



Defining Ethical Guidelines for Augmented Intelligence: 2019 Workshop Report

Lindsey Jarrett, PhD

Matthew Pjecha, MS

Introduction

What is artificial intelligence?

At the simplest level, artificial or augmented intelligence (AI) is a branch of computer science focused on the simulation of intelligent behavior in computers, thereby enabling computers to make decisions like or better than humans. AI uses machine learning tools to train algorithms to identify subtle patterns in massive data sets, patterns so subtle that humans can't detect them on their own. The training process teaches the algorithm to use knowledge from those patterns to make decisions.

Until recently, ethics and AI in healthcare have rarely been addressed together. This is peculiar given that the most famous code of ethics, the Hippocratic Oath, has guided doctors for centuries. With emerging AI technologies poised to play an important role in healthcare, we have the opportunity now to leverage its strengths and limit its potential risks.

Ethical Risks of AI in Healthcare

Machine learning techniques can be used to produce algorithms that are sensitive to subtle and complex patterns found in large data sets. These AI algorithms can then be used to make clinical predictions and determinations that human providers would otherwise be unable to make. While this presents a significant opportunity to use these AI tools to improve the quality of care, it can also present certain ethical risks.

AI tools have already proven better than unassisted humans at detecting some conditions and providing more timely and accurate diagnoses, and the FDA has already begun approving some of these AI systems for clinical use. AI can interpret data in electronic medical records and identify patients at risk

for hospital-acquired infections or match them with available clinical trials. In the future, it could provide access to care for thousands who currently lack it. It could enhance evidence-based decision making. It could recommend individualized treatment plans based on clinical need. It could save time and money, enabling clinicians to devote more time to humanistic aspects of care. By using metrics for state of mind and mood, there is even evidence that AI could be a game-changer in mental health care.

The ethical risks develop at the intersection of three factors. The first factor is commonly known as the black box problem; because AI systems are designed to be sensitive to patterns that are often too complex or subtle to otherwise detect, it can be difficult to observe or explain how a given AI system arrives at a determination based on a given set of inputs. This becomes a problem when the determinations are aberrant or otherwise unexpected.

The second factor is the capacity for AI systems to reproduce systemic biases found in the data sets used to train them. Algorithms used in criminal justice have been shown to produce systematically different outputs for black and white subjects. Algorithms used in recruitment have been shown to penalize applications submitted by women. Algorithms used in facial recognition have been shown to fail more frequently for users of some ethnicities. Many of these biases are consonant with systemic biases in our society, running the risk of deepening existing inequities.

The third factor is a cultural association of computers and quantitative systems with increased objectivity. In some cases, AI systems are being developed to automate processes that are otherwise handled by humans. Algorithms are already being used to allocate Medicaid benefits in some states. One argument for this automation is an attempt to make these processes more objective, by removing the subjective perspective of the human. Increasing the perceived objectivity of a process makes it more authoritative, making it more resilient to outside questioning. While this is not necessarily a problem unto itself (in fact, many may consider it a benefit of AI systems), it can undermine attempts to question or object to problematic outputs produced by AI systems.

Taken together, these three factors present the risk of AI systems producing problematic outputs that are likely to deepen existing societal inequities while simultaneously being positioned as more authoritative actors than the humans they replace. Yet there is promise in intelligence: adding convenience, accessibility and potentially removing barriers. AI will soon be pervasive in healthcare. How do we embrace this inevitable shift while navigating the accompanying uncertainty? As stakeholders, we can agree we need guidelines for safety, but what about other things like fairness and accountability?

Institutions across the private, public and government sectors that are creating tools in healthcare technology are now, arguably more than ever, at an important intersection. The path they take will determine their success as their users are demanding products to not just be efficient and effective, but also check ethical boxes such as fairness, justice, safety, and inclusion. Unfortunately, in the current state these ethical requests are not clearly defined and vary across sectors and platforms for numerous reasons that are still being unpacked by those researching this intersection. For example, companies are evaluating the need for ethical consideration in their products and processes. Internal governance structures at most technology companies are failing to ensure accountability for AI systems, possibly due to the lack of structures that go beyond ethics guidelines, inadequate representation across

advisory boards, and the negligence of transparency and audit efforts (Whittaker et al., 2018). AI is a broad term and spans across several technological domains, touching our everyday lives in how we communicate, how we work, and how we mobilize our bodies and minds. All of these applications present risks and opportunities. Specifically, this convergence of technology and healthcare could dramatically increase access to care, more timely diagnosis, support evidence-based decision making, and individualized treatment plans. It is amidst these incredible improvements on the horizon that we must responsibly handle ethical risks as developers seek guidance when creating AI solutions.

Methods

In 2018, recognizing the need to leverage AI's benefits and mitigate risks, the Center for Practical Bioethics (CPB)*, in collaboration with Cerner Corporation** and other leading healthcare institutions, started working together to develop ethical AI strategies tailored to the unique needs of healthcare. In 2019 these organizations designed a one-day workshop to convene stakeholders across healthcare and technology to lay the groundwork for ethical AI strategies.

The workshop had three main objectives:

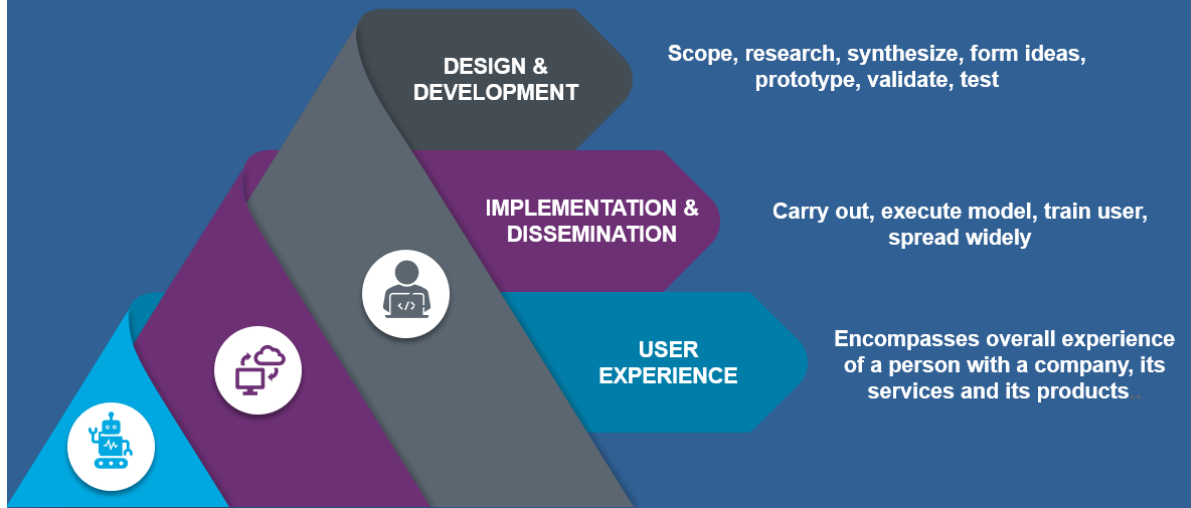
- 1) For individuals to share what each of six ethics keywords looks like from their professional and personal perspective. These keywords were fairness, inclusion, accountability, transparency, data privacy & security, and reliability & safety.
- 2) For each group to discuss and record what each keyword should look like across design/development, implementation/dissemination and user experience pillars of healthcare intelligence.
- 3) To achieve consensus on recommendations for incorporating ethics across the pillars of healthcare intelligence.

The workshop recruited stakeholders representing three broad categories or “pillars.”

- **Design and Development** – Engineers, researchers, computer scientists, informatics specialists and others who create AI tools.
- **Dissemination and Implementation** – Clinicians, providers, payors, and others who use AI tools.
- **Patients and Community** – Those who would be subject to (and potentially at risk from) AI tools, from patients, their advocates, and other beneficiaries to corporate and governmental policymakers.

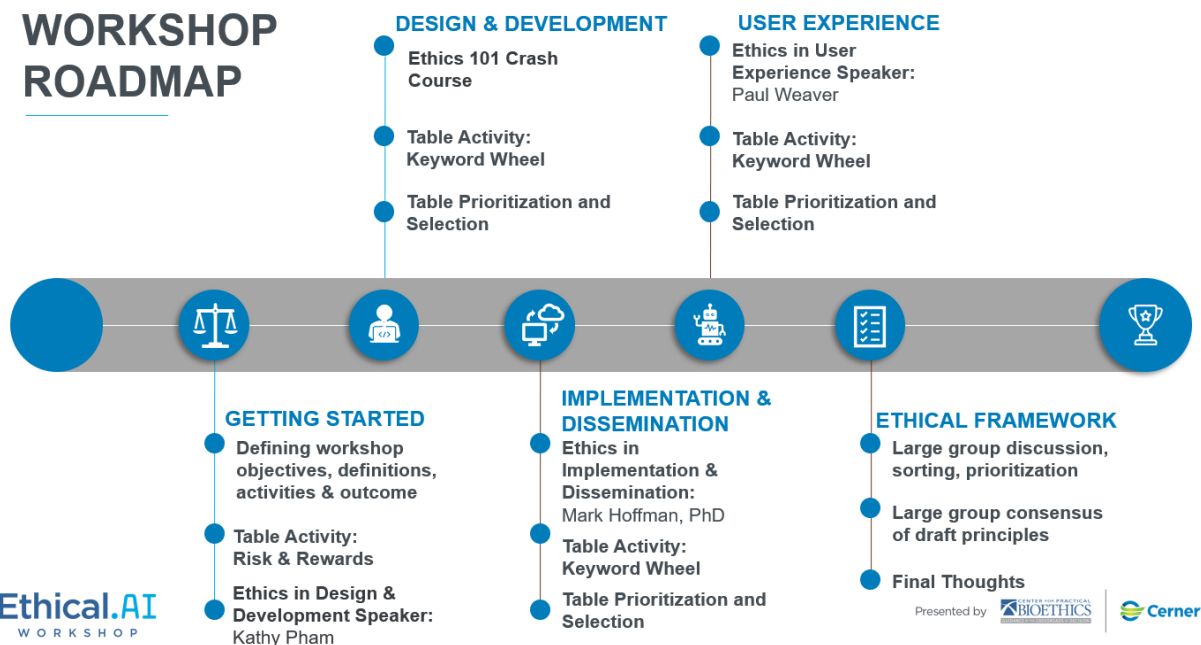
These individuals have experience working across the pillars as they are common phases of AI Tool creation and can be found organized in the following way across the healthcare technology sector.

PHASES OF AI TOOL CREATION



By gathering stakeholders with interdisciplinary expertise, we created an opportunity to look at various concerns and propose solutions that consider the impact of AI tools in healthcare. The process included a foundation of epistemic humility, or an agnostic base that keeps the group humble with our own assumptions. The roadmap of the one-day workshop is outlined below.

WORKSHOP ROADMAP



Stakeholders received materials in advance of the workshop to lay the groundwork for their participation. The workshop was organized into 10 workgroups/tables, each composed of individuals from all three pillars. Workshop founders and facilitators social engineered the tables prior to the workshop to ensure diverse experience and backgrounds at each table. The groups were not engineered

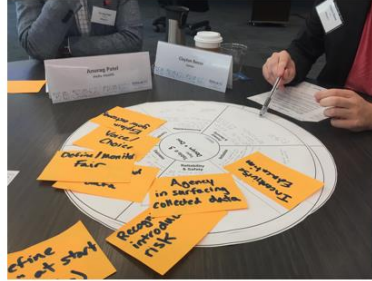
by demographic factors, such as race, ethnicity or gender, though, as this data was not collected in the initial workshop. Workgroups began by asking members to identify their professional roles and experiences. The goal of this exercise was to recognize points of view not represented, to strive to accommodate those perspectives in their absence, and to commit to including them in future work.

The workshop contained 4 short “TED talk style” presentations by leaders in the healthcare and technology industries to cover the breadth of work across Design and Development (Kathy Pham), Implementation and Dissemination (Mark Hoffman, PhD) and User Experience (Paul Weaver). These particular talks were vital to introducing and defining the pillars, especially since the participants varied in expertise and knowledge. Additionally, to prepare participants for conversations related to ethics, Ryan Pferdehirt, D. Bioethics, HEC-C, a clinical ethicist, presented an “Ethics Crash Course.” In it, he introduced the four principles of bioethics (i.e., justice, autonomy, beneficence and nonmaleficence) and the concept of applied ethics (i.e., applying the four principles to real-world situations). Ryan then explained how to incorporate the six keywords (i.e., fairness, inclusion, accountability, transparency, data privacy and security, and reliability and safety) into an ethical decision-making model for AI in healthcare. This was intended to lay a foundation for ethical reflection and standard definitions for participants to leverage throughout the day.

Ethical.AI Workshop Speakers

			
Computer Scientist & Product Leader Harvard and Mozilla	Director of Membership & Ethics Education Center for Practical Bioethics	Chief Research Information Officer Children's Mercy Hospital	VP of User Experience & Human Factors Research Cerner
Kathy Pham, MS	Ryan Pferdehirt, D. Bioethics	Mark Hoffman, PhD	Paul Weaver

The activities centered around collaborative discussion and the data collection process was repeated three times for each pillar (i.e., Design/Development, Dissemination/Implementation, and Patients/Community). After a short overview of the pillar from the speakers outlined above, workgroup members were instructed to express what each keyword would look like from their perspective. For example, a member might say, “In my role as an engineer, transparency should look like AI systems in healthcare ought to be explainable and understandable.” Finally, each of the 10 workgroups were asked to identify their group’s top two statements for each keyword within each pillar, with an emphasis on those statements most relevant and critical to healthcare in Greater Kansas City.



Additionally, an associate of Mozilla Foundation participated in the event, and collaborated with the workshop founders on interactive modules that participants could visit during breaks and lunch. The artifacts pictured below were located on each table to provide information for the modules located in the hallway of the event space. The demos showcased were created by Mozilla's Creative Media Award winning web-based simulations (*Survival of the Best Fit*, *A Week with Wanda*) and interactive film (*Stealing Ur Feelings*) that can be found here: <https://blog.mozilla.org/blog/2018/10/24/keeping-ai-accountable-with-science-fiction-documentaries-and-doodles-plus-225000/>.

Findings



 1 DAY August 2019	 54 Participants	 10 Tables
4 Industry Leaders	 <ol style="list-style-type: none"> 1. Ethics 101 2. Design & Development 3. Implementation & Dissemination 4. User Experience 	
7 Facilitators	Themes <ol style="list-style-type: none"> 1. Biases & Diversity 2. Auditability 3. Feedback Loop/ Process 4. Limitations in models 5. Humans and AI 6. Data Use/Ownership 	
 360 Sticky Notes		

Ethical.AI WORKSHOP Presented by  

In August 2019 at Cerner Corporation in North Kansas City, MO, fifty-four professionals across numerous fields of healthcare and technology from the US, including engineering, medicine, social work, research, data science, user experience, nursing and other related fields gathered to examine various concerns regarding ethics in AI and propose solutions. Questions related to keywords including fairness, inclusion,

accountability, transparency, data privacy and security, reliability and safety were addressed and leveraged to establish a healthcare intelligence framework for discussion.

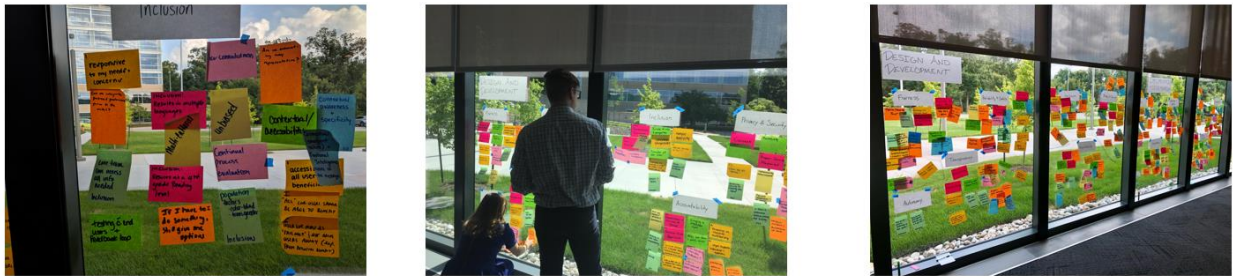
The tables were intended to have seven participants at each table, but some participants cancelled, bringing the total number to 54 participants. The facilitators reworked the tables to ensure as much diversity as possible and a balanced number at each table. Participants included employees from Amazon, BioNexus KC, Bradford & Galt, Center for Practical Bioethics, Cerner Corporation, Children’s Mercy, Donnelly College, First Hand Foundation, Helix Health, Kansas City University, KC Digital Drive, ModRN Health, Mozilla, MySidewalk, Pulmonaer Analytics, RKT Consulting, Rockhurst University, Thrive Health Connection, Truman Medical Center, UMKC School of Medicine, University of Kansas (Medical Center and the Center on Developmental Disabilities), and WellSky. Additional participant details (company/credential types, level of proficiency) are included in tables below. This data was collected prior to the conference at registration. Responses to the level of proficiency question guided us in the engineering of the tables, alongside their credentials and employer.

Company Type	Participants (n=54)
Healthcare Technology Companies	30
Educational Institutions (Academic Medical Centers, Private colleges/universities, Private & Public Medical Schools)	11
Non-Profit Institutions/Organizations	9
Healthcare Consulting Firms/Companies	4

Credential Type	Participants (n=54)
Health Care Provider	14
Master’s Degree (Non-Provider)	13
Bachelor’s Degree (Non-Provider)	13
Doctoral Degree (Non-Provider)	9
<i>Not Listed</i>	5

Level of Proficiency for “ethics in healthcare technology”	Participants (n=54)
Intermediate (Practical Application)	19
Novice (Limited Experience)	12
<i>Unanswered</i>	9
Advanced (Applied Theory)	7
Fundamental Awareness (Basic Knowledge)	5
Expert (Recognized Authority)	1
<i>Not Applicable/NA</i>	1

As the conversations flowed at each table, the seven facilitators from Cerner’s User Experience organization who volunteered for the day, not only helped to keep tables on track during activities, but also organized the prioritized thoughts into connected ideas. These connected ideas were then transcribed and entered into an excel database for analysis. Data Analysts from Cerner volunteered to create a code book for further analysis post workshop. To ensure privacy, identifiable information of the participants was not included in the database used for analysis. Their findings are articulated below.



Examination of patterns

Since this was our initial workshop, it also operated as a pilot to determine the feasibility of the approach, implementation and outcomes. To examine the approach, we looked at how our design played out in practice through the activities and tested some of our assumptions through evaluating patterns in the codes of the sticky notes as they travelled from notes to priorities. Due to the design of the workshop, we assumed that most groups would write comments on the table and the sticky notes would repeat the comments seen on the table. Based on this assumption, we wanted to understand which concepts were prioritized on the sticky notes and which concepts were left on the table. This could point to consistently prioritized concepts discussed throughout the workshop. To test this, we

parsed the concepts into codes by table, phase, and keyword. We found 700 codes in the table activity, which was the first stage of documentation for the participants as they discussed each phase. When examining the patterns, we actually found that some codes (155 total) appeared for the first time on the sticky note, without being mentioned on the table. A possible interpretation could be that concepts documented on the table were most likely posited by an individual. When prioritizing ideas as a group, those concepts could have evolved in complexity. This supports the need for collaboration in this space: codes appearing for the first time in the sticky notes were consistent in the most frequent codes suggesting iteration and collaboration are important components of this work.

Examination of code frequencies across phases

The codes defined in the codebook assisted the team in connecting similar ideas across sticky notes and documentation at each table and provided the team with the ability to standardize this mixed methodological approach of the workshop. From the analysis we found that the codes *consent* and *patient* were the top two codes overall. When looking at each phase, these 2 codes both appeared in the 5 most frequent codes in Design and Development (DD); consent appeared in the 5 most frequent codes in Implementation and Dissemination (ID), while patient occurred in the 5 most frequent codes in User Experience (UX). It is not surprising that in a workshop about ethics and AI, the patient and consent were the top focuses. They were a common thread in every phase and tended to be emphasized as evident by their appearance on sticky notes. In looking at the overall occurrences, consent tended to be prioritized at a higher rate than other codes. This suggests that the experts at the workshop place a lot of value on ensuring patient consent. This could point to an area that is deserving of more guidelines and awareness. The overall occurrence of the patient code on the sticky notes was relatively consistent with the averages, though slightly higher in DD than UX. There are many possible explanations for this. First, the industry emphasizes patient-centered design, so it is possible there is a natural inclination to prioritize the patient at that phase. Second, one could argue the top 3 codes in the UX phase – *exposing/sharing contributing factors*, *accessibility*, and *trust* – are grounded in patient-centered principles and, as a result, the patient code was not prioritized as often.

We also found certain codes to be present in each phase.

Design and Development (DD): *diversity in design* was coded frequently in this phase but not in other phases. It was frequently prioritized, indicating that diversity is one of the most important concepts in the design phase.

Implementation and Dissemination (ID): *outcomes* was the top code in this phase and among the top overall codes. This suggests that communication, implementation, and dissemination should prioritize clearly stating the outcomes this technology provides. We also found that this code was present after the collaborative portion of the activity, suggesting that this concept grew from group discussion and synthesis.

User Experience (UX): *expose or share contributing factors* was the top code in this phase. This points to an important ethical priority in AI. Across this group, it was frequently emphasized that patients should

know what factors are being considered in an algorithm. This implies that a patient should have autonomy to intervene on a factor or dispute the output of an algorithm.

Additionally, it is important to consider that *diversity in design* and *exposing contributing factors* were the top codes in one phase but not in the top 5 overall. This could suggest that these concepts are important in the context of that phase, but there could be an opportunity to emphasize diversity and human factors throughout the AI process. For example, stakeholders of implementation and dissemination should be asking if an algorithm was designed for diverse populations. In the design phase, stakeholders should consider how patients can interact with the factors included in an algorithm. Finally, these codes are patient-centered, while the top code in ID (*outcomes*) is systems-focused. This implies DD and UX prioritize the individual, while a priority of ID is the system.

Our team also examined the codes that were not coded on the table but appeared on the sticky notes (as priorities). Four codes (*standardized terms and conditions*, *disclosure*, *secure systems and interoperability*) appeared for the first time on sticky notes with more frequency than they did on the tables. One possibility is that these are all codes that we tend to associate with technology, so we may see these concepts mentioned at tables, or more clearly embedded in discussion if we had more representation from groups of highly technical backgrounds. This is something to explore in future workshops and conversations across the field.

Lastly, the codes *backup* and *liability* were never mentioned on a sticky note despite each being mentioned at a table ≥ 10 times. These concepts are typically relied on after AI fails. Backup is used to indicate what will keep a system going after a failure, while liability was used to indicate legal responsibility or consequences. There can be significant risks with not prioritizing these concepts. However, any intervention designed to address liability will likely have costs associated with it, placing it in tension with the potential costs of liability itself.

In summary, we found that the conversations and subsequent documentation throughout the day pointed to several areas of improvement, including a clear lack of diversity across workshop participants and the need for community engagement outside of those working in the tech industry. It is widely known that tech companies, especially those considered tech giants (Amazon, Apple, Facebook, etc.), have made little progress in hiring minorities, which could be argued is widening the gap in understanding the diversity of their users. This is problematic, especially as we think about this in the space of clinical practice. This evidence and the results from our workshop led us into increasing our community engagement and evaluating our partnerships across the industry.

Current Community Involvement

In addition to the 2019 workshop, we have had the opportunity to present locally to Kansas City Digital Drive, KC Tech Council's Tech on Tap and to the Women in Security Kansas City group. We were also featured in KCPT's Flatland magazine in December 2019. In October of 2020, we conducted our second workshop (virtually) at the annual American Congress of Rehabilitative Medicine conference. The

workshop was intended as an in-person workshop, similar to the one held in Kansas City, but due to COVID, the original workshop content was shortened to fit a virtual environment and assessments were formatted for online use. The modifications to the design did create new ideas for how the workshop could be scaled to virtual attendees, including our methods around data collection.

Through developing and implementing the workshop, we have created a network of subject matter experts who face daily challenges of ethical decision making and AI and have shared their expertise through presentations, participation in the workshop and as advisors to the Ethical.AI Project. Current collaborators include BioNexus KC, Children's Mercy, UMKC and Cerner. We have also worked with several start-ups, universities, and healthcare providers across the Kansas City area. These stakeholders have a distinct opportunity to create a collaborative space across healthcare, academia and intelligence to a) assess opportunities and risks in the development of and use of AI in healthcare; b) identify areas requiring investigation and, potentially, intervention; c) allocate appropriate resources for project development and d) engage policy and protocol by constructing policy briefs and best practices. We will collaborate with these partners across the industry in curating and creating content to initially serve as a foundational library of resources for healthcare IT companies and other stakeholders.

Proposed Next Steps

As we have outlined, there is a lack of adequate governance, oversight or accountability to respond to the rapid expansion of AI systems. Social systems that cross over public and private sectors, especially those related to health care have their own historical and regulatory frameworks that a national AI safety entity or general standards will struggle to meet the necessary sectoral requirements. As discussed, we could leverage some specific healthcare examples but have an opportunity to lean on the experiences of those in other fields. AI Now Institute argues, "we need a sector-specific approach that does not prioritize the technology but focuses on its application within a given domain (Whittaker et al., 2018)." Due to the positive feedback and support from the 2019 pilot Ethical.AI Workshop in Kansas City, Missouri, we have an opportunity to continue these workshops and expand our offerings into a larger project that spans various communities over the next year. The Ethical.AI Project will lead to the integration of ethical principles into the design, development, dissemination, and implementation of AI tools so that these ethical principles are demonstrated in the results.

Conclusion

In conclusion, by examining the literature and industry reports we have a responsibility to understand how ethics impact AI tools, especially in the healthcare space. AI tools will grow to play a vital role in healthcare, and we must be prepared for the benefits as well as the consequences. These tools will assist providers in evaluating, diagnosing, planning, intervening and monitoring care and we must enter this new era with caution alongside excitement. Some of the largest tech companies across the globe, such as Google, IBM, Amazon, and Microsoft are working to create processes and protocols to ensure ethical decision making across their evolving product line. These processes and protocols seem to be aligned to regulatory guidance but lack the necessary foundation required for best practices to be translated across expertise and levels of leadership. There is a need for community-based

infrastructures to enable public participation on AI accountability issues (Whittaker et al., 2018). As the Ethical AI project continues to grow and be grounded in community engagement, we anticipate increased connection and accountability across those who design, develop and implement AI across the healthcare industry. The workshop and its findings pave the way in understanding the complexities, risks and opportunities of connecting technology to clinical practice. The knowledge derived from this connection will perhaps be where we see the ultimate rewards in medicine (Mukherjee, 2017). Thus, developing standards and processes that consider the ethical implications of AI will be a stronger match to the sensitivity of clinical decisions, in which people, processes and data intersect.

Acknowledgements:

We would like to thank our facilitators from the Cerner Corporation User Experience organization who volunteered countless hours to the success of the planning, implementation and output of the workshop. We also would like to thank our panel speakers, who continue to assist the Ethical AI project team in developing and presenting content for our workshops and community presentations. The data analysis for this report was completed by volunteers, Kate Seiwert, MPH and Marina Daldalian, MPH. They currently donate their time to ensure that participants' voices are heard and disseminated. We are incredibly grateful for everyone who has supported this work from CPB, Cerner and our community collaborators.

References

- Dastin, J. (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. U.S. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08GIBM>
- Watson Oncology Clinical Trial Matching - Overview. (2020). IBM. <https://www.ibm.com/products/clinical-trial-matching-oncology>
- Lecher, C. (2018, March 21). A healthcare algorithm started cutting care, and no one knew why. The Verge. <https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy>
- Castelvecchi, D. (2020, November 18). Is facial recognition too biased to be let loose? Nature. <https://www.nature.com/articles/d41586-020-03186-4>
- Mukherjee, S. (2017). A.I. Versus M.D. - The New Yorker. Retrieved January 10, 2019, from <https://www.newyorker.com/magazine/2017/04/03/ai-versus-md>
- STAT. (2020, March 12). AI has arrived in medical imaging. Now the FDA needs to monitor it. <https://www.statnews.com/2020/02/28/ai-medical-imaging-fda-monitor-impact-patients/>
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., ... Schwartz, O. (2018). *AI Now Report 2018*. Retrieved from <http://www.ainowinstitute.org>
- *The Center's mission is to raise and respond to ethical issues in health and healthcare. Founded in 1984 as a free-standing, independent nonprofit, CPB's focus on healthcare in Greater Kansas City and its emphasis on "practical" (i.e., the real world) issues is what distinguishes it from nearly all other bioethics centers across the country. Most other centers are based in academic institutions and focus primarily on theoretical issues. CPB applies bioethics to the lived experiences of patients, providers and policymakers.
- **Cerner Corporation is an American supplier of health information technology services, devices and hardware. Cerner's mission is to relentlessly seek breakthrough innovation that will shape health care of tomorrow. The headquarters for Cerner are in Kansas City, MO with employees providing services across the globe.