

- **Cover Page**
- **Executive Summary**
 - Summary of the problem we looked at #note: this is just a skim of an introduction
 - Possible solutions and action plan for the client
 - Research we conducted
 - Google form
 - General Public
 - People in the Medical Field
 - Difficulties and Limitations
 - Potential issue with wording of black box, as people seemed to believe it was almost “hiding” the decision, when in fact it is just algorithms too complex for human readability.
- **Introduction**
 - Summary of the problem in more detail
 - Background information to understand the report and to understand the AI.
 - structure of our report
- **Ethical Issues or value tradeoffs addressed by our solutions**
 - **Ethical Issues and Value Tradeoffs our report addresses**
 - Data Protection and Privacy
 - Safety and Transparency
 - Complexity
 - Algorithmic fairness and Biases
 - Liability

- **Ethical Issues and Value Tradeoffs our report doesn't address**
 - Cyber Security
- **Analysis of our data we collected**
- **Proposed conclusions and recommendations**
 - No black box, people are more concerned with transparency than privacy based on our data (transparency)
 - AI use training should be funneled through several hospital representatives to the rest of the facility (sociotechnical system)
 - Very very transparent on how all data is kept, used, and saved, and on accuracy of the system. Information should not only be transparent but readily available. (transparency)
 - Come up with a slow introduction to warm people up to the use of AI, because even when told information straight up it often does not sink in, or they do not believe it (sociotechnical system)
 - Emphasize that the system is to support, not replace medical staff (transparency and sociotechnical)
 - Campaign stressing the amount and diversity of testing and training data (biases)
- **Resources**
 - Annotated Bibliography

Artificial Intelligence
Used as Diagnostic Tools
in the Medical Field

Authors: Dagmawe Hailelassie,
Duncan Forbes, Timothy Roback

5/20/21

Executive Summary

This paper will address some critical ethical concerns pertaining to the use of artificial intelligence as a diagnostic tool in the medical field. The main artifact of interest is an ai called Audia, created by Corti.ai. This software presents an interesting ethical question pertaining to the use of blackbox and voice recognition softwares. The main problems arising from the blackbox type of software is the lack of transparency, a tradeoff for increased security. Voice recognition on the other hand creates biases that arise from accents, leading to unfair, discriminatory behaviour or mistaken decision making and problems that arise from the utilization of this AI by medical professionals. In order to find solutions to these problems, further assessments and research were conducted on similar ai, and a comparison survey was created in order to gain insight on the public opinion between these forms of ai. Furthermore, information on the laws constricting such a software was needed in order to propose a solution that fits within the current bounds. Since laws pertaining to softwares that deal with private data vary wildly from place to place, the research was restricted to the United States. Finally, an interest in how medical professionals view these forms of diagnostic tools was desired. Research on how the users and the ai interacted was conducted as well as a questionnaire made specifically for medical professionals and those who wish to work in the field. Once these data were collected, an assessment was created in an informal fashion. After the information was obtained, a few options were selected for solving the identified problems. The first pertains to the lack of transparency given from blackbox systems. It seems that this trade-off for more privacy is generally disliked, and so removing the blackbox system is recommended. Maintaining as much transparency as

possible while also keeping some form of security on data is necessary. It is also recommended that each company maintains high transparency for their consumers. Clarity on what data is kept, where it is kept, who will have access to this information, and most importantly, the use and accuracy of diagnostic software used. Additionally, give patients the option to utilize the ai diagnostics system. Allowing control over patients' own well being will create a more comfortable atmosphere. Now for the concerns about the interaction between the medical professionals and the diagnostics ai. In order to maintain a sense of comfort for the user, it has been indicated that there needs to be human to human interaction. With this information, it is necessary that a medical professional deals with the ai, not the patient. When considering this interaction between the ai and the medical professional, it is important that the training for utilization of these diagnostic softwares be facilitated through representatives of the hospital, so that the medical professionals also maintain a form of human to human interaction during the learning process. An additional recommendation pertains to the patients themselves. Even after the medical staff becomes accustomed to using these devices to make diagnoses, there should also be a patient introduction to these diagnostic softwares. While full understanding of how they work is not entirely needed, a demonstration of their capability is highly recommended. Finally, to address the biases that arise from the use of voice recognition softwares, it is important to first obtain a large enough pool of training data from all forms of accents, ethnic groups, geographical locations, speech impediments, and the like in order to have a sufficient success rate for any individual that may call. Once this is established, test and stress to patients that the pool has been edited to encompass all people with different backgrounds both in a cultural and medical sense. With these recommendations, a step forward can be made to solve just a few of the ethical problems present.

Introduction

Artificial intelligence has become synonymous with support and efficiency in the medical community. AI in medical diagnosis helps with medical decision making, management, automation, admin and workflows. Over the past few years artificial intelligence in medical diagnosis has shown promise in changing the standards of medical care alongside reducing the extreme pressures felt by the medical industry. This paper will address some critical ethical concerns pertaining to the use of artificial intelligence as a diagnostic tool in the medical field. The main artifact of interest is an AI called Audia, created by Corti.ai. This AI assesses patients on a regular basis via text or call. Through Audia's automated patient engagement, users receive tailored support and Medics can track patient progress; receive alerts when patients need more support and gather pharmacovigilance and real world evidence data. Audia can have a natural conversation with patients, adapting to variations in conversation, such as answering patients' questions or asking more in-depth questions. This software presents an interesting ethical question pertaining to the use of blackbox softwares and voice recognition. The main problems arising from this type of software is the lack of transparency, a tradeoff for increased security, voice recognition biases that arise from accents, leading to unfair, discriminatory behaviour or mistaken decision making and problems that arise from the utilization of this AI by medical professionals.

In order to assess these dilemmas, further research was conducted on similar AI. One of the secondary artifacts of interest is an AI called Canary Speech. Canary Speech is a speech and language company that specializes in the area of identifying disease and human condition

through speech. Since there are many muscular diseases that affect speech patterns, the founders thought it possible to be able to detect certain diseases. Although they are both voice recognition medical softwares that are very similar, Corti AI and canary speech are different in the design of their AI systems. Canary Speech is not a blackbox software. The other secondary artifact of interest is an AI called Infermedica. Infermedica is a software that specializes in diagnosis. The methodology is different and is helpful in comparison against the main socio-technical environment, especially when looking at the utilization of this AI by medical professionals and the problems that arise.

Research Conducted and Data Collected

Our research topics fall into three categories. The first of these is information on the sociotechnical system pertaining to ai in the medical field. This category is incredibly broad, so further narrowing was required of the scope. This led to specifically diagnostics ai, and the interaction between these different forms of ai and the medical professionals utilising the equipment. The second pertains to research on the current ethical issues about software use in the medical field. This included concerns on privacy, patient transparency, and where the liability lies when a misdiagnoses occurs that involves a diagnostic software. Finally, research was conducted on the legal terms of the medical field as well as laws on data privacy and diagnostics in the US. After conducting this research, enough knowledge was gained in order to create a helpful questionnaire for the public and medical professionals to fill out. This questionnaire was multifaceted; giving insight on the public opinion of the current use of diagnostic software, as well as any possible misconceptions or fears that the public may have. Note that the majority of the general public has not had experience with the use of diagnostic ai in the medical field. The same is true of medical professionals, both for identifying any misconceptions or fears that

medical professionals may have and their interactions with diagnostic ai. These data could help the team later identify what problems need to be addressed for public reasoning and which need to be addressed for safety reasoning. After having identified these different problems and opinions on the matter, we can then compare them to the facts of how a diagnostic software interacts with medical professionals and patients.

Problems Addressed

By utilizing three different diagnostic software, the team created a way to compare different options of performing the same task. In doing so there is a way to isolate each problem and consider this method's pros and cons. Furthermore, isolating Audia as the main artifact allows for a critique of this particular ai instead of forcing the creation of an entirely new ai diagnostic software. Since the main problems in Audia pertain to the voice recognition algorithm, and the choice of a black box software, it was necessary to pick a software without a blackbox and a software that does not utilize voice recognition. By doing so, it becomes easier to identify what effect each facet of Audia has on its socio technical system. First we will begin by assessing the choice of a blackbox system. This kind of system utilizes a complete blackout of information. That is to say, any information received cannot be accessed throughout the whole computing process. This is where the concern of user-transparency arises.

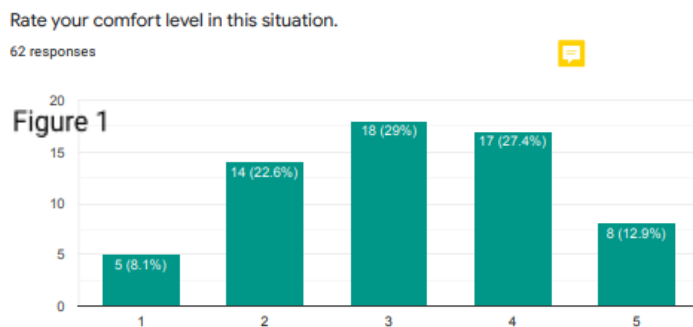
In normal circumstances, if a patient were to ask about why a diagnosis was given, the medical professional would be able to give some form of answer. Whether or not the patient would be able to understand the explanation is another matter, what is important is that there is an explanation that can be offered. In explicit use of blackbox systems, there is no way to know why a diagnosis was given over another. With no form of verification process on the diagnosis, the result becomes much more hazardous. Both CanarySpeech and Infermedica allow

user-transparency in its diagnosis, and while this is not required for every diagnosis, this option is incredibly important for patient transparency. Since Audia also utilizes a voice recognition system, the team included another ai that utilizes voice recognition but does not have the added complication of a blackbox system. It was also important to include a system that had no automated input whatsoever, as to assess the amount of control needed in order for a patient and medical professional to feel comfortable. The main problem that arises from the use of voice recognition software is its inability to recognize different accents. This bias often lies in a lack of data from different regions of the world, thus creating a xenophobic system. Finally there is the overarching problem between all of these systems, that of where liability lies during misdiagnoses. During a normal diagnosis procedure, the liability lies within the doctor for any mistakes made. However, when a diagnostics ai is utilized, there must be a distinction made between liability. Would the public be satisfied with faulting the ai, or should the doctor be blamed for deciding to follow the ai diagnosis? This final problem is the most pertinent when considering the relationship between the patient, medical professional, and ai.

Analysis of the data we collected

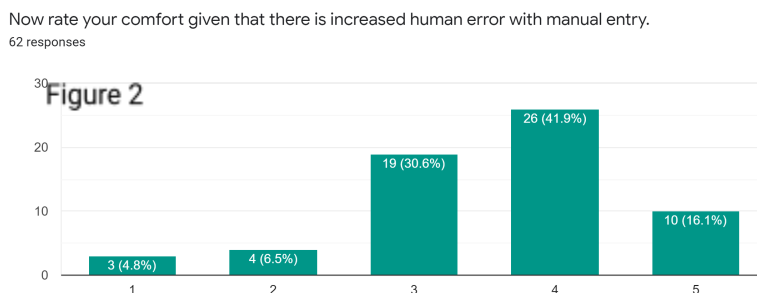
The team gathered our data through a survey sent out to the general St. Olaf student body, as well as specifically students and faculty in related fields. We asked the general student body two sets of questions each based around a proposed telehealth call scenario pretending as if they are the patient caller. The survey we directed at faculty and staff in the medical field we asked these same questions with the same scenarios, but then we added a third scenario in which the respondent was asked to imagine that they were on the caregiver end of the telehealth call. In this section of the report we discuss the results that we found most relevant and interesting to the previously outlined ethical issues. In the first scenario the team asked respondents to suppose

they are feeling sick, and they call the hospital for a tele-health option. While on the call respondents are told that an AI is listening in to help by using voice recognition software. The AI will create a diagnosis report based on the library of data at it's predisposal, and the responses to the questions posed by the representative. When the AI has reached a conclusion it will inform the representative of the hospital. The representative then tells them that the AI has shown a diagnosis. The representative then tells patients standard procedure for this diagnosed issue. In this case, the diagnosis is that of a common health issue that is self-treatable and a visit to the hospital is not recommended. Instead the representative provides respondents with the name of an over the counter medication that is commonly used to treat this condition.



The responses to general comfort level with the first proposed situation appear to resemble a normal distribution, which would suggest that implementation of AI along these lines would already be acceptable (figure 1). In the next question

we modified the situation to include manual entry of data by a hospital representative instead of automatic entry via audio recorded by the AI. With this new situation in mind, even when reminded that human error is on average higher than machine error, the responses now appear to



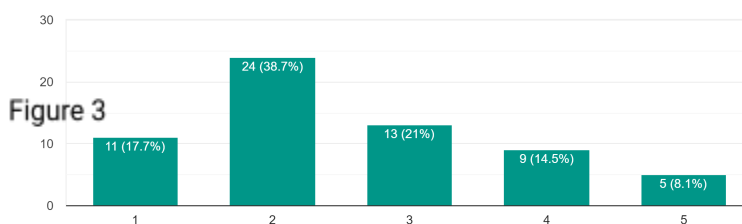
resemble a left-skewed distribution (figure 2). This information suggests that human input alongside AI is of high priority when considering the comfort of a patient. When asked their further

opinions for this major difference, respondents believed that either “The AI can’t pick up extremely nuanced health situations as many health problems can be, but if an AI was used in tandem with a human professional opinion then it becomes a lot more beneficial.” One other respondent mentioned that “My ideal telehealth option would not include AI. I am a rare disease patient and I would fear any AI would be programmed to look for the most likely diagnosis and run me in circles when that is not actually my issue at all. I already had to struggle with actual doctors to find my rare diagnosis and I think AI would actually complicate that process more.”

The team then introduced the idea of a black box system. The upside of having a black box system is that any data used in the processing of a result cannot be accessed, creating a high level of patient privacy. The downside is the lack of transparency in how the system arrives at its decision. Neither the patients nor the medical professional will know how the AI reached its conclusion. The team also noted that this process maintains and possibly improves the accuracy of the resulting diagnosis.

When asking their comfort level in this situation from one to five the team has found that

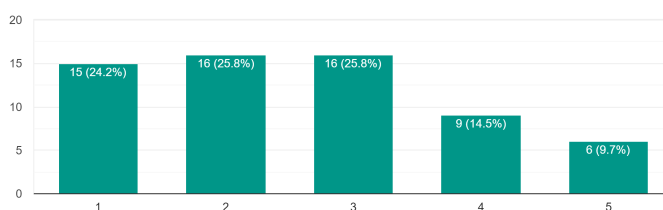
Rate your comfort level based on this new situation.
62 responses



respondents were overall more uncomfortable with this situation, with 56.4% reporting a comfort level of 2 or less (figure 3). When contrasted with the normally distributed general comfort levels

from the previous version of this situation in which black box technology had not yet been

Rate your comfort level if this form of AI was used on 911 phone calls. Keep in mind this is a medical emergency.
62 responses

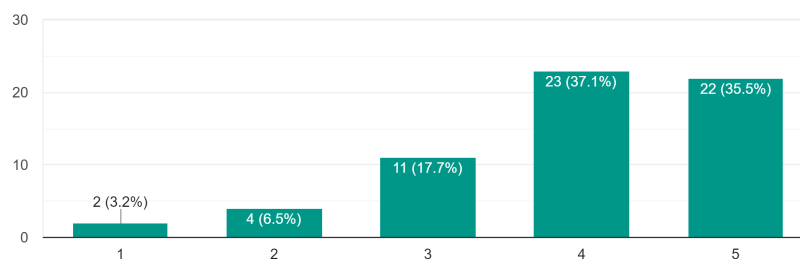


introduced, the right skew of this distribution suggests that the

transparency tradeoff is actually less favorable among patients. Finally we asked respondents to rate their comfort levels of using the proposed technology on 911 phone calls. During a medical emergency the majority of respondents were again uncomfortable with the situation, with exactly 50% reporting 2 or less, and more specifically 24.2% reporting no comfort at all (figure 4). This information leads us to believe that most of the respondents trust a human more than an AI when it comes to medical diagnosis especially during a 911 call. The team was also able to see that many of the respondents believed that a blackbox system is very dangerous and unpredictable. “If the reasons the AI reaches a conclusion aren't available for inspection, or at the very least able to explain why it got to conclusion X, then I would have very little faith in it.”

In the second scenario we asked respondents to “Suppose you are feeling sick, you go to the hospital and while you are checking in, the nurses let you know that either they or your doctor will-- given your permission-- enter your signs, symptoms, and critical health complaints into an AI-based device that is statistically more accurate than human doctors. The AI then analyzes your health data and compares your symptoms against the issues that they are commonly associated with and past diagnoses with similar symptoms. The doctor then uses this AI diagnosis in his decisions about your recommended treatment.”

Rate your comfort level on this situation.
62 responses



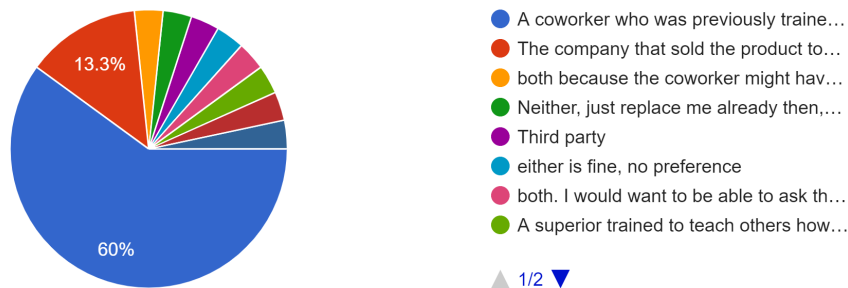
The distribution of responses to this new proposed situation is heavily left skewed. This further supports the hypothesis we made during the previous situation in which we believed that one of

Figure 4

the most highly valued variables in comfort level with AI is the amount of human interaction and oversight. From our data it would seem as though people are overall comfortable with AI as long as the process is being overseen and carefully checked by a medical professional.

Finally, a result that stood out to us from the questions directed specifically at those involved in or studying medicine was from the question “who would you rather be trained by to use this AI?” We gave respondents the options to be trained by either a representative from the company, a coworker who had previously been trained by the company, or to write their own response. We had initially expected that being trained by the company that created the AI would

Who would you rather be trained by to use this AI?
30 responses



make the most sense, but it appears that we underestimated the value that respondents place in trusting who they obtained their information from, as the overwhelming majority (60% from figure 6) responded that they would rather be trained by a coworker who had previously been trained.

Proposed conclusions and recommendations

Audia’s main problems have been sorted into three categories, and each should be solved independently. The first of these concerns to be addressed will be the lack of transparency

created by blackbox systems. According to the research we conducted, this problem is a significant one. The tradeoff for higher security on private information for no transparency is not worth it to the average citizen. This form of system would have no serious future without a major change of perspective from the population. Thus, it is recommended that blackbox systems not be utilized in the United States for use of medical diagnoses. Another recommendation is to keep a high level of transparency not only where legally required, but also in terms of the diagnostics software. Transparency here is required in two parts; between the company that developed the software, which is a legal requirement, and transparency between the hospital and the patient. This is a current grey area in a legal aspect. It is recommended that all data about the process of interaction between the patient, medical professional, and ai is not only accessible, but offered up front. That is to say, there should be a legal requirement to inform patients about the use of such diagnostic ai as well as some information on how the ai works and its accuracy level. A deep understanding of how the ai should not be offered. Instead, offer a simplistic version. If an example is desired, our questionnaire offers one in scenario 2.

The second concern to be addressed is that of the sociotechnical system that the ai resides in. The first recommendation is slightly tied to a previous recommendation of transparency between the medical professional, patient, and software. However, this recommendation instead focuses on the utilization of the AI by medical professionals. This interaction is the part of the AI that is not explained to the patient that needs to be understood by the medical professionals utilizing the software. It is recommended that firstly, there is a form of training for all staff who intend to interact with the software. Additionally, this training should be facilitated through representatives of the hospital. This will maintain an understanding of the AI while also adhering to any expectations the hospital may have. Another recommendation is that of patient trust in the

AI. It has become apparent that the general populace needs more than pure statistics on the performance of a software in order to develop trust. It is recommended that there be a slow introduction to the use of diagnostic AI as well as possible demonstrations. Furthermore, a choice should be allowed by the patient to include or exclude this form of AI in their diagnosis. Those who would be willing to introduce this aspect into their medical process and those who would eventually be willing will be a part of this, and those who are adamantly against would be seeing no decrease in excellence.

The final concern to be addressed is that of the xenophobic tendencies of voice recognition softwares. These types of softwares are entirely reliant on the form of training data given to the AI, and with a biased data set, there will be a biased performance. Currently, this is a very common practice. A training data set could have a significantly disproportionate amount of data for each kind of accent, ethnic group, those with speech impediments, and nations worldwide. The only real way to fix this is to introduce a large enough data set from each category. However, due to the nature of accents and their ability to mix together, it is nearly impossible to perfect this concept. Thus, there will be biases concerning this form of data input. With this known, it is only a matter of professionalism to add to the software a way to identify if the voice recognition system is able to understand what this individual is saying. Something of the form of being able to recognize when the success rate of converting those spoken words to machine language is below a specified threshold. If that condition is met, an error should appear and indicate that this software should not be utilized at this time, and that further training data is needed in order to serve the caller. While this is not a good solution in terms of being able to cater to any customer, it is a fair solution in terms of recognizing limitations of such a software. An alternative to this would be to have an individual manually check this transcription, adding

another sociotechnical check to the process in order to verify that the input data from a user is meeting the necessary requirements to make a proper diagnosis. It is necessary to point out that a human observer should be well versed in both the algorithm used in the AI and what information is more important than others in a medical sense. This kind of person is an invaluable resource, and is also incredibly difficult to find. A team of observers could be a solution to this problem, and is the final recommendation on the topic.

Resources

The conclusions reached were affected heavily by the laws in the medical field as well as the laws pertaining to data privacy and software use in the medical field. Additionally, current ethical concerns were identified in the field and addressed in order to create recommendations to problems that would have a significant impact. The last piece needed is information on the physical interactions between the medical professionals, the patients, and the diagnostics ai. A small amount of research was conducted on professional opinions on the topic paired with a questionnaire to gather information from the general population and others who work or will work in the medical field. Below is a list of sources used.

Annotated Bibliography

Socio-Technical System

Arslan, J., Benke, K. K., & Baird, P. N. (2020). Artificial Intelligence in Medicine: What Are the Implications for the Medical Practitioner? *Australasian College of Nutritional & Environmental Medicine Journal*, 39(2), 21–26.

This journal article brings up the potential conflict of interest that could occur if we ever reach a point in which AI begins replacing the functions of doctors and rendering them obsolete. It also looks at the other side of this issue, and realizes that by taking some of the more time consuming menial tasks of doctors, AI could allow doctors to focus more on other aspects of their work. These include things such as hospital management and emotionally supporting patients. In the article these two aspects are weighed against each other as to which is the more likely outcome.

David, M. (2019). AI and the Illusion of Human-Algorithm Complementarity. *Social Research*, 86(4), 887–908.

This article takes a deeper look into the dynamics between humans and AI. The article goes on to argue that the type of AI in combination with medical staff systems that we are investigating do not have a positive overall dynamic. It suggests that AI may actually undermine human skills and expertise rather than work nicely in tandem as the systems we are investigating claim that they can do. It also touches on this topic specifically in the context of medical diagnosis, thus is even more crucial of a resource to us.

Hui-Chen Lin, Yun-Fang Tu, Gwo-Jen Hwang, & Hsin Huang. (2021). From Precision Education to Precision Medicine: Factors Affecting Medical Staff's Intention to Learn to Use AI Applications in Hospitals. *Journal of Educational Technology & Society*, 24(1), 123–137.

This article is valuable to us because it provides insights on the attitudes of medical professionals in regards to AI, and specifically in their intentions to learn about AI. Using information collected from 285 surveys given out to medical professionals that aimed to rate each professional on five major factors the researchers connected with AI learning intentions. Using statistical analysis it was then uncovered which of the factors best predicted the others, and this provides valuable information that could potentially be used for recommendations on how to best accommodate the intentions of medical professionals and most effectively educate on the uses of AI.

Ishii, E., Ebner, D. K., Kimura, S., Agha-Mir-Salim, L., Uchimido, R., & Celi, L. A. (2020). The advent of medical artificial intelligence: lessons from the Japanese approach. *Journal of Intensive Care*, 8(1), 1–6.
<https://doi.org/10.1186/s40560-020-00452-5>

The specific information in this article that stood out as potentially useful to us was the discussion of the growing need for an AI-workforce to accommodate the increase in AI use, and how this demand outpaced the growth of AI use training programs and general knowledge of AI. We are interested in potential fixes to this that may have been implemented in Japan, and how successful these were. Hopefully from this we can gain ideas for recommendations on training protocols for the AI that we are more closely studying.

Moulton, Carol-anne E., et al. "Slowing Down When You Should: A New Model of Expert Judgment." *Academic Medicine*, vol. 82, no. Suppl, Oct. 2007,

doi:10.1097/acm.0b013e3181405a76.

This article is useful to us because it examines and discusses the literature around expertise, and when experts need to switch from automated routines to more carefully analyzed problem solving. They discuss how when one becomes an expert they often operate mostly in an autopilot-like mode if things seem normal because they know it so well. However it is the process of when and how they recognize when things are outside the norm, and thus demand a more critical eye in order to properly assess the issue. This is particularly crucial in our area of AI diagnostics, because they will make it incredibly easy to slip into the autopilot mode, and simply give inputs and take outputs. The skill of recognizing when the output seems off and properly questioning it then becomes of utmost importance.

Paranjape, K., Schinkel, M., Nannan Panday, R., Car, J., & Nanayakkara, P. (2019). Introducing Artificial Intelligence Training in Medical Education. *JMIR medical education*, 5(2), e16048. <https://doi.org/10.2196/16048>

This article touches on almost all of the aspects of medical AI that we are examining. First and foremost it is an article about medical education in regards to new technology and AI. It both assesses the current state of education and also makes recommendations about the improvements for the future of medical information. This is valuable in helping us create our own recommendations to incorporate the AI we are looking at into the current socio-technological system surrounding medical diagnostic AI. The article also touches on ethical issues such as transparency, liability, and the use of black boxes.

Pinto dos Santos, D., Giese, D., Brodehl, S., Chon, S. H., Staab, W., Kleinert, R., Maintz, D., & Baeßler, B. (2019). Medical students' attitude towards artificial intelligence: a multicentre survey. *European Radiology*, 29(4), 1640–1646. <https://doi.org/10.1007/s00330-018-5601-1>

This article provides insight on the thoughts of medical students regarding AI. The article discusses the results of a survey that went out to students at three major medical schools, and received 263 responses. These questions were specifically aimed at AI use in radiology but we believe results can still be applied to our situation. Some interesting metrics mentioned were that 68% of responders stated that they were not aware of the technology involved in radiology AI, and only about 71% believed that AI training needed to be included in their studies. The responders did also widely agree that they would not be replaced by AI which refutes many other sources that cite this as a possible concern.

Ethical issues

Kluge, Eike-Henner, et al. "Ethics Certification of Health Information Professionals." Yearbook of Medical Informatics, Georg Thieme Verlag KG, 22 Apr. 2018, www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0038-1641196.

This article overlaps on ideas from our socio technical system articles and provides a model for ensuring the ethical acceptability of the AI in healthcare and more specifically diagnostic AI in healthcare. It builds on work done by the Global Protection of Health Data Project within the Security in Health Information Systems. This article concludes that there has been little to no attention to specify the qualifications required for health informatics professionals. We thought this would be relevant when looking at the topic of transparency between the patient as well as the medical professionals who are by law required to describe the procedure if something were to go wrong.

Morley, Jessica, et al. "The Ethics of AI in HealthCare: A Mapping Review." Social Science & Medicine, Pergamon, 15 July 2020, www.sciencedirect.com/science/article/pii/S0277953620303919#bib101.

This article presents a mapping review of the literature concerning the ethics of artificial intelligence in healthcare. It effectively summarizes current debates of ethical concerns faced in the medical AI industry. It also provides an overview of the epistemic, normative and overarching ethical concerns related to the algorithmic use in healthcare. Which mainly includes inconclusive evidence, Misguided evidence, unfair outcomes and traceability. The primary reason we wanted to use this article was the vast literature it reviewed to identify healthcare examples for each of the concerns previously stated. The hope of this article and ultimately our research is to encourage a discussion of the ethical implications of AI-health at individual, interpersonal, group institutional and societal level.

Vayena, Effy, et al. "Machine Learning in Medicine: Addressing Ethical Challenges." PLOS Medicine, Public Library of Science, 6 Nov. 2018, journals.plos.org/plosmedicine/article?id=10.1371%2Fjournal.pmed.1002689.

The article starts off by mentioning that a recent UK survey reports that 63% of the adult population is uncomfortable with allowing personal data to be used to improve healthcare and is unfavorable to AI replacing doctors/nurses. One of the major issues that this article addresses is biases, fairness and transparency. Poorly representative training data sets can introduce biases to MLm trained algorithms. Two examples that were provided to support this claim was a population data bias in the overdiagnosis of schizophrenia in African Americans as well as an algorithm that is trained mostly on data from older white men.

This article also looks at the lack of transparency in a black box algorithm. We thought this would be a very good resource to address ethical issues in the themes of biases as well as data and information transparency.

Vollmer, Sebastian, et al. "Machine Learning and AI Research for Patient Benefit: 20 Critical Questions on Transparency, Replicability, Ethics and Effectiveness." ArXiv.org, 21 Dec. 2018, arxiv.org/abs/1812.10404.

Written by multiple authors this article addresses the transparency issue that arises from healthcare AI. The article further acknowledges that although there is no consensus on what the best practice of these issues looks like there should be critical questions for health related AI technology. The article tries to answer questions like when and how should patients be involved in data collection, analysis, deployment and use of medical AI? Is there organizational transparency about the flow of data? On what basis is data accessible to other researchers? And many more. This article helped us navigate the field of diagnostic AI and the transparency as well as ethical issues that come along with it. Our main goal is to use this article and compare the points raised alongside the AI examples we provide later on in our research project and this bibliography.

W. Nicholson Price II, JD. "Potential Liability for Physicians Using Artificial Intelligence." JAMA, JAMA Network, 12 Nov. 2019, jamanetwork.com/journals/jama/fullarticle/2752750.

This article looks at diagnostic AI and it's liability for physicians as well as the responsibilities and duties of physicians from an ethics point of view . Points including taking active steps to evaluate pract specific algorithms in diagnostic AI are brought up in this paper which we thought would be beneficial for our research

Legislation and other Artificial Intelligence

Canary Speech: Home. 2021. *Canary Speech*. [online] Available at: [<https://www.canaryspeech.com/>](https://www.canaryspeech.com/) [Accessed 18 March 2021].

Canary Speech is a new artificial intelligence based on voice recognition. Since there are many muscular diseases that affect speech patterns, the founders thought it possible to be able to detect certain diseases and conditions through recognition from speech. This is the main website for advertisement and gives some insight on their marketed goals, their team, some surface level project work, as well as the individual diseases they are able to detect.

Infermedica.com. 2021. *Infermedica Home Page*. [online] Available at: <<https://infermedica.com/case-studies/pzu-zdrowie>> [Accessed 19 March 2021].

Infermedica is another artificial intelligence software that specializes in diagnosis. The methodology is different and is helpful in comparison against the main socio technical environment. This is the business' home page, and general information about the company can be found here.

Käde, L., & von Maltzan, S. (2019). Towards a Demystification of the Black Box--Explainable Ai and Legal Ramifications. *Journal of Internet Law*, 23(3), 3–13.

An article considering the effects that Black Box artificial intelligence will have on softwares. This considers specifically biased parameters, which can lead to untold problems in the socio economic setting, as well as to various forms of lawsuits. While the article considers the effects specifically on discrimination between people by the Ai's inability to adjust for any individual's differences, this information can be extrapolated towards all sorts of biases, not only with people but with the environment that the Ai exists in as a whole.

Kanowitz, S., 2021. *AI-enabled voice assistance for responders -- GCN*. [online] GCN. Available at: <<https://gcn.com/articles/2019/09/25/ai-audrey-responders.aspx>> [Accessed 19 March 2021].

This is a report on an in-progress artificial intelligence software with the intention of being able to help first responders on-scene. While the software is not in use yet, the socio technical environment is explicitly highlighted in the article. This will be another comparison point for the main artifact.

Loc.gov. 2021. *Regulation of Artificial Intelligence*. [online] Available at: <<https://www.loc.gov/law/help/artificial-intelligence/americas.php#us>> [Accessed 18 March 2021].

The Library of Congress is a website that updates the current concerns of congress, organized by category. This is the most recent update to artificial intelligence action taken by the Americas and the Carribean, as well as comparisons from other countries. While the article never individually brings up artificial intelligence in medicine explicitly, it does bring up privacy and transparency laws pertaining to AI, as well as the regulations for federal institutions.

Ncsl.org. 2021. *Legislation Related to Artificial Intelligence*. [online] Available at:
<<https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx>> [Accessed 18 March 2021].

National Conference of State Legislature is a website that records the recent bills and resolutions that were introduced at a state level. This shows the priorities of each state in recent years, as well as what laws have been enacted, which have failed, and which have yet to be decided.

Reynolds, M., 2021. *Health insurer calls analysed for signs of disease in your voice*. [online] New Scientist. Available at:
<<https://www.newscientist.com/article/2120426-health-insurer-calls-analysed-for-signs-of-disease-in-your-voice/>> [Accessed 19 March 2021].

Canary Speech gathered data through a relatively controversial method. This is an article describing the methodology, and while the process was legal, there were certainly questions raised as to whether this data collection method should be used.

Tobia, K., Nielsen, A. and Stremitzer, A., 2020. When Does Physician Use of AI Increase Liability?. *Journal of Nuclear Medicine*, [online] 62(1), pp.17-21. Available at:
<<https://jnm.snmjournals.org/content/62/1/17.long>> [Accessed 18 March 2021].

This is a journal article on the impact that artificial intelligence use by physicians will have on the liability of doctors through the eyes of jurors. Since one of the sources of judgements in a court case comes from jurors, it is important to see the opinion of the general public of a mis-diagnosis from a doctor with the use of an artificial intelligence software compared to a mis-diagnosis without the use of such a software. This article describes a study that attempts to answer that exact question.

A link to the questionnaire can be found here:

https://docs.google.com/forms/d/e/1FAIpQLSeuriaJ1_TXioUzOq732no_KG9p_Fj_e2XMopNykTKLM8YNGQ/viewform?usp=sf_link