (2019) quantified the racial bias in a widely used AI software for follow-up healthcare. They found that white patients were disproportionately prioritized by the algorithms over Black patients, with white patients occupying 82% of the medical center's high-risk group (Obermeyer et al., 2019). However, if the algorithm reflected patients' true sickness, white patients comprised 53%, and Black patients consisted of 47% (Obermeyer et al., 2019). Black patients, a large proportion of whom had lower incomes and inferior insurance compared to the white patients, received fewer follow-up healthcare resources at the same medical center due to the algorithmic bias. We do not know, however, of data on public awareness or attitudes toward such bias in COVID-

19-related AI use, nor how such realities on the ground and news coverage and public discourse on the topic might shape public opinions.

Attitudes toward data privacy and AI in healthcare

As mentioned above, one of the more prominent topics for which there is more data is on attitudes concerning data privacy, including for AI in healthcare. A 2018 online survey of 1,114 American adults found that willingness to share anonymized personal health data largely depended on who the data-receiving body was, with personal doctors being the most trusted data-receiving body (by 61% of Americans), followed by one's family (41%), academic or medical research institutions (26%), and pharmacists (25%) (Ghafur et al., 2020). The least-trusted were tech companies using the data for commercial purposes (4%) and any other commercial companies (who only 3% of Americans trust). These data are particularly interesting because much of the health data is in the hands of private companies. Additionally, 15% of respondents indicated they would not be willing to share their anonymized health information with anyone. The trend suggested that respondents' willingness to share anonymized personal health information decreased as the data-receiving institution became more commercially oriented. This finding was corroborated by another 2019 survey of U.S. adults which found that consumers were least willing to share health data with tech companies whose commercial models relied on it (Day et al., 2019). Respondents aged 35 years and older were less likely to trust any data-receiving institutions than their younger counterparts.

Another 2019 survey of 4,000 U.S. adults on their adoption of digital health technologies found similar results: 73% of respondents indicated they were willing to share their health information with their physicians, followed by their health insurance companies (52%), their pharmacies (48%), research institutions (34%), health tech companies (23%), pharmaceutical companies (19%), government organizations (12%), and tech companies (10%). Older respondents were significantly more willing to share health information with their physicians than younger respondents, whereas younger respondents were more willing to share health information with health tech companies than their older counterparts. Those who had previously downloaded a healthcare app were more willing to share their health information with tech companies (Day et al., 2019).

In a 2019 Pew survey, respondents were asked whether they accept their genetic, physical, and mental health data being used for different purposes (Pew Research Center, 2019b). That survey found that about half of white (51%) and Hispanic (50%) Americans accepted DNA testing companies sharing their customers' genetic data with law enforcement agencies in order to help solve crimes. Black Americans (46%) showed less acceptance than other races. Additionally, educated Black Americans with a college degree (38% "accept," 37% "don't accept," 25% "not sure") showed much lower acceptance than their Black counterparts who attended some college (45% "accept") or high school or less (50% "accept"). There was also a divide between wealthy Black Americans and poor Black Americans, with 36% of those with

income above \$75,000 and 51% of those with income lower than \$30,000 accepted the data sharing for crime-solving.

Regarding makers of a fitness tracking app sharing their users' data with medical researchers seeking to better understand the link between exercise and heart disease, Pew data show that white Americans expressed less acceptance than their Black and Hispanic counterparts (Acceptance: 40% white, 47% Black, 49% Hispanic; Unacceptance: 40% white, 31% Black, 27% Hispanic). Hispanic Americans with some college (58%), bachelor's degree or more (60%), or middle income (\$30,000 - \$74,999; 64%) showed the highest acceptance of sharing data with medical researchers on studying exercise and heart disease (Pew Research Center, 2019b).

Public opinion is divided on social media companies' monitoring its users' posts for signs of depression in order to identify people who are at risk of self-harm and connect them to counseling. The Pew survey presented racial gaps in the acceptance of this application. The majority of white Americans (50%) did not accept, while only a quarter (24%) accepted, and another quarter was not sure (26%). Black Americans showed a bit higher acceptance (41%) than unacceptance (34%). Within the Black American group, those with high school graduation or less showed higher acceptance (44% "accept," 20% "don't accept," 36% "not sure") than others who had attended some college (39% "accept," 45% "don't accept," 17% "not sure") or gained a college degree or more (37% "accept," 43% "don't accept," 20% "not sure"). Likewise, low-income Black Americans (45% "accept," 20% "don't accept," 36% "not sure") were the only group whose acceptance outweighed unacceptance (45% "accept," 24% "don't accept," 31% "not sure"), much higher than their counterparts with middle-income (37% "accept," 48% "don't accept," 14% "not sure") and high-income (36% "accept," 31% "don't accept," 33% "not sure").

Regarding differences across urban and rural residents, there do not appear to be major differences in attitudes. The 2019 Pew survey data did not show a salient urban-rural divide in terms of attitudes toward AI applications related to health data use (Pew Research Center, 2019b). More specifically, about half (49%) of Americans from both metropolitan and non-metropolitan areas accepted the use of DNA testing companies sharing genetic data with law enforcement agencies to help solve crimes. Around four-in-ten Americans from metropolitan (43%) and non-metropolitan (37%) areas said it was acceptable for a fitness tracking app to share data for medical research on heart disease. In addition, about a quarter of Americans from metropolitan (27%) and non-metropolitan (25%) areas accepted the use of social companies' tracking their users' posts to detect signs of depression.

This Pew survey did find some partisan differences, however, in attitudes toward these diverse applications. Democrats or Democratic-leaning were more likely to accept the use of physical data for heart disease research (47%) and social media data for depression detection (31%) than Republicans or Republican-leaning (fitness tracking app: 37% "accept"; social media posts: 23 "accept"). In contrast, Republicans or Republican-leaning (52%) were slightly more accepting of genetic data sharing for crime-solving than Democrats or Democratic-leaning (47%).

A 2019 survey of 4,000 Americans found that most respondents wanted control over their health data, as 82% of respondents expressed desire for controlling who might access their health data. About 81% of respondents wanted to be informed about the kind of health data being collected about them. Older respondents (aged 45 years and above) especially wanted to be informed of the health data being collected about them, as compared with those aged 18-24 years. Those who had previously downloaded a healthcare app expressed less interest in being told what health information was being collected about themselves (Day et al., 2019).

In terms of the personal right to have medical data collected by a health provider about individuals be permanently deleted by these providers, about seven in ten Americans (71%) agreed that all Americans should have the right. White Americans (74%) were more likely to agree that they should have the right to remove personal medical data, compared to their Black (60%) and Hispanic (66%) counterparts. There was a prominent difference within the Black American community, with educated Black Americans who had a bachelor's degree (79%) caring more about their right to delete medical data than those with high school graduation or less (46%). Similarly, high-income (\$75,000 or above) Black Americans (74%) cared more about data deletion rights than their low-income (\$30,000 or lower) counterparts (50%; Pew Research Center, 2019b).

Attitudes toward robot caregivers in healthcare

A 2014 poll (Pew Internet and American Life Project Poll, 2014) of 3267 Americans asked the respondents whether they thought it would be a change for the better or a change for the worse if lifelike robots became the primary caregivers for the elderly and people in poor health. The responses were broken down along demographic lines: gender-wise, slightly more men (30%) than women (26%) thought it would be a “change for the better,” whereas more women (68%) than men (61%) thought it would be a “change for the worse;” 6% of men and 3% of women thought it was “both good and bad.” Other ethnic groups including Hispanics (35%) were more positive than non-Hispanic whites (24%) and non-Hispanic Blacks (26%) that the change would be for the better, while more non-Hispanic whites (68%) and non-Hispanic Blacks (65%) than other ethnic groups (58%) thought it would be a change for the worse. The differences in people’s opinion across age groups, educational attainment and household income were very small. However, there were clear partisan and geographical gaps in terms of how people responded to the survey item: 32% of Democrats and 30% of Independents thought it would be a change for the better, compared with only 20 % of Republicans who said so. In contrast, 73% of Republicans said it would be a change for the worse, compared with 58% of Democrats and 64% of Independents who indicated as such. Finally, respondents from the Midwest and the South were clearly less positive than respondents from the Northeast and the West about a scenario where lifelike robots became the primary caregivers for the elderly and people in poor health, as only 24% of Midwesterners and 25% of Southerners said it would be a change for the better, compared with 32% of Northeasterners and 33% of Westerners who said so. More Midwesterners (72%) and Southerners (67%) said it would be a change for the worse than Northeasterners (59%) and Westerners (57%).

Another 2015 survey (Monmouth University Polling Institute, 2015) of 1,010 Americans asked the respondents whether they thought it was a good or bad idea to use AI for “robotic nurses for bed-ridden patients that can diagnose situations and decide when to administer medicine.” Overall, about two in three respondents thought it was a bad idea, although it was not without demographic nuances: Asians and Hispanics were much more positive about the AI application than whites and Blacks, as 44% of Asians and 49% of Hispanics said it was a good idea, compared with only 27% of whites and 33% of Blacks. 53% of Asians and 49% of Hispanics thought it was a bad idea, whereas the overwhelming majority of whites (68%) and Blacks (63%) said so.

There was also a clear gender gap, too, and men were more favorable toward the idea of using AI for robotic nurses than women: 38% of men said it was a good idea, whereas only 24% of women indicated as such; 57% of men said the application was a bad idea, compared with 73% of women who said so. When breaking down the data along lines of educational attainment, it became clear that there was a gap between those whose highest grade completed was 8th grade or less and all other respondents: only 7% of those who only had 8th grade or less said the application was a good idea, while for each of the other groups, about 23-35% of respondents from each group said it was a good idea. An overwhelming majority (79%) of those who completed 8th grade or less said the application was a bad idea, whereas for each of the other groups, only 61-70% of the group members thought it was a bad idea (except for the “don’t know/ refuse to answer” group of which half said the application was a bad idea). This could be, however, purely a side effect of the small sample size of the “8th grade or less” group (as well as the “don’t know/ refuse to answer” group which only had 4 respondents) since only 14 people fell under this category. Finally, respondents who described themselves as “somewhat liberal” were notably more favorable of the application, as 43% of them said it was a good idea, whereas for each of the other groups only around 26-31% of respondents thought it was a good idea.

A few surveys have focused specifically on experts’ views of AI, which we did not summarize here. For further reading and an annotated bibliography of these surveys, please see the *Further Reading* section at the end of this report.

Summary - Areas for future public opinion work on AI

In this section, we identify broad areas that need to be filled in with more and better public opinion research, and describe how future surveys can help collect public inputs and generate insights necessary for policymaking and public engagement efforts on AI in the context of public health.

While representative surveys at the national and global levels can be a useful tool for understanding views of large diverse publics, unfortunately, we have limited survey data on public views of AI to date, with especially little work focused on the key action areas of concern. Few studies exist focused on public opinion of AI in the U.S. and those that do tend to focus on Americans more broadly, without breakdowns, focus, or oversampling of different publics

based on demographic and geographic distinctions. Polls also tend to focus on specific, sometimes very hypothetical or obscure-seeming potential applications of AI, making it difficult to piece together an overall picture of public opinion regarding AI in general and in healthcare in particular.

Some broader trends do appear, however. Despite low overall levels of familiarity with AI as a concept or through specific applications, many Americans do have awareness and concern about lack of control over how their data is collected and used, and by whom. Polls also indicate that most Americans, regardless of their gender or race, are hesitant to think that AI with “true” intelligence or decision-making power in healthcare contexts would be a good idea. Women tend to be less optimistic than men about the benefits of AI applications, despite equal levels of use of smart phones with AI voice recognition capabilities and other general computer and technology use in daily life.

Across surveys, the trends by race or ethnicity are also less clear and vary by application and context. We identify preliminary, general trends, however, where Hispanics and Asians hold slightly more positive attitudes toward applying AI technology in public healthcare, whereas whites and Blacks have less positive attitudes. When it comes to AI and data privacy, different AI applications seem to trigger different responses across racial groups. For example, whites have lower acceptance of social media companies and more concerns about insurance providers using AI technology and personal data in predicting signs of depression, whereas Blacks show lower acceptance and more concerns about DNA testing companies sharing customers’ genetic data with law enforcement agencies to solve crimes. Educational attainment appears to be positively associated with individuals’ awareness of AI applications in healthcare, their perceptions of the likelihood that AI will help improve individuals’ health, and their concerns over using personal data in various AI healthcare applications.

There is also a lack of survey data on how diverse publics perceive the potential risks and benefits that come with AI applications in various healthcare domains. For example, few of the survey items we found tapped into public views of the benefits and risks of using AI in triage decision-making; nor did they allow us to assess how the public balances collective and personal health against concerns about privacy violations when it comes to the use of AI in contact tracing to combat COVID-19. The few items we found that touched on risk perceptions of AI focused on non-healthcare contexts such as human jobs being eliminated by AI in the future, and AI eventually posing a threat to the existence of the human race (Gallup Organization, 2017; Monmouth University Polling Institute, 2015). This lack of data significantly limits our ability to identify areas of AI development in the healthcare domain that the public values and would hope to see expanded, as well as fields that could evoke strong public concerns and should thus be developed with great caution (if at all).

Moreover, questions related to AI usage and applications during and after the COVID-19 pandemic are lacking, despite the fact that the pandemic has further expanded both research and applications advances and attention to the area of AI in healthcare. As we have reviewed in our legal implications and policies section above, four broad areas of pandemic preparedness

including disease surveillance, improved diagnostics and triage, vaccines, treatment, and equipment development, and ongoing healthcare see especially promising AI uses at present and in the near future. Future public opinion surveys should therefore be designed with these areas in mind to help policymakers determine how the public understands the implications associated with these applications and which applications might be more prone to controversies.

Although questions about responsibility and rights involved in data collection and usage are becoming increasingly important, these are reflected in only a few existing public opinion surveys. In addition, the current laws do not apply in contexts where much of the data used to inform AI applications are collected (such as through cellphone applications and consumer buying behavior), posing legal gaps in responsible development of AI. Though experts frequently mentioned potential inequalities caused by AI in healthcare (Pew Research Center, 2018), the existing surveys rarely tapped into the public's views on these issues. Furthermore, the general public are likely to rely on their trust in different institutions and social groups to make decisions about complex science issues such as AI. Little of the survey research that we found, however, included items measuring the American public's trust levels in institutions and social groups involved in AI development (although a few of them assessed trust in different health data-receiving institutions). To fill in these gaps, future research should tap into the public's views regarding how, why, and for and by whom personal health data should be collected and used, as well as the public's trust levels in different agencies and social institutions guiding the development of AI.

Finally, a systematic approach to designing survey questions to collect public inputs on the stages of the AI system lifecycle, namely views and awareness related to 1) design, data collection, and model building, 2) verification and validation, 3) deployment, and 4) operation and monitoring of AI systems, could be valuable, as each of these stages can have unique implications and need for public engagement. Understanding the variation in views across demographic differences and experiences with AI and with science and institutions more broadly, however, will also be vital to successful engagement. However, existing surveys are rarely designed with these phases in mind, nor with a focus on being able to understand views across diverse publics.

Future public opinion surveys can help us understand the unique opportunities and challenges underlying each of these stages of the AI lifecycle, if adopting such a systematic approach. Last but not least, future surveys may also pay special attention to minority groups and groups that are traditionally digitally marginalized as these are the groups that are more susceptible to negative outcomes of AI usage and applications (for example, racial bias embodied in AI systems). Survey items gauging the AI digital divide among the U.S. public--issues related to algorithmic literacy, AI awareness, understanding, access, affordability, availability, and adaptability--would be especially helpful for filling the gaps in AI fairness and equity and how we want to use AI, or not, to address the COVID-19 pandemic and other issues we face going forward.

Further Reading

Ethics, Human Rights, and AI

Bradley, C., Wingfield, R., & Metzger, M. (2020). *National artificial intelligence strategies and human rights: A review*. Retrieved from <https://cyber.fsi.stanford.edu/gdpi/content/national-artificial-intelligence-strategies-and-human-rights-review>

The authors review existing National AI Strategies (NASs), with a specific focus on how these documents have incorporated human rights. An NAS is a “strategy, ordinarily developed by a government (although there are examples of governments adopting or endorsing NASs developed by other organizations) which sets out its broad approach to AI, specific areas of focus, and activities that it will undertake which relate to AI” (p. 5).

While many of the NASs mention human rights, “the depth of subsequent engagement in the human rights framework and its application to AI varied considerably” (p. 21). Of those that mention do human rights, direct engagement with the issues of privacy and equality or non-discrimination are most frequent. Other rights mentioned included the right to an effective remedy, rights to freedom of thought, expression and access to information, and the right to work. The authors recommend that governments:

- Include human rights explicitly and throughout the strategy
- Outline specific steps to be taken to ensure human rights are protected
- Build in incentives or specific requirements to ensure rights-respecting practice
- Set out grievance and remediation processes for human rights violations
- Recognize the regional and international dimensions to AI policy
- Include human rights experts and other stakeholders in the drafting of NASs

Donahoe, E., & Metzger, M. M. (2019). *Artificial intelligence and human rights*. *Journal of Democracy*, 30(2), 115-126. doi:10.1353/jod.2019.0029

The authors argue for a human rights-based approach to the governance of AI as an alternative to many current guidelines that are grounded in ethics. They argue that “human-rights considerations must make themselves felt in the design, application, and evaluation of AI products, as well as in all government procurement and deployment of AI. This can increase the likelihood that the massive potential of AI will benefit societies, without causing harm to the human beings who compose them” (p. 123). The authors outline four benefits of a human rights approach that draws on the framework of the Universal Declaration of Human Rights (UDHR):

- Human rights are focused on human impacts
- Human rights principles comprise many of the ethical issues raised by AI
- Human rights frameworks apply to both governments and to the private sector
- International human rights laws such as the UDHR are established and recognized

Examples of rights from the UDHR that are relevant to AI:

- Article 2: The right to equal protection and nondiscrimination (e.g., bias in data and fairness in decision-making).
- Article 3: The right to life and personal security (e.g., autonomous weapons)
- Articles 8 –11: The right to effective remedy for violations and infringements of rights and the right to due process of law (e.g., transparency, fairness, accountability)
- Article 12: The right to privacy (e.g., data privacy)
- Article 19: The right to freedom of expression (e.g., algorithm-based content)
- Articles 20 and 21: The rights to association and peaceful assembly and the right to democratic participation in government (e.g., human agency)
- Articles 23 and 25: The rights to work and to enjoy an adequate standard of living (e.g., job loss and displacement)

Morley, J., Machado, C., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2019). The debate on the ethics of AI in healthcare: a reconstruction and critical review. SSRN. doi:10.2139/ssrn.3486518

The authors conducted a literature review to identify ethical issues related to applications of AI in a healthcare context. Importantly, they focused on those that are *unique* to the use of AI, not just healthcare more broadly. They identified six broad ethical concerns related to:

- *Inconclusive evidence*: soundness and causality in diagnosis
- *Inscrutable evidence*: lack of transparency or clarity in the data
- *Misguided evidence*: reliability of data and concerns about bias
- *Unfair outcomes*: justice and fairness in who receives treatment
- *Transformative effects*: privacy and loss of autonomy in decision-making
- *Traceability*: accountability and responsibility for AI decisions

The authors also explored how each issue varies by level: individual, interpersonal, group, institutional, sectoral, societal. They note that “protecting people from the harms of AI-Health goes beyond protecting data collection and applying a valid model. Normative frameworks need to contemplate the complexities of the human interactions where these technologies will be introduced and their emotional impacts” (p. 18).

Raso, F., Hilligoss, H. H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). Artificial intelligence & human rights: opportunities & risks. Berkman Klein Center Research Publication.

The authors explore the positive and negative human rights considerations for six different applications of artificial intelligence: criminal justice (risk assessments), finance (credit scores), healthcare (diagnostics), content moderation (standards enforcement), human resources (recruitment and hiring), and education (essay scoring).

Human rights are a useful framework for AI because rights “provide an agreed set of norms for assessing and addressing the impacts of the many applications of this technology, while also providing a shared language and global infrastructure around which different stakeholders can engage” (p. 8).

There are a number of human rights that pertain directly to the use of AI for medical diagnostics. These include:

- Right to Life, Liberty, and Security of Person—increasing access to high-quality medical diagnostics
- Right to Privacy—concerns about individual privacy from data collection
- Right to Desirable Work—could reduce the number of individuals excluded from work for medical reasons
- Right to Adequate Standard of Living—early detection and diagnosis leading to better quality of life
- Right to Education—could reduce the number of individuals excluded from education for medical reasons

Expert attitudes toward the development of AI in healthcare

Lai, M. C., Brian, M., & Mamzer, M. F. (2020). Perceptions of artificial intelligence in healthcare: Findings from a qualitative survey study among actors in France. *Journal of Translational Medicine*, 18(1), 13.

This study interviewed stakeholders about applications of AI in healthcare in France, including health professionals and researchers, industrial partners, regulatory agencies, and individuals with no conflict of interest. The diverse stakeholders shared many concerns about AI and saw the necessity of defining responsibilities to guide AI-based health applications, though their focuses were slightly different based on their positions.

- Industrial partners and researchers were highly committed to the development of AI tools for health practitioners
- While industrial partners were concerned about legal difficulties to access individual data of employing health applications, health researchers and members of regulatory agencies viewed themselves as having important roles in developing the regulation of these applications to guarantee the considerations of public values
- Health practitioners showed enthusiasm for useful and safe practice for their patients, however, they were cautious about ethics and the evidence of efficiency
- Individuals outside of the field of health, also raised more concerns about potential risks, social justice, and data protection.

Pew Research Center (2018). *Artificial Intelligence and the Future of Humans*.

Pew Research Center interviewed experts about how they view the consequences of AI applications, including those in the field of public health.

- Optimistic views
 - Healthcare and the many possible applications of AI in diagnosing and treating patients or helping senior citizens live fuller and healthier lives.
 - Contributing to broad public-health programs built around massive amounts of data that may be captured in the coming years about everything from personal genomes to nutrition.
 - Emergency healthcare and aid work are areas where AI can be applied to benefit the majority.
- Concerns
 - *Human agency*. Decision-making may become more and more dependent on code-driven, “black box” processes which people have little control over or knowledge of.
 - *Data abuse*. Most AI tools are in the hands of for-profit companies without values or ethics being built in the systems. In addition, these globally networked systems can be difficult to regulate.
 - *Increased inequity*. The high efficiency and other economic advantages of AI tools may disrupt all aspects of human work, resulting in mass job losses. Use of AI systems may also worsen economic and healthcare inequity, widening divides between “the haves and have-nots” and bringing about social upheavals.
 - *Reduction of cognitive, social and survival skills*. Reliance on AI tools may erode human abilities to think for themselves, operate autonomy, and carry out effective social interactions.
 - *Increased vulnerabilities*. Autonomous military applications may result in much greater life losses. Use of weaponized information, lies and propaganda may dangerously destabilize human groups.
- Solutions:
 - Improve human collaboration across borders and stakeholder groups.
 - Develop policies to assure that development of AI will be directed at augmenting humans and the common good.
 - Shift the priorities of economic, political and education systems to empower individuals to stay ahead in the “race with the robots.”

References

- AARP. (2017). AARP poll: 2017 human enhancements survey (Version 2) [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research. doi:10.25940/ROPER-31115174
- Al-Almoudi, I., & Latsis, J. (2019). Anormative black boxes: artificial intelligence and health policy. In I. Al-Amoudi & E. Lazega (Eds.), *Post-human institutions and organizations: confronting the Matrix* (pp. 119-142). London: Routledge.
- Beduschi, A. (2020). Human rights and the governance of artificial intelligence. *Geneva Academy of International Humanitarian Law and Human Rights*. Retrieved from <http://hdl.handle.net/10871/41078>
- Bradley, C., Wingfield, R., & Metzger, M. (2020). *National artificial intelligence strategies and human rights: a review*. Retrieved from: [https://www.gp-digital.org/wp-content/uploads/2020/04/National-Artificial-Intelligence-Strategies-and-Human-Rights%E2%80%9494A-Review .pdf](https://www.gp-digital.org/wp-content/uploads/2020/04/National-Artificial-Intelligence-Strategies-and-Human-Rights%E2%80%9494A-Review.pdf)
- Calo, R. (2017). Artificial intelligence policy: a primer and roadmap. *University of California, Davis, Law Review*, 51, 399-435. doi:10.6092/issn.2531-6133/8670
- Castro, D., & Johnson, A. (2020). *The most significant AI policy developments in the United States in 2019*. Retrieved from: <https://www.datainnovation.org/2020/02/the-most-significant-ai-policy-developments-in-the-united-states-in-2019/>
- Chander, A. (2016). The racist algorithm. *Mich. L. Rev.*, 115, 1023. Retrieved from <https://repository.law.umich.edu/mlr/vol115/iss6/13>
- Day, S., Seninger, C., Kothari, M., Fan, J., Pundi, K., Perino, A., Turakhia, M., McGuinness, D., Zweig, M., & Evan, B. (2019). *Digital health consumer adoption report 2019*. Retrieved from <https://rockhealth.com/reports/digital-health-consumer-adoption-report-2019/>
- Davies, S. E. (2019). Artificial intelligence in global health. *Ethics & International Affairs*, 33(2), 181-192. doi:10.1017/S0892679419000157
- De-Arteaga, M., Herlands, W., Neill, D. B., & Dubrawski, A. (2018). Machine learning for the developing world. *ACM Transactions on Management Information Systems (TMIS)*, 9(2), 1-14.
- Donahoe, E., & Metzger, M. M. (2019). Artificial intelligence and human rights. *Journal of Democracy*, 30(2), 115-126. doi:10.1353/jod.2019.0029
- Executive Order 13859: *Maintaining American leadership in artificial intelligence*, 13859 C.F.R. (2019).

Fast, E., & Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. AAAI 2017. Retrieved from <https://arxiv.org/abs/1609.04904>

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). doi:10.1162/99608f92.8cd550d1

Ghafur, S., Van Dael, J., Leis, M., Darzi, A., & Sheikh, A. (2020). Public perceptions on data sharing: Key insights from the UK and the USA. *The Lancet Digital Health*. doi:10.1016/S2589-7500(20)30161-8

Grant, A. W., Seruwagi, L. Z., & Dodd, M. S. (2011). *Artificial Intelligence Through the Eyes of the Public*. Retrieved from <https://digitalcommons.wpi.edu/iqp-all/153/>

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399. doi:10.1038/s42256-019-0088-2

Lai, M. C., Brian, M., & Mamzer, M. F. (2020). Perceptions of artificial intelligence in healthcare: Findings from a qualitative survey study among actors in France. *Journal of Translational Medicine*, 18(1), 13. doi:10.1186/s12967-019-02204-y

Latonero, M. (2018). *Governing artificial intelligence: Upholding human rights and dignity*. Retrieved from <https://datasociety.net/library/governing-artificial-intelligence/>

Lewis, J. E. (2020). *Indigenous protocol and artificial intelligence position paper*. Retrieved from Honolulu, Hawaii: <https://www.indigenous-ai.net/position-paper/>

Malgieri, G., & Niklas, J. (2020). Vulnerable data subjects. *Computer Law & Security Review*, 37, 105415. doi:10.1016/j.clsr.2020.105415

Marson, J., Stupp, C., & Hinshaw, D. (2020). U.S. and Europe Turn to Phone-Tracking Strategies to Slow Spread of Coronavirus. Retrieved from <https://www.wsj.com/articles/u-s-and-europe-turn-to-phone-tracking-strategies-to-halt-spread-of-coronavirus-11585906203>

McCullom, R. (2020). *Artificial intelligence, health disparities, and Covid-19*. Retrieved from <https://undark.org/2020/07/27/ai-medicine-racial-bias-covid-19/>

Monmouth University Polling Institute. (2015). Monmouth University national poll: March 2015 (Version 2) [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research. doi:10.25940/ROPER-31113655

Morley, J., Machado, C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2019). The debate on the ethics of AI in health care: a reconstruction and critical review. *SSRN*. doi:10.2139/ssrn.3486518

Niiler, E. (2020). An AI Epidemiologist Sent the First Warnings of the Wuhan Virus. Retrieved from <https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/>

NSCAI. (2019). *Initial report*. Retrieved from: <https://epic.org/privacy/NSCAI-initial-report-073119.pdf>

NSCAI. (2020a). *Mitigating economic impacts of the COVID-19 pandemic and preserving U.S. strategic competitiveness in artificial intelligence*. Retrieved from: <https://drive.google.com/file/d/1vSRRfqV6S4xGoMxueC-CMG0tYYTGa1Ly/view>

NSCAI. (2020b). *Privacy and ethics recommendations for computing applications developed to mitigate COVID-19*. Retrieved from: <https://drive.google.com/file/d/1m0AT21dS2XJ6JIGMgo7SuLSLveWIO8WK/view>

NSCAI. (2020c). *The role of AI technology in pandemic response and preparedness: recommended investments and initiatives*. Retrieved from: https://drive.google.com/file/d/153DUHToD4zoM_GXe9MWGNKzend7Tsl2o/view

OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. Retrieved from: <https://www.fsmb.org/siteassets/artificial-intelligence/pdfs/oecd-recommendation-on-ai-en.pdf>

OECD. (2020). *OECD principles on AI*. Retrieved from <http://www.oecd.org/going-digital/ai/principles/>

Pew Research Center. (2014). Pew Research / Smithsonian -- Science Futures Survey February 2014 (Version 2) [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research. doi:10.25940/ROPER-31096266

Pew Research Center. (2017). Pew Research Center: June 2017 science and news survey (Version 2) [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research. doi:10.25940/ROPER-31114469

Pew Research Center. (2018). Artificial intelligence and the future of humans. Retrieved from https://www.elon.edu/docs/e-web/imagining/surveys/2018_survey/AI_and_the_Future_of_Humans_12_10_18.pdf

Pew Research Center. (2019a). Pew Research Center's American trends panel poll January 2019 wave 42 (Version 2) [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research.

Pew Research Center. (2019b). *Americans and privacy: Concerned, confused and feeling lack of control over their personal information*. Retrieved from <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>

Pew Research Center. (2020). *Most Americans don't think cellphone tracking will help limit COVID-19, are divided on whether it's acceptable*. Retrieved from <https://www.pewresearch.org/fact-tank/2020/04/16/most-americans-dont-think-cellphone-tracking-will-help-limit-covid-19-are-divided-on-whether-its-acceptable/>

Rainie, L., Anderson, J., & Page, D. (2017). Code-dependent: Pros and cons of the algorithm age. *Pew Research Center*, 8.

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). Artificial intelligence & human rights: opportunities & risks. *Berkman Klein Center Research Publication*. Retrieved from <http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439>

Rosenblat, A., Wikelius, K., boyd, d., Gangadharan, S. P., & Yu, C. (2014). *Data & civil rights: health primer*. Paper presented at the Data & Civil Rights Conference. Retrieved from: <https://datasociety.net/wp-content/uploads/2014/10/Health.pdf>

Ross, C. (2020). As Covid-19 surges, researchers amass lung scans to build AI systems for patient care. Retrieved from <https://www.statnews.com/2020/07/27/covid19-ai-researchers-amass-patient-lung-scans/>.

SCIMEP. (2020). U.S. Public Attitudes on Artificial Intelligence. Unpublished data.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, version 2*. Retrieved from <http://standards.ieee.org/develop/indconn/ec/autonomoussystem.html>

The White House. (2020). *Artificial intelligence for the American people*. Retrieved from: <https://www.whitehouse.gov/ai/>

The White House, & Office of Science and Technology Policy. (2020). *American artificial intelligence initiative: year one annual report*. Retrieved from: <https://www.whitehouse.gov/wp-content/uploads/2020/02/American-AI-Initiative-One-Year-Annual-Report.pdf>

United Nations. (2020). *Human Rights Law*. Retrieved from <https://www.un.org/en/sections/universal-declaration/human-rights-law/index.html>

Vanity Fair/60 Minutes. (2016). 60 Minutes/Vanity Fair Poll: March 2016 (Version 2) [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research.

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. Retrieved from London: <https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf>

Yeung, K., Howes, A., & Pogrebna, G. (2020). AI governance by human rights-centred design, deliberation and oversight: an end to ethics washing. In M. Dubber & F. Pasquale (Eds.), *The Oxford handbook of AI ethics*: Oxford University Press.

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. Oxford, UK: Center for the Governance of AI, Future of Humanity Institute, University of Oxford.