APPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR IMPROVING HEALTHCARE IN THE WASHINGTON METROPOLITAN AREA

ANALYZING TECHNOLOGICAL INFLUENCES ON HEALTHCARE WORKER BURNOUT

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Andrew Lucktong

October 27, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Rider Foley, Department of Engineering and Society

Rosanne Vrugtman, Computer Science

Introduction

There is a common saying in healthcare that your zip code is a better predictor of your health than your genetic code. The County Health Rankings model, created by the University of Wisconsin Population Health Institute, shows that around 80% of health outcomes are linked to socioeconomic factors, physical environment factors, and behavioral factors (Figure 1). Many non-clinical factors, known as the social determinants of health (SDoH), can be used to determine potential health risks and outcomes. The factors can be grouped into five distinct groups: economic stability, education access/quality, health care access/quality, neighborhood/built environment, and social/community context (Centers for Disease Control and Prevention [CDC], 2023).

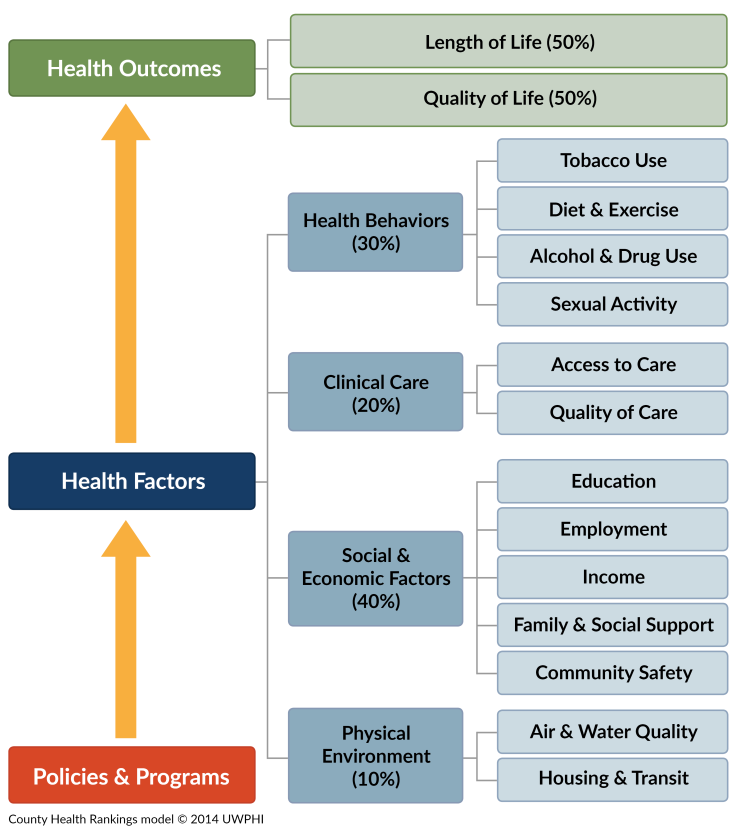

Figure 1: County Health Rankings Model (County Health Rankings Model, n.d.)

In 2018 the Virginia Commonwealth University Center on Society and Health conducted a study to compare life expectancies in the Washington Metropolitan Area based on census tracts of the region (See Figure 2). The area, comprised of Washington D.C. and parts of Maryland and Virginia, is only around 5,500 square miles, but the study found that life expectancies varied by as much as 27 years (Woolf et al., 2018, p. 4).

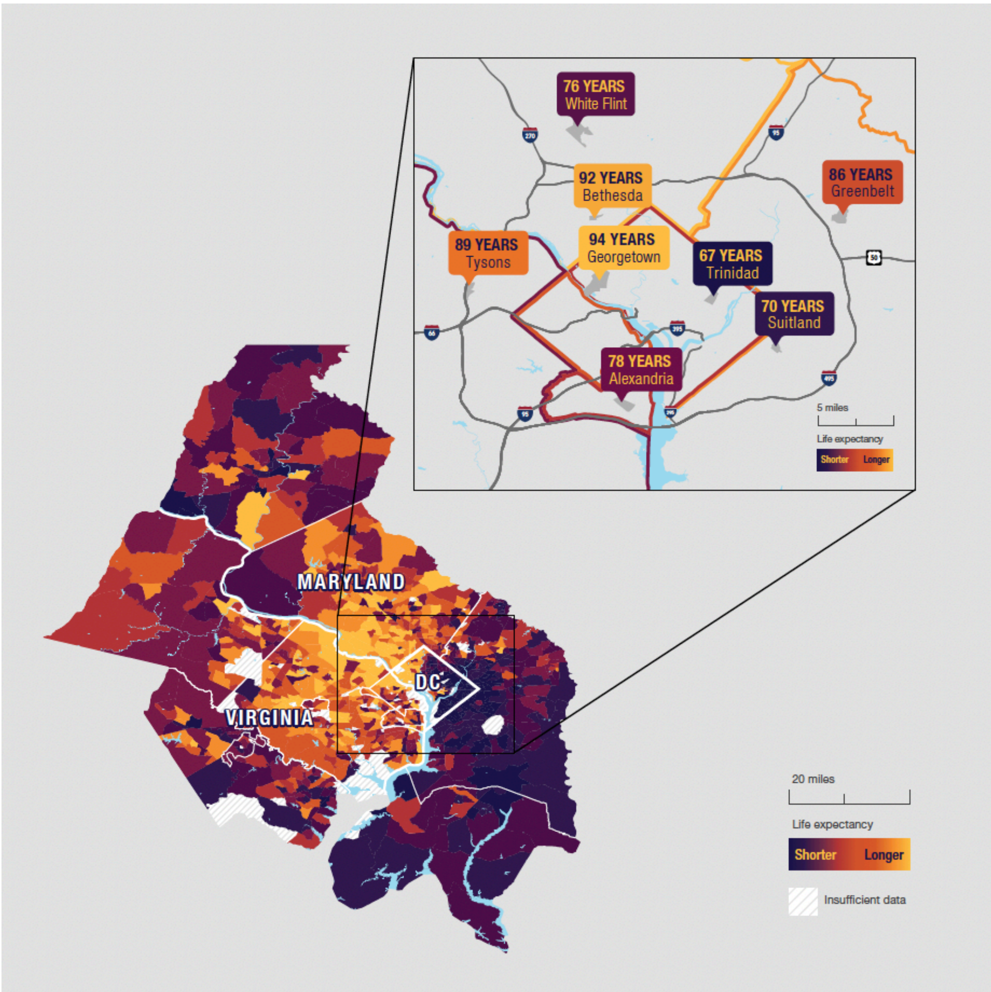

Figure 2: Life expectancies in the Washington Metropolitan Area (*Large Life Expectancy Gaps*, 2019)

A 2019 analysis by researchers at the Department of Population Health at New York University School of Medicine found a similar figure, with Washington D.C. having a life expectancy gap of 27.5 years, the second largest gap in the US behind Chicago at 30.1 years (*Large Life Expectancy Gaps*, 2019). These differences in health are largely attributed to the social determinants of health, with neighborhood disparities being more prevalent in larger cities. Exploring and analyzing social determinants of health and determining their effects on patient outcomes is important in providing equal and equitable care, but current data collection methods lack structure and uniformity.

A 2023 survey of medical professionals conducted by NORC at the University of Chicago revealed that nearly eight in ten respondents collected SDoH data at their organization but experienced troubles when it came to integration and standardization of this information (*Social Determinants of Health Data, 2023)*. Artificial intelligence (AI) could provide a solution for efficiently processing and analyzing patient data to predict at-risk groups and individuals based on the social determinants of health. Natural Language Processing (NLP) is a subset of AI that gives computers the ability to understand and interpret human language in a meaningful way (IBM, n.d). Applying NLP to medical records could allow a computer model to extract key information from a patient's chart, providing more structured and organized data. This data could then be used as input for a machine learning (ML) clinical decision support system (CDSS) to provide personalized care plans and insights to health care providers caring for patients in areas negatively impacted by SDoH. This technical project focuses on applying NLP to extract SDoH information from electronic health records (EHRs) and using machine learning to provide decision support for physicians with the goal of incorporating SDoH recommendations into personalized care plans for individuals in the Washington Metropolitan Area.

Artificial Intelligence in Medicine

Current software applications with the ability to extract specific clinical data from unstructured EHR text mostly focus on a specific area of the social determinants of health. A 2021 systematic review of NLP systems used for extracting SDoH data found that smoking status, substance use, homelessness, and alcohol use were the most frequently studied subcategories (Patra et al., 2021). Many studies have also proposed AI-based solutions to predicting health outcomes based on SDoH and some have had promising results. In a 2020 study, machine learning was used to predict if patients would have emergency department or inpatient utilization within 90 days based on health determinant data. This study found that the model was able to predict utilization correctly 83% of the time when given unseen testing data which is promising as machine learning is anticipated to improve (Chen et al., 2020).

Despite the advantages these technologies may provide, there still seems to be a lack of full implementation into healthcare systems. While 70% of SDoH data collection is done electronically, collection is not entirely automated, increasing workload for physicians and nurses. Designing an integrated system that is both intuitive and simple to use is important for the scope of this technical project. Natural language processing greatly increases simplicity of data collection by interpreting language in an electronic health record like a human would. Figure 3 shows a simplified version of named-entity recognition (NER), an NLP technique that can identify important words or phrases based on a pre-defined set of named entity categories (*A Comprehensive Guide to Named Entity Recognition (NER)*, n.d.). This technique could be used to identify and extract the many different social determinants of health, each with a specific label or category. Collecting this information using NLP also gives the advantage of having structured data, making statistical analysis and research more possible.

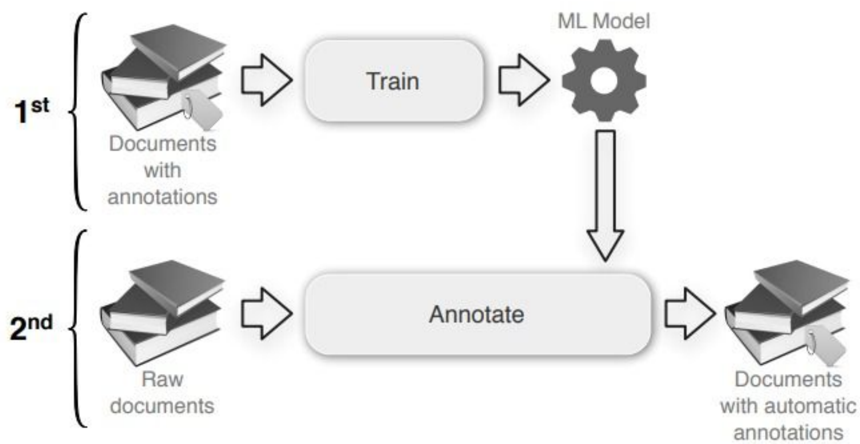

Figure 3: Simple example of a machine learning based NER model (Dathan, 2021)

Clinical Decision Support Systems (CDSS) utilize clinical data to assist physicians in making recommendations for patients. Decision support is deeply embedded into EHR systems by taking advantage of patient data. As of 2021, nearly 9 in 10 U.S. office-based physicians have adopted some form of an EHR (*Office-Based Physician Electronic Health Record Adoption*, n.d.). In a traditional knowledge based CDSS, the system uses a predefined rule set to recommend the appropriate action based on the clinical data. However, there has been a recent increase in interest of non-knowledge based clinical decision support systems which use AI to create suggestions instead of using a predefined ruleset (See Figure 3). These systems do not require rules for the inferencing and can work with incomplete information, making them advantageous for analyzing SDoH data where there is no real ruleset for care plans (Sutton et al., 2020). Machine learning algorithms also tend to improve performance with more data, allowing this type of system to make better estimates of SDoH related issues over time.

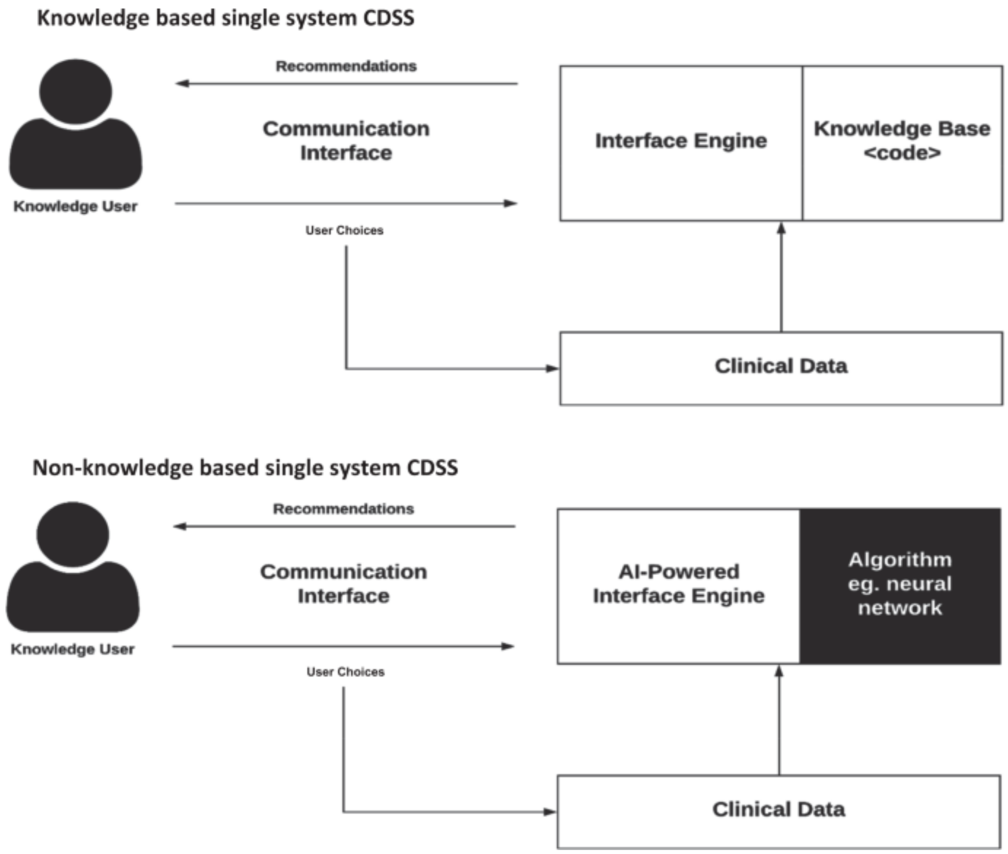

Figure 4: Knowledge based vs. non-knowledge based CDSS (Sutton et al., 2020)

This type of system could be adapted to provide specific action-based care plans based on the data extracted via an NLP model. JvionAI, a clinical AI company, has created a similar model called Jvion CORE that is able to provide specific recommendations for at-risk patients based on several different factors analyzed by their machine learning model. The system takes advantage of a combination of EHR and publicly available data and integrates seamlessly with existing healthcare technologies (*Implementing Advanced Technology in Healthcare*, n.d.). The goal of using machine learning is to have a computer think like a clinician, providing insights to healthcare workers to help them make the best decision for the patient. The model for this project would be trained on prior health data, analyzing social determinants of health, and determining what actions or treatments were most effective based on outcomes. This would allow the system to provide patient-centered care plans, reducing health disparities with proper semi-automated intervention.

Applications of Infrastructure to Combat Physician Burnout

Even before the COVID-19 pandemic in 2020, 35-45% of nurses/physicians in the United States were experiencing burnout according to a survey conducted by the National Academy of Medicine (National Academies of Sciences, Engineering, and Medicine, 2019). In his 2022 advisory, U.S. Surgeon General Vivek Murthy explains the negative impacts that physician burnout has patient care and worker livelihood. One of the main forces influencing physician burnout is the quantity of administrative work (Murthy, 2022). Murthy emphasizes that healthcare technology companies need to “design technology to serve the needs of health workers, care teams, and patients across the continuum of care” and “strengthen integration of data across different platforms and health sectors” to lessen the burden of technology based administrative tasks. Using Susan Leigh Star’s framework of the Ethnography of Infrastructure (1999), we can begin to analyze how assistive AI technologies can help to increase productivity and reduce healthcare worker burnout.

Infrastructure is a dynamic component that plays an important role in influencing how technology is used and how societies are affected. While conducting fieldwork with a group of biologists and a computer scientist, Star found that using standard participatory design principles to analyze human needs was much less effective than considering infrastructural aspects for creating a working and usable system. This led her to develop a relational definition of infrastructure with nine key properties for analyzing social and technical factors based on her findings. The main goal of AI technologies in healthcare is using computers to aid physicians by reducing monotonous or time-consuming tasks. When it comes to infrastructure, the most important aspects for analyzing AI technologies are embeddedness and embodiment of standards. Embeddedness explains the invisible interconnectedness of infrastructure and how it is sunk into other technologies and systems where subcomponents are often not considered but are rather just "in" it. Embodiment of standards defines how infrastructures work with other systems by using common interfaces and how these standards shape the scope of the system which Star applied by embodying many standards of the biological and academic community.

Designing an AI system to analyze the social determinants of health requires integrating with health system data and electronic health records. Current work by Microsoft is underway to integrate Azure OpenAI into Epic Health Systems electronic health record system to improve patient care using AI technologies (Boyd, 2023). Much like subcomponents of any other system, this system would not be visible or impeding to a physician's workflow but would work in the background to increase productivity. Systems like GitHub Copilot use AI to aid programmers by autocompleting code and are directly built into development environments. Using properties of infrastructure to analyze how these AI technologies are effectively embedded into previous technologies would help determine how an AI-based system could be applied to healthcare.

Much like biology, medicine has many standards that need to be considered when creating a usable system. Applying embodiment of standards ensures that AI technologies work seamlessly with medical technologies. Working to build a system that integrates medical standards would allow for intuitive use and simplifies connectivity of this system to current systems. The Office of the National Coordinator for Health Information Technology publishes and maintains the Interoperability Standards Advisory (ISA) which contains a comprehensive list of health information standards. These standards help keep implementation simple by providing specific guidelines on how information should be handled and processed (*2023 Interoperability Standards Advisory*, 2023). It is also important to consider patient and technological standards, such as data privacy implications. Data privacy was one of the main concerns patients had when asked about the use of AI in healthcare, and creating a system that ensures privacy will be beneficial for both physician and patient (Richardson et al., 2021).

Research Question and Method

Designing effective technology to serve the needs of people requires human-centered development (HCD). The International Organization for Standardization (ISO) describe HCD as "an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques" (*ISO 9241-210*, n.d.). To design the most human-centered technology, the question must be asked: What are health care workers' opinions on technology and in what ways has technology either increased or decreased their symptoms of burnout?

To answer this, focus group interviews of Virginia healthcare workers will be conducted. Interviewing the primary users of the proposed system will help provide a better understanding of how to implement a system that both improves physician productivity and patient care. Questions should give an overarching view of system requirements, including how technology has either increased or decreased symptoms of burnout, how workers use technology daily, what technologies healthcare professionals would like to see in the future and their opinions on upcoming technologies like artificial intelligence and machine learning. Gathering this information from different organizations in Virginia allows me to see different challenges that certain healthcare workers may have with computer-based technologies.

Additionally, I plan to spend time with healthcare teams to observe how technology is used during a regular workday. This will allow me to better understand the technologies used and how automation could be applied to increase productivity and decrease administrative burden. It is important to not only consider what the challenges are with the current technology but also understand how these challenges could be addressed using AI and ML.

Conclusion

As AI systems become increasingly integrated into everyday life, it is important to consider the applications AI may have in the healthcare industry. Analyzing the social determinants of health with AI and creating a machine learning algorithm to provide patient-centered care can help work toward the goal of reducing health disparities and inequities. This is especially important in areas where social factors have a large impact on healthcare outcomes like in the Washington Metropolitan Area but has the potential to impact more than just a single community. The research outlined within this prospectus aims to determine how technology can be used to decrease physician burnout. The insights from this research will provide design principles to develop the most human-centered technology, focusing on the technologies that do and do not work. I hope to learn what kinds of technologies healthcare workers are looking for and applying my own expertise to determine how any issues could be addressed using AI.

Resources

2023 Interoperability Standards Advisory. (2023). Office of the National Coordinator for Health Information Technology.

A Comprehensive Guide to Named Entity Recognition (NER). (n.d.). Retrieved October 27, 2023, from https://www.turing.com/kb/a-comprehensive-guide-to-named-entity-recognition

Boyd, E. (2023, August 22). *Microsoft and epic expand AI collaboration to accelerate generative AI's impact in healthcare, addressing the industry's most pressing needs*. The Official Microsoft Blog. https://blogs.microsoft.com/blog/2023/08/22/microsoft-and-epic-expand-ai-collaboration-to-accelerate-generative-ais-impact-in-healthcare-addressing-the-industrys-most-pressing-needs/

Centers for Disease Control and Prevention. (2023, September 15). CDC - Social Determinants of health - STLT gateway. https://www.cdc.gov/publichealthgateway/sdoh/index.html

Chen, S., Bergman, D., Miller, K., Kavanagh, A., Frownfelter, J., & Showalter, J. (2020). *Using Applied Machine Learning to Predict Healthcare Utilization Based on Socioeconomic Determinants of Care*. *26*.

County Health Rankings Model. (n.d.). University of Wisconsin Population Health Institute. County Health Rankings & Roadmaps 2023.

Dathan, A. (2021, November 3). A Beginner's Introduction to NER (Named Entity Recognition). *Analytics Vidhya*.

Ergonomics of human-system interaction. (n.d.). ISO 9241-210:2019. Retrieved October 26, 2023, from https://www.iso.org/standard/77520.html

IBM. (n.d.). *What is natural language processing?*. IBM. https://www.ibm.com/topics/natural-language-processing

Implementing Advanced Technology in Healthcare. (n.d.). Jvion. Retrieved October 27, 2023, from https://jvion.com/approach/technical-implementation/

Large Life Expectancy Gaps in U.S. Cities Linked to Racial & Ethnic Segregation by Neighborhood. (2019, June 5). NYU Langone News. https://nyulangone.org/news/large-life-expectancy-gaps-us-cities-linked-racial-ethnic-segregation-neighborhood

National Academies of Sciences, Engineering, and Medicine. 2019. *Taking action against clinician burnout: A systems approach to professional well-being*. Washington, DC: The National Academies Press.

Office-based Physician Electronic Health Record Adoption. (n.d.). HealthIT.Gov. Retrieved October 27, 2023, from https://www.healthit.gov/data/quickstats/office-based-physician-electronic-health-record-adoption

Patra, B. G., Sharma, M. M., Vekaria, V., Adekkanattu, P., Patterson, O. V., Glicksberg, B., Lepow, L. A., Ryu, E., Biernacka, J. M., Furmanchuk, A., George, T. J., Hogan, W., Wu, Y., Yang, X., Bian, J., Weissman, M., Wickramaratne, P., Mann, J. J., Olfson, M., … Pathak, J. (2021). Extracting social determinants of health from electronic health records using natural language processing: A systematic review. *Journal of the American Medical Informatics Association: JAMIA*, *28*(12), 2716.

Richardson, J. P., Smith, C., Curtis, S., Watson, S., Zhu, X., Barry, B., & Sharp, R. R. (2021). Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ digital medicine*, *4*(1), 140.

Social Determinants of Health Data: Survey Results on the Collection, Integration, and Use. (2023). American Health Information Management Association (AHIMA).

Star, S. L. (1999). The ethnography of infrastructure. American behavioral scientist, 43(3), 377-391.

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *Npj Digital Medicine*, *3*(1), Article 1.

Woolf, S., Chapman, D., Hill, L., Schoomaker, H., Wheeler, D., Snellings, L., & Lee, J. H. (2018). *UNEVEN OPPORTUNITIES*. Virginia Commonwealth University (VCU) Center on Society and Health.