DIGITAL TRANSFORMATION OF HEALTHCARE USING ARTIFICIAL INTELLIGENCE

by

Jennifer Gogova

A Thesis Submitted to the Faculty of

The College of Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, FL

May 2023

DIGITAL TRANSFORMATION OF HEALTHCARE USING ARTIFICIAL INTELLIGENCE

by

Jennifer Gogova

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Oge Marques, Department of Electrical Engineering and Computer Science, and has been approved by all members of the supervisory committee. It was submitted to the faculty of the College of Engineering and Computer Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Oge Marques (Apr 12, 2023 12:10 EDT)

Oge Marques, Ph.D.
Thesis Advisor

Borko Furht (Apr 19, 2023 12:56 EDT)

Borivoje Furht, Ph.D.

Nataliia Neshenko

Nataliia Neshenko (Apr 19, 2023 12:58 EDT)

Nataliia Neshenko, Ph.D.

Sareh Taebi

Sareh Taebi, Ph.D.

Hanqi Zhuang, Ph.D.
Chair, Department of Electrical Engineering and Computer Science

Stella Batalama

Stella Batalama (Apr 25, 2023 08:29 EDT)

Stella Batalama, Ph.D.
Dean and Professor, The College of Engineering and Computer Science

William David Kalies

William D. Kalies, Ph.D.
Interim Dean, Graduate College

April 25, 2023
Date

ACKNOWLEDGEMENTS

I would first like to thank Dr. Oge Marques for being the kindest, most supportive and understanding person I have had the pleasure to work with. His expertise within the field of healthcare and artificial intelligence combined with his guidance and wisdom, allowed me to broaden my knowledge and excel academically. Thank you for trusting and believing in me. Your dedication and commitment to my success, as well as your patience and unwavering encouragement, have been a constant source of motivation.

I would also like to thank my thesis committee, Dr. Furht, Dr. Neshenko, and Dr. Taebi, for their invaluable comments and insightful feedback that enabled me to refine and improve the quality of my thesis. I am very grateful for your time, effort, and continued support.

I would like to express my gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. None of these opportunities and accomplishments would have been possible without them. A special thank you to my younger sister, for being my ray of sunshine, and making this journey that much more pleasant!

ABSTRACT

Author:	Jennifer Gogova
Title:	Digital Transformation of Healthcare Using Artificial Intelligence
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Oge Marques
Degree:	Master of Science
Year:	2023

Digital transformation is rapidly changing the healthcare industry, and artificial intelligence (AI) is a critical component in this evolution. This thesis investigates three selected challenges that might delay the adoption of AI in healthcare and proposes ways to address them successfully. Challenge #1 states that healthcare professionals may not feel sufficiently knowledgeable about AI. This is addressed by Contribution #1 which is a guide for self-actualization in AI for healthcare professionals. Challenge #2 explores the concept of transdisciplinary teams needing a work protocol to deliver successful results. This is addressed by Contribution #2 which is a step-by-step protocol for medical and AI researchers working on data-intensive projects. Challenge #3 states that the NIH *All of Us* Research Hub has a steep learning curve, and this is addressed by Contribution #3 which is a pilot project involving transdisciplinary teams using *All of Us* datasets.

DIGITAL TRANSFORMATION OF HEALTHCARE USING ARTIFICIAL INTELLIGENCE

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1.1 MOTIVATION

AI has the potential to revolutionize healthcare, but it requires access to vast amounts of data for successful implementation. AI is data-hungry! Data is used to train models, by teaching them how to recognize patterns, make predictions or classify data based on specific criteria. Most of today's AI uses machine learning and much of what goes by AI today is deep learning which essentially, is a subset of machine learning. But how is this data collected and where does it come from?

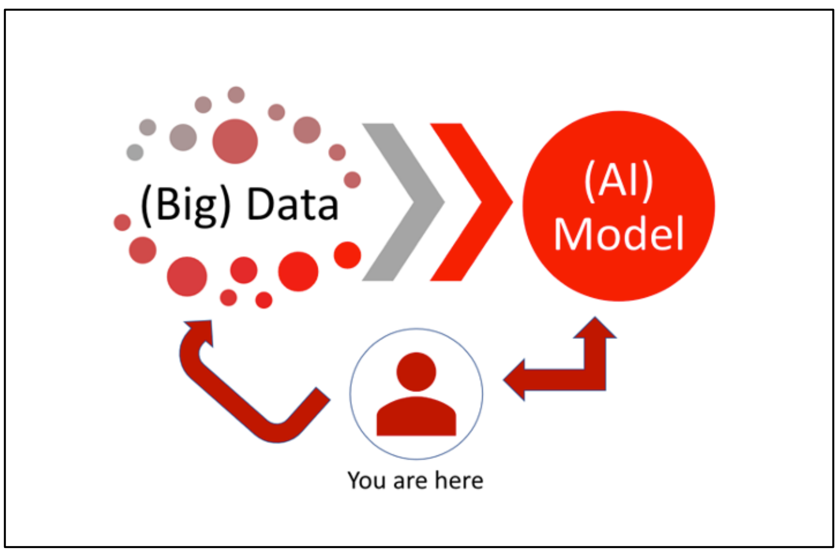


Figure 1 - AI models require large amounts of data that come from multiple sources, including the users of those models.

Data is everywhere. It can be found in digital devices, online platforms, public and historical records, scientific experiments, and through human input. It is a continuous cycle that includes the collection of data, training of the model, and evaluating and fine-tuning based on the results of the evaluation. This cycle continues until the team is satisfied with the results, and they can use and implement the model within the industry.

Digital transformation includes the incorporation of digital technology which has the potential to improve efficiency by allowing medical professionals to focus their efforts on patient care, while the completion of any administrative or autonomous tasks is already taken care of through the implementation of AI. This transformation can also improve patient care, innovation and add overall value. Easy access to data and healthcare records become more prominent as everything is stored in cloud-based data systems that can be accessed at any time and place. Through digital transformation and AI, there is a potential for improvement in diagnoses and treatment plans which ultimately benefits both healthcare professionals as well as patients. The concept holds many promises of potential advancements and artificial intelligence will play a big role in this transformation.

Despite the potential of artificial intelligence, it is still met with so much resistance. This could be due to a variety of reasons:

- Medical professionals lack knowledge and some fear that AI could potentially take over their jobs.
- Datasets can be questionable due to their size, quality, privacy, and representativeness/bias.

- Transdisciplinary teams are needed but that also comes with its own set of challenges – differences in culture, language, and skill set. Therefore, the implementation of a protocol is needed, which is a contribution explored in this thesis.

Although there are numerous challenges, the digital transformation of healthcare is expected to continue to evolve and disrupt the industry in the coming years. As artificial intelligence and machine learning technologies continue to advance, there will be more opportunities for medical professionals to expand their knowledge and ultimately improve patient care.

1.2 PROBLEM STATEMENT

As healthcare continues to evolve, the use of artificial intelligence is rapidly becoming a critical component of exploration. While AI has the potential to revolutionize healthcare delivery by improving accuracy, reducing costs, and enhancing efficiency, there is a limited comprehension of the challenges and opportunities posed by this transformation. Therefore, this thesis hopes to contribute to a better understanding of the potential behind the implementation of AI in healthcare and inform policymakers, healthcare providers, and patients about the implications of this transformation.

Although AI holds enormous potential for medical applications, its use has not become prevalent in everyday clinical practice. Several complex and interrelated challenges are impeding the translation of research into practical application. These challenges have created a lack of trust in AI-based solutions among key stakeholders such as healthcare professionals, patients, and regulators, which in turn reduces the likelihood of their adoption.

In this thesis, we choose to focus on a few selected challenges that might delay the adoption of AI and other technologies in healthcare.

1. Healthcare professionals may not feel sufficiently knowledgeable about AI.
2. Transdisciplinary teams need a work protocol for delivering successful results.
3. The NIH *All of Us* research platform (described in Chapter 4) has a steep learning curve.

1.3 MAIN CONTRIBUTIONS

This thesis demonstrates that the digital transformation of healthcare using artificial intelligence and data analytics is possible and provides concrete examples and guidelines on how to achieve/implement it using the NIH *All of Us* research platform.

In summary, the main contributions of this thesis are:

1. A **guide** for self-actualization in AI for healthcare professionals. This guide addresses Challenge #1 and is described in detail in Chapter 3.
2. A **step-by-step protocol** for medical and AI researchers working on data-intensive projects. This protocol addresses Challenge #2 and is described in detail in Chapter 5.
3. A **pilot project** involving transdisciplinary teams using *All of Us* datasets. This project addresses Challenge #3 and is described in detail in Chapter 5.

The work presented in this thesis is being adapted into two publications (in progress):

Paper 1

Authors: Megan Coffee, Kenny Moise, Oge Marques, Jennifer Gogova, Adam Wyatt, Janet Robishaw, and Michèle Retrouvey

Title: Hands-On AI Learning Pathway for Medical Professionals

Target journal and submission date: JAMA Network Open, May 2023

My contributions:

- Provided a curated reading list of recommended books.
- Writing and reviewing selected sections.
- Assisting with the bibliography.

Paper 2

Authors: Oge Marques, Tessa Harland, Jennifer Gogova, Adam Wyatt, Mindy Knowles, Christian Garbin, Nha Tran, Nicholas Marques, Chelsea Zuvieta, Daniel Cuentas, Richard Acs, and Matthew Acs

Title: Statistical Modeling of Prevalence and Social Determinants of Chronic Pain

Target journal and submission date: IEEE Access, June 2023

My contributions:

- Appointed as the liaison between the medical team and the DS/CS/AI team.
- Gathered key information to enable the beginning of the Workspace process.
- Assisting with the writing and reviewing of selected sections.

1.4 OVERVIEW OF THESIS

- Chapter 2 focuses on the background and context of digital transformation. It discusses the reasons behind the need for digitization and highlights how it is already being implemented. The chapter explores the breakdown of current digital transformation trends and does an in-depth analysis, examining the benefits of such technological advancements within the medical field. Furthermore, it traverses through the COVID-19 pandemic whilst conveying ways in which AI prevented the spread of disease, saved lives, and regulated the economy. Lastly, the chapter dives into medical education and the advantages of adopting artificial intelligence within the curriculum.
- Chapter 3 highlights the crucial role of bringing healthcare practitioners up to speed with the latest technologies, particularly those related to AI, and provides a carefully curated reading list for their self-actualization (Contribution #1).
- Chapter 4 provides a comprehensive description of the NIH *All of Us* initiative and guides readers on how to utilize the platform effectively.
- Chapter 5 focuses on transdisciplinary research in medical AI and argues for the judicious use of artificial intelligence, machine learning, and big data analytics. The chapter proposes a framework for integrating medical teams and data science teams to solve medically relevant problems using AI and publicly available datasets, such as *All of Us*. It accentuates the need for interdisciplinary teams, breaks down the interactions into a series of steps, offers templates for effective communication (Contribution #2), and provides resources to assist with the onboarding process of new researchers. The chapter also highlights an ongoing research project on “Statistical modeling of prevalence and social determinants of

chronic pain" and explains how the framework was successfully used to guide the work in that project (Contribution #3).

- Lastly, Chapter 6 provides lessons learned and concluding remarks.

CHAPTER 2

BACKGROUND AND CONTEXT

This chapter reviews the state of the art in technology adoption in healthcare and medicine. It also makes the case for the judicious use of AI, machine learning, and big data analytics along with the need to train professionals and narrow the gap between medical teams willing to carry out hypothesis-driven research and the computer science and data science (CS/DS) professionals who will mediate access to the datasets.

2.1 THE HEALTHCARE INDUSTRY IS IN NEED OF DIGITAL TRANSFORMATION

It has become clear that the digital transformation of healthcare is inevitable. In the medical field, AI is constantly evolving by offering numerous possibilities, including mass screening and accurate diagnosis. Technological advancements have led to complex computing abilities that allow for pattern recognition, relationship establishment based on input data, and comparison with the motivation of inducing performance standards. The use of devices for early disease screening or detection can reduce the need for trained medical professionals. Furthermore, digital transformation of healthcare can enhance patient care by providing accurate, efficient, and timely care through electronic health records (EHRs), telemedicine, mobile health apps, and wearables such as Fitbit. With various technological advancements, healthcare professionals can monitor patients in real-time, offer personalized care plans, and respond to emergencies much faster.

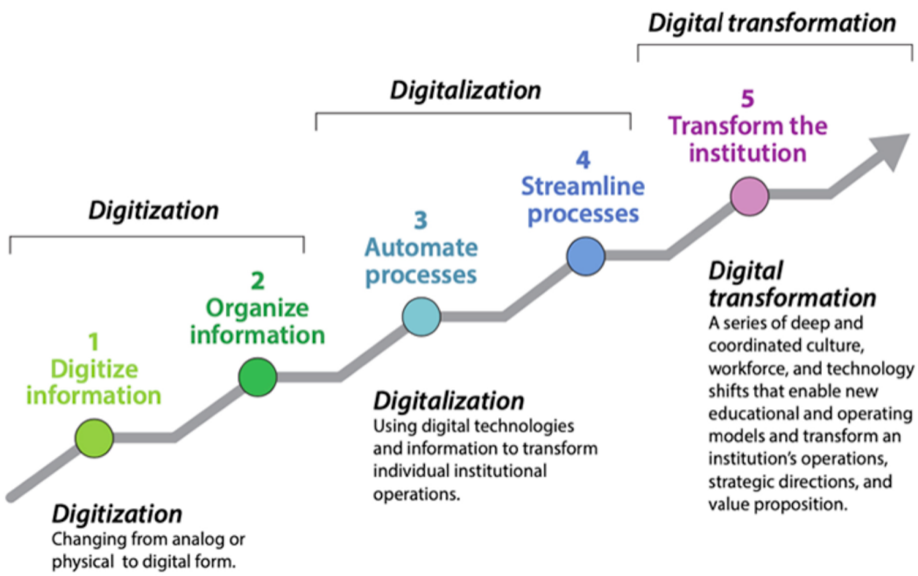


Figure 2 - A visual representation of the transition from Digitization to Digital Transformation [1]

Digital transformation has the potential to provide cost-effective healthcare by streamlining operations, reducing costs, and improving efficiency. For instance, using data analytics to analyze digitized medical records can help medical professionals identify high-risk patients, reduce readmission rates, and optimize care delivery. This benefits both the patient and the healthcare sector. Additionally, digital transformation can improve access to healthcare, particularly in remote or underserved areas where traditional medical facilities may be inaccessible due to distance, cost, or mobility issues. Telemedicine and remote monitoring technologies, such as wearables, can be particularly valuable in such circumstances. Moreover, healthcare providers have the ability to communicate and collaborate more efficiently, whether with their colleagues or patients. For example, EHRs allow healthcare professionals to have easy access to patient information and share this information with other medical professionals, resulting in improved decision-making and patient care/outcomes.

Technological advancements have made the digital transformation of healthcare possible. Although the concept is promising, it remains complex to fully implement. Several factors contribute to this evolution. The rapid increase of mobile devices has produced a connected ecosystem that enables real-time interaction between healthcare providers and patients.

Digital transformation within the healthcare industry can have a substantial impact on medical professionals, institutions, and organizations with the goal of providing better treatment for patients and improving overall operations.

2.2 BREAKDOWN OF DIGITAL TRANSFORMATION TRENDS

Health wearables have become a significantly valuable tool in monitoring health metrics. Medical professionals now encourage patients to wear technology devices such as Apple watches or Fitbit to track key vitals. These include things like oxygen saturation, energy expenditure, blood pressure, and sleep patterns. They can provide valuable information in terms of a patient's vitals and medical professionals can then review this data and analyze what might have caused an abnormality. This is possible due to digital transformation. A patient's vitals might be normal during a clinical visit so having a connected wearable device can aid healthcare professionals in capturing essential medical information whilst obtaining a more accurate assessment of a patient's health. Eventually, this improves overall treatment and outcomes. There was an instance where an Apple watch saved a man's life after he had been involved in an accident [3].

His watch alerted respective authorities by sharing his location and they were able to find him and save his life. Technological advancements not only collect and share valuable information that can improve one's health, but they can also play a big role in people's safety.

Patient portals are online platforms that enable individuals to have easy access to their health records, communicate with their healthcare providers, and schedule appointments in an easy and time-efficient manner. As both the patient and doctor have access to the portal, it promotes transparency and convenience. Furthermore, these portals include information about past examinations, prescriptions, and clinical visits that can be used to gather valuable information and improve diagnoses. The healthcare providers have access to their patient's prior visits and medical records as well as any visit notes from other medical professionals that can be used to better assess and treat the patient. Also, through digital transformation, there is no need to transfer or share medical records manually which is favorable for both parties – patients don't need to memorize or recall technical information provided by past examiners and doctors can easily access other healthcare professionals' notes to gain a better understanding of their patient's conditions. In addition, there's a part in the portal where the patient completes a combination of questionnaires to further explain their symptoms. Even though this part is deemed "subjective" by physicians reviewing the data, having access to previous examinations, test results, and feedback from healthcare providers, allows for more accurate diagnoses, better treatment plans, and more efficient consultations. According to a survey carried out by the Center for Connected Medicine, a large majority of healthcare professionals (82%)

conveyed by Figure 3, view patient portals as a crucial technological innovation for interacting with patients [4].

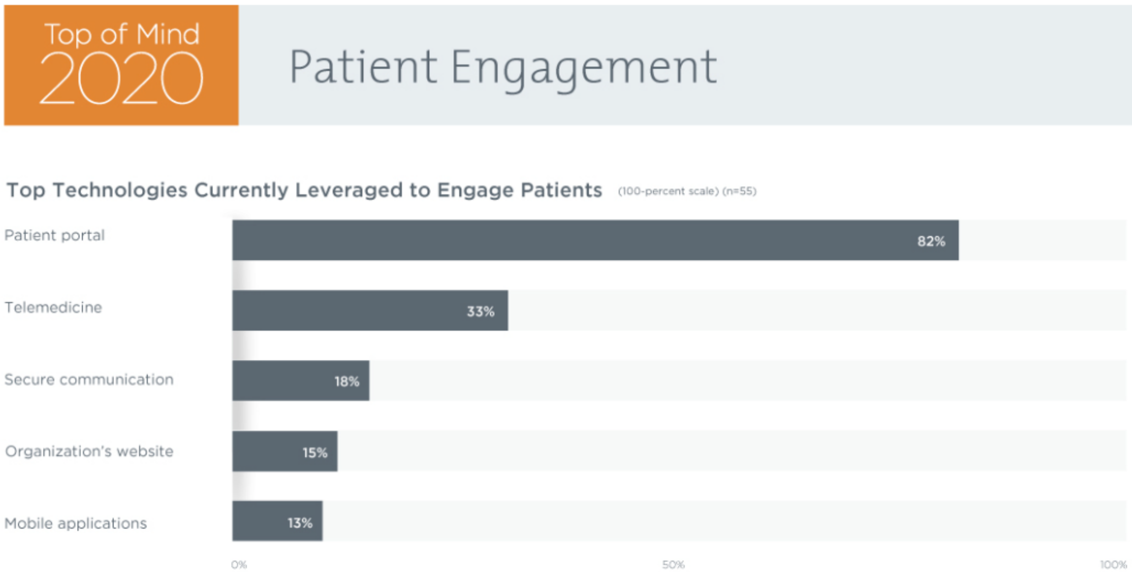


Figure 3 - Survey results investigating the Top Technologies Currently Leveraged to Engage Patients [4].

Furthermore, the portal allows physicians to review a patient's medical chart (chief complaint, notes from current consultation, history of present illnesses, associating symptoms, prescriptions, lab/bloodwork, X-rays, treatment plan, and follow-up schedule), which ultimately allows for the application of the best treatment plan. For instance, primary care physicians (PCP) may sometimes follow up with patients once or twice a year, therefore having access to the portal enables them to offer better service as they have had the chance to review the patient's data and information prior to their appointment. Maintaining a well-structured patient portal has proven to be very valuable for healthcare professionals and it resulted in clinics hiring medical scribes that take notes and organize the medical charts.

The development of education, overall knowledge, and access to resources has enabled the digital transformation within the healthcare sector to take place. Digital literacy allows patients to access and manage their electronic health more effectively. This process has increased patient empowerment as individuals have greater access to their personal health information, allowing them to make more thoughtful decisions about their health and lifestyle choices. They can easily communicate with healthcare professionals and discuss any concerns or queries they might have in a much more convenient and time efficient way.

Data aggregation allows hospitals to make efficient and insightful decisions without missing out on any valuable information. The healthcare industry consists of vast amounts of data that is received in multiple different ways, such as patient-provided data, lab results, EHRs, health metrics from wearables, and imaging data. The aggregation of data improves patient care whilst lowering overall costs. Moreover, it manages to create patient profiles which leads to shorter wait times as the time spent going through large quantities of data is significantly reduced. IBM Watson is an analytical data collection platform that recently partnered with Pfizer to conduct genomic data from cancer patients with the initiative of developing a new cancer drug based on their natural immune responses. By interpreting intricate genomic data, two companies were able to make predictions regarding distinct types of cancer as a result of IBM Watson's initiative.

Artificial intelligence screening is a growing trend of digital transformation within the healthcare field. AI screening can be quite broad. It can include things like medical screening, recruitment screening, security screening, financial screening, or social media screening. For instance, medical AI screening enables healthcare professionals to detect trends that could lead to early diagnosis of potentially serious conditions. On the other hand, the adoption of AI-based chatbots and voice systems reduces the administrative load on the medical staff. These innovations redirect patients to a variety of lines, such as appointment scheduling, lab results, pharmacy, and so on. Ultimately, this allows the medical staff to deal with more urgent queries and provide better care in a much more efficient manner. IBM Watson can be used to build, deploy, and optimize an AI chatbot. During the pandemic, Mass General Brigham hospital launched a COVID-19 Screener Chatbot that helped patients determine whether they required a COVID test. They found that the introduction of the chatbot reduced the number of people getting tested which ultimately prevented the spread of the disease.

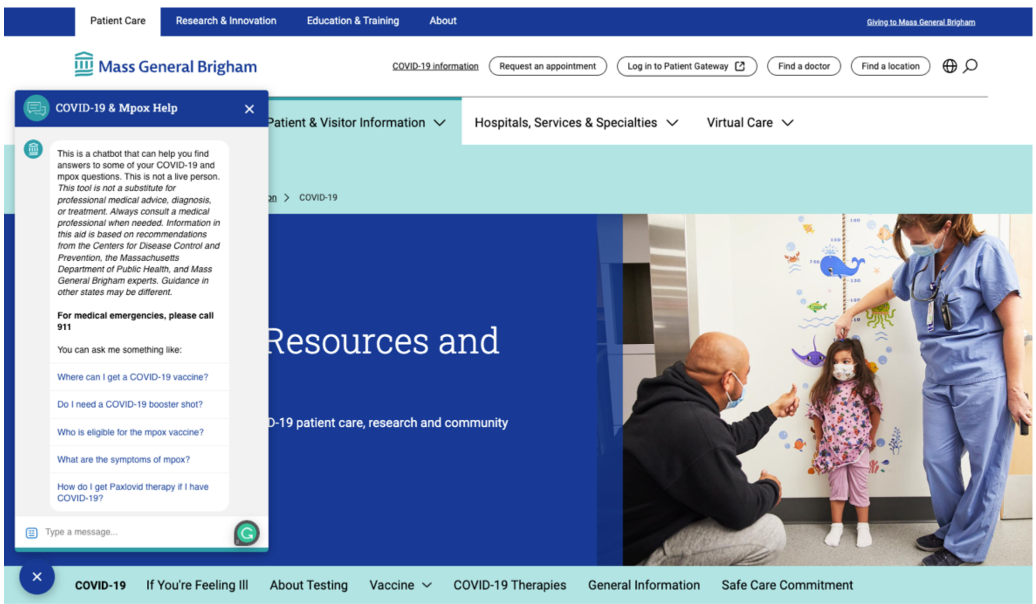


Figure 4 - Screenshot of the Mass General Brigham COVID-19 chatbot.

Digital transformation enables healthcare providers to utilize data analytics, artificial intelligence, and machine learning to identify key patterns and trends. This can potentially allow for early diagnosis of serious conditions. With the evolvement of technology, providers can better customize treatment plans that meet patients' needs and preferences. Data analytics includes an in-depth analysis of raw data and extraction of insightful verdicts. The availability of powerful data analytics tools has made it possible to collect, store, and analyze vast amounts of patient data. Within the healthcare sector, medical professionals can identify patterns, trends, and anomalies that can improve clinical decision-making and ultimately better patient care/treatment. Cloud computing is closely linked to data analytics, and it provides the required capacity to store and retrieve data. Digital transformation allows healthcare professionals to conveniently store and access patient data from the cloud at any time and from anywhere. This has significantly increased the opportunities for remote care provision and easy collaboration with other providers from various locations.

This thesis focuses on the digital transformation brought about by artificial intelligence. Advances in AI and machine learning have enabled this transformation in many industries, including healthcare. Medical professionals can now analyze patient data and make predictions about conditions or care plans, leading to more personalized care and better identification of high-risk patients. The use of wearables and IoT devices has also become more prominent, enabling healthcare professionals to monitor patients remotely and collect real-time data about their health. This can be extremely valuable in identifying and diagnosing conditions before they become critical or life-threatening. Electronic health records allow healthcare professionals to store, share, and access patient data in real-time leading to more efficient use of time and informed decisions about patient care.

2.3 ARTIFICIAL INTELLIGENCE AND COVID-19

March 2020 marked the beginning of a very significant transition in digital transformation and the world as a whole. The COVID-19 pandemic was very abrupt, and it had a tremendous impact on the healthcare industry. The rapid spread of infection, lack of resources, and the limitation of knowledge resulted in the inability to react which ultimately led to thousands of deaths.

Throughout the pandemic, scientists and healthcare professionals conducted endless research which led to the collection of large amounts of valuable data. COVID-19 resulted in the creation of coronavirus-related datasets as well as published papers that aided innovation in drug discovery, vaccine development, and public awareness. Artificial intelligence was used to detect and extract information of great worth. Machine learning, deep learning, and data analytics all played a role in finding trends and patterns that allowed medical professionals to discover various treatment options and vaccine development. Deep learning is under the AI umbrella, and throughout the pandemic, it helped accelerate the process of drug discovery through predictive modeling. Predictive models can take existing or new drugs and find certain properties that could potentially treat COVID-19. Artificial intelligence fulfilled its vast potential as it became one of the most used techniques to prevent the spread of disease, save lives and regulate the economy.

In 2020, research into AI became more prominent as organizations believed that technological advancements would enable medical professionals to tackle the virus. In the United States, the White House Office of Science and Technology Policy, the *COVID-19 Open Research Dataset (CORD-19)* contained over 29,000 machine learning readable articles regarding the coronavirus [5]. The White House strongly believed that AI was the key to combating the spread of the virus, so they actioned the nation's AI specialists to assist with any queries or developments around COVID-19 [5].

COVID-19 caused an unexpected shock to the world. The high infection rates as well as the rapid spread of the virus resulted in increasingly high mortality rates. Medical professionals couldn't take action as the virus was incomparable to other viral illnesses such as the flu, which evidently presented a lot of obstacles in getting the situation under control. There was an urgency to create a vaccine in order to combat the spread and in 2019, the National Institute of Allergy and Infectious Diseases funded the first clinical trial of an AI-based flu vaccine in the United States which was created by scientists at Flinders University [6]. At first, they used a program they created, called synthetic chemist, in order to create trillions of synthetic compounds, which was then accompanied by another AI program called Search Algorithm for Ligands (SAM) to identify suitable vaccine candidates [6], [7]. The use of artificial intelligence shortened the development of the vaccine by several years which is an incredible accomplishment within the medical field.

Artificial intelligence was used to screen available drugs at the time whilst simultaneously, identifying generic medications that could have potentially been used to treat COVID-19. Having datasets like the CORD-19 can be extremely beneficial when used to train AI algorithms. This is because it can use machine learning to screen for drugs that may demonstrate effectiveness in treating COVID-19.

Leveraging the power of AI allows for the rapid innovation and development of healthcare. Whether it involves the identification of drug and vaccine candidates or the prevention of economic crises, technological advancements have the potential in aiding clinicians and patients in extremely valuable ways.

2.4 ARTIFICIAL INTELLIGENCE IN MEDICAL EDUCATION

The lack of incorporation of artificial intelligence within the medical school curriculum can result in healthcare professionals falling behind in the rapid progression of patient care. Limited knowledge or comprehensibility of AI and machine learning can hinder physicians and their medical practices.

Training may not be provided for several reasons:

- Limited knowledge and uneven skillsets among faculty members.
- Insufficient research-based justification for meeting the increasing demand of students to study artificial intelligence.
- Absence of recommendations or regulations from the Liaison Committee on Medical Education regarding the integration of artificial intelligence into medical education.

However, misconceptions from skeptical individuals can prevent digital transformation and its applications from improving healthcare practices. Research shows that artificial intelligence can put an end to medical staff shortages and bring access to healthcare in remote areas [8]. This is one of the many motivations behind the integration of AI within medical school curriculums.

The application of AI has already provided numerous benefits within the healthcare industry. Although the power of AI for image recognition is evident in fields like radiology and pathology, research shows that it is also being utilized across many other fields of expertise [9]. As previously mentioned in section 2.1, data collected by wearables, smartphones, and other mobile monitoring sensors across different medical domains can be used to train deep learning algorithms and ultimately draw insightful conclusions that can improve patient care [10].

Certain advancements in AI present challenges in the healthcare industry, in particular, the resistance from healthcare professionals who may fear that the progression of digital transformation can lead to machines dominating the industry. However, the lack of understanding, along with the distrust towards technological upgrades, can obstruct the adoption of AI within the medical field. Consequently, this could lead to companies being at a competitive disadvantage due to their inability to adopt and accept change [10]. On one hand, this is understandable as AI will potentially change the traditional relationship between clinicians and patients, creating a more complex dynamic that raises ethical, legal, and financial concerns [11]. Medical education must evolve with the effort of preparing future healthcare professionals to deal with new responsibilities and manage technical tasks [12]. Moreover, medical schools must adapt to provide students with the knowledge and skills needed to thrive in this technologically dominated landscape, while also emphasizing unique human abilities that give physicians a competitive edge over machines [13], [14]. It is key to educate future physicians on how to interact with this technology. This way, physicians might be more open to the transition as they will be working alongside digital transformation without fear of losing their jobs.

For this integration to happen, undergraduate medical education (UME) is necessary. Medical education deans and governing bodies should discuss and plan an AI curricular reform, and the Liaison Committee on Medical Education should provide guidelines for incorporating AI into medical education [15]. Future medical professionals need to have a comprehensive understanding of the advantages, disadvantages, liabilities, and ethical considerations of AI-based implementations within healthcare. This can ensure that physicians will be able to make informed decisions and deliver quality treatment and care to their patients.

Medical schools should consider the incorporation of a curriculum that explores and discusses artificial intelligence tools. Medical students will highly benefit from understanding the frameworks of engineering and designing AI solutions related to their specific fields of study. Having a solid foundational knowledge of the role that data plays in the development of AI innovations becomes extremely valuable and aids students in making critical decisions that ultimately improve their performance and overall patient care. It is anticipated that AI will enhance its ability to analyze various types of data and offer practical advice to healthcare professionals [16]. Furthermore, it is expected that the progress in technology and digital transformation will enable AI efforts to be integrated into a majority of medical aspects. Students can help drive change by reading about AI advancements in medical journals and exposing themselves to AI in the clinical setting. Chapter 3 provides a carefully curated reading list for medical professionals looking to make the transition into AI and achieve self-actualization.

CHAPTER 3

AI KNOWLEDGE FOR MEDICAL PROFESSIONALS: A GUIDE FOR SELF-ACTUALIZATION

This chapter explores challenge #1 which states that healthcare professionals may not feel sufficiently knowledgeable about AI. It provides a guide for self-actualization in AI for healthcare professionals and discusses some benefits that stem from the adoption of AI within medicine. The motivation behind contribution #1 stemmed from the following papers from our research group:

- Educating Future Physicians in Artificial Intelligence(AI): An Integrative Review and Proposed Changes [17] (Authors: Joel Grunhut, Adam TM Wyatt, Oge Marques)
- Needs, Challenges, and Applications of Artificial Intelligence in Medical Education Curriculum [18] (Authors: Joel Grunhut, Oge Marques, Adam TM Wyatt)

Both papers address the need to integrate AI to further develop the medical education curriculum along with improving physicians' overall knowledge and skill set. They discuss the demand to educate medical professionals in order to increase the value of patient care and the overall development of medicine.

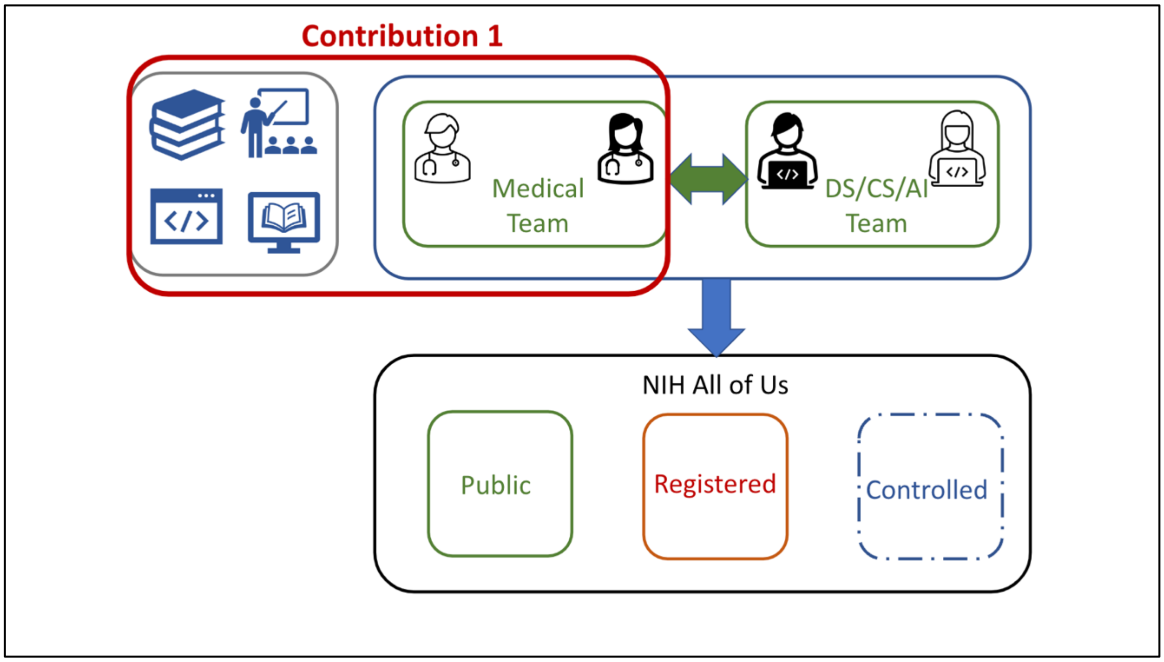


Figure 5 - An overview of the vital and continuous interaction between the Medical Team and the DS/CS/AI Team, who work together and use the NIH All of Us datasets to answer the Medical Team's research questions. Contribution #1 is a guide for self-actualization for medical professionals.

3.1 BENEFITS FOR THE MEDICAL PROFESSIONALS

As artificial intelligence (AI) becomes more prominent in healthcare, practitioners must stay up to date with the latest technologies for a variety of reasons. Knowledge of AI-related concepts can potentially improve patient care by allowing medical professionals to provide better care, including early diagnosis, personalized treatment plans, and remote patient monitoring. If machine learning can detect the early onset of a condition, medical professionals can promptly configure a plan of action that is in the best interest of their patient's health. Late diagnosis is a contributing factor to fatalities, but with digital transformation and overall technological advancements, medical professionals can prevent this from happening. Additionally, AI can help healthcare professionals work more efficiently by automating certain tasks revolving around administrative work or image analysis.

If doctors are trained effectively, they can save time and put more effort towards helping patients rather than doing autonomous tasks. AI has the potential to allow medical professionals to make better decisions by providing them with good quality data which will eventually enable them to make more informed decisions regarding treatment plans and future research recommendations. Due to the undergoing digital transformation within the healthcare sector, medical professionals who understand how to work with AI systems will become more competitive within their job market. By learning and expanding their knowledge of artificial intelligence, medical professionals improve their skill set, career prospects, and patient treatments. Medical professionals being exposed to AI can also be valuable when considering the ethical implications of the technologies being utilized. By learning and researching AI further, medical professionals can ensure that they are using these technologies in an ethical and responsible manner. At the end of the day, medical professionals have a responsibility to keep their patients safe and prioritize their privacy in any way possible.

3.2 A CURATED LIST OF RECOMMENDED BOOKS

To better familiarize myself, I reviewed the literature and produced a summary of eight books on the intersection of AI and healthcare, organized into two groups:

(i) Broader conceptual knowledge and overview of hype, hope, perils, and words of advice (Table 1).

(ii) Coding-oriented for those interested in implementing real AI solutions for selected healthcare problems (Table 2).

Book	Takeaways
Topol, 2019 [19]	Highlights a shift from shallow to deep medicine in healthcare while discussing the benefits of using AI in the industry. Although AI has its limitations, doctors can use it to improve diagnoses, especially those based on pattern recognition. It can be beneficial for doctors who don't work with patterns as AI can be used to perform routine tasks. AI can also reform health systems, improve research, personalize medicine and diets, and automate clinical functions to allow doctors to focus on patient care.
Meskó, 2019 [20]	The evolution of artificial intelligence has greatly impacted the healthcare industry from a futuristic promise to an innovation benchmark. Emphasis on the need to discuss and understand the contributions of AI in medicine and to anticipate its potential use cases. Also addresses the importance of recognizing and controlling the risks and implications of AI in healthcare.
Bohr and Memarzadeh, 2020 [21]	Discussion of various data techniques in healthcare data analysis, including machine learning and data mining. Provides examples of their applications and challenges in designing, implementing, and managing intelligent systems and healthcare data networks. Includes case studies across all areas of AI in healthcare data.
Matheny et al., 2020 [22]	Explores the opportunities offered by the emergence of AI in healthcare, including improved patient outcomes, reduced costs, and better population health. Looks at potential risks around user disillusionment, another AI winter, and exacerbation of existing disparities. There is a need for thoughtful, inclusive AI applications that consider potential unintended consequences and balance profit motives by managing and reducing risks.
Holley and Becker, 2021 [23]	Explores a variety of topics related to the use of artificial intelligence in healthcare. Discusses myths and realities of AI, the importance of human-centered AI, and the use of AI technologies beyond precision medicine. Considers the use of the internet of things (IoT) as well as ambient computing with AI to deliver patient care. It looks at how AI can help reduce waste in healthcare and the importance of an AI strategy to identify high-priority AI applications.

Table 1 - Recommended books (part 1: basic/conceptual)

Book	Takeaways
Anshik, 2021 [24]	Investigates how machine learning and deep learning tools can be applied to the healthcare industry. Includes a clear coverage of algorithms and techniques with case studies. Explores different problem areas within the industry and solves them using a code-first approach. Covers advanced topics such as multi-task learning, transformers, and graph convolutional networks. The target audience for this book is data scientists and software developers interested in machine learning and its application within the healthcare industry.
Shimonski, 2021 [25]	A comprehensive analysis of essential topics in healthcare applications of artificial intelligence, such as healthcare IT, AI clinical operations, project planning, metrics, reporting, automation, cloud operations, and the future of AI. Written in an accessible and straightforward style, making it valuable for IT managers, administrators, and engineers in healthcare settings, as well as anyone with an interest or stake in healthcare technology.
Nguyen, 2022 [26]	Explores different types of healthcare data, including electronic health records, clinical registries and trials, digital health tools, and claims data. Discusses the challenges of working with such data, particularly when aggregating it from multiple sources. Analyses current options for extracting structured data from clinical text and making tradeoffs when using tools and frameworks for normalizing structured healthcare data. Highlights the importance of harmonizing healthcare data using terminologies, ontologies, mappings, and crosswalks.

Table 2 - Recommended books (part 2: technical/applied)

3.3 ADDITIONAL REMARKS

It should be noted that some MDs are also PhDs with coding proficiency, Kaggle grandmasters, and other skills. They have taken the lead in educating their peers on AI, machine learning, deep learning, and related topics with numerous courses, certificates, events, and specialized columns in medical journals, such as the Magician's Corner in the RSNA *Radiology: AI* Journal. However, there are also a lot of healthcare professionals who are resistant to the adoption of AI. There is a misunderstanding and fear that these technological advancements within the field will take away jobs and completely transform patient care. Therefore, the incorporation of AI within the medical education curriculum or a simple guide to self-actualization can play a vital role in allowing physicians to have early exposure to advanced frameworks specific to engineering and designing AI solutions within their field of practice. Medical professionals who are knowledgeable about the development of AI innovations, and can comfortably utilize a variety of resources, will become extremely valuable assets to the healthcare field as their skill set will improve overall patient care and allow for the adoption of AI to become more prominent.

CHAPTER 4

THE NIH *ALL OF US* INITIATIVE

This chapter will provide a representative overview of the NIH *All of Us* initiative including a detailed breakdown of the main components and how they can be utilized by medical professionals.

4.1 GUIDED TOUR TO *ALL OF US*

All of Us is an NIH (National Institute of Health) research program [27], building one of the largest biomedical data resources. It stores health data from a diverse group of participants from across the United States. Registered researchers can access *All of Us* data and tools to conduct studies and ultimately improve and develop their understanding of human health. There are several benefits to using the *All of Us* Research Hub [28]. Authors have noted that size and diversity, along with the availability of multiple coding templates and sample code from other demonstration projects, were the most significant attributes when utilizing the platform. *All of Us* can be used to identify, validate, and most importantly learn how to better address health disparities. This research program is inviting a variety of people from all over the United States with the goal of building the most diverse health database system in the world. Currently it includes 609,000+ participants, 358,000+ electronic health records and they have received 438,000+ biosamples so far [29]. The data is used to learn how our biology, lifestyle and environment affect our overall health. The *All of Us* research program is hoping to develop and improve research with the goal of finding ways to prevent and treat diseases. Thanks to the blooming of technology and research, the medical field can progress in extraordinary ways.

In the Summer of 2022, I was among the first researchers to gain registered tier access into the *All of Us* Researcher Workbench. The *All of Us* is very strict and protective over their participants' privacy so the security of data is key. In order to become a registered tier member, I had to undergo training and complete a knowledge-based exam prior to gaining access to the registered tier data. The program stands out through its inclusivity and diversity. This early exposure enabled me to better familiarize myself with the platform and work with Dr. Marques on projects that revolved around the development of workflow tutorials as well as a guide on how to utilize the platform and datasets to their full potential. This exposure established a solid foundation that ultimately allowed the implementation and completion of all three contributions in this thesis.

4.2 MAIN COMPONENTS

4.2.1 DATA BROWSER

The Data Browser provides an interactive environment where views are exchanged about the participants' data from the *All of Us* research program. Participant provided information includes data from surveys, wearables, physical measurements (taken at the time of enrollment) and electronic health records (reported by healthcare providers, not participants). Protection of data is extremely important and to ensure there won't be a breach of privacy, the data is de-identified, limited to rounded aggregate data counts of 20, and only the summary demographic information is included.

4.2.2 DATA SNAPSHOTS

The Data Snapshots [29] provide an overview of the participants' characteristics and the types of data collected from the participants. They showcase the scale and diversity of the *All of Us* Research Program participant cohort and the snapshots provide information regarding participants' demographics, geographic distribution, and more. It must be noted that this information is updated daily.

Currently, the data points to the following:

- More than 919,000 people [29], have registered with the program.
- The highest number of participants are from California (13.63%) and Arizona (13.87%).
- Florida, Alabama, Texas, Illinois, Wisconsin, Michigan, Pennsylvania, New York, and Massachusetts also have significant participation rates.
- The map reflects the number of participants in each state who have completed the initial steps of the program. Counts are updated daily.

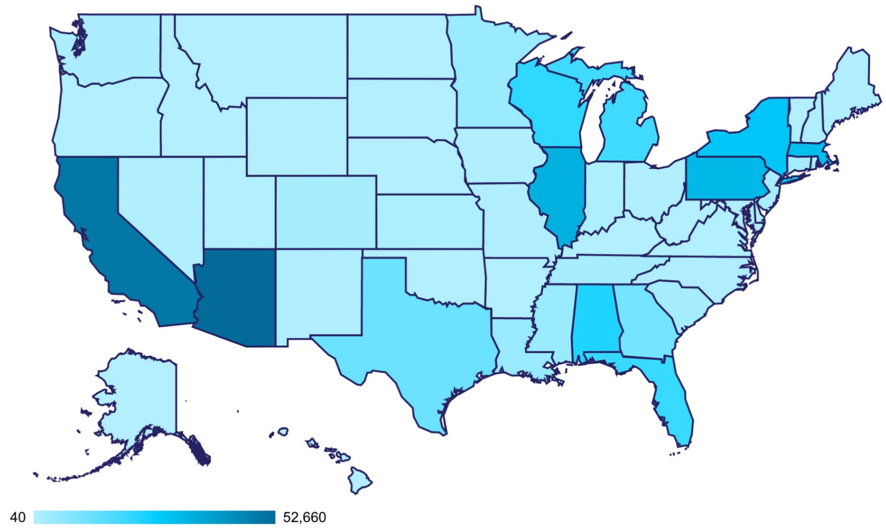


Figure 6 - Map reflecting the number of participants in each state having completed the initial steps of the program.

- Diversity includes 50%+ racial and ethnic minorities as well as sexual and gender minorities, people with low income or limited education, and other groups. It encompasses 80%+ of underrepresented groups in biomedical research.
- The graph below represents the self-reported races and ethnicities of participants who have completed the initial steps of the program. The information is gathered from the program's Basic Survey [30].

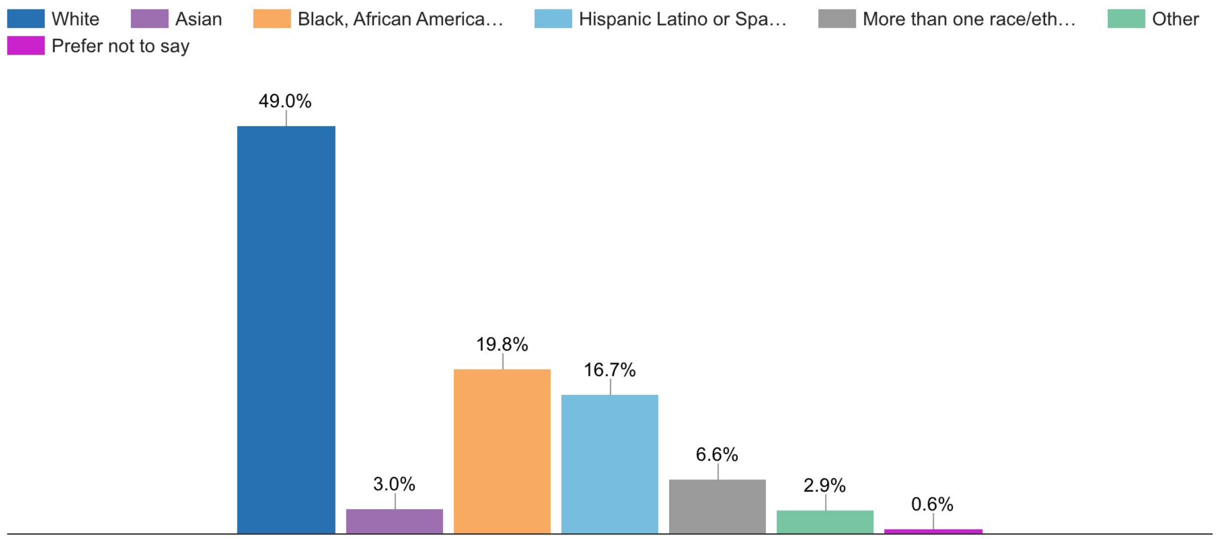


Figure 7 - Bar graph used to represent the self-reported races and ethnicities of participants who have completed the initial steps of the program.

- The graph below represents the self-reported gender identities of participants who have completed the initial steps of the program. The information is based on participants' responses to the Basic Survey [30].

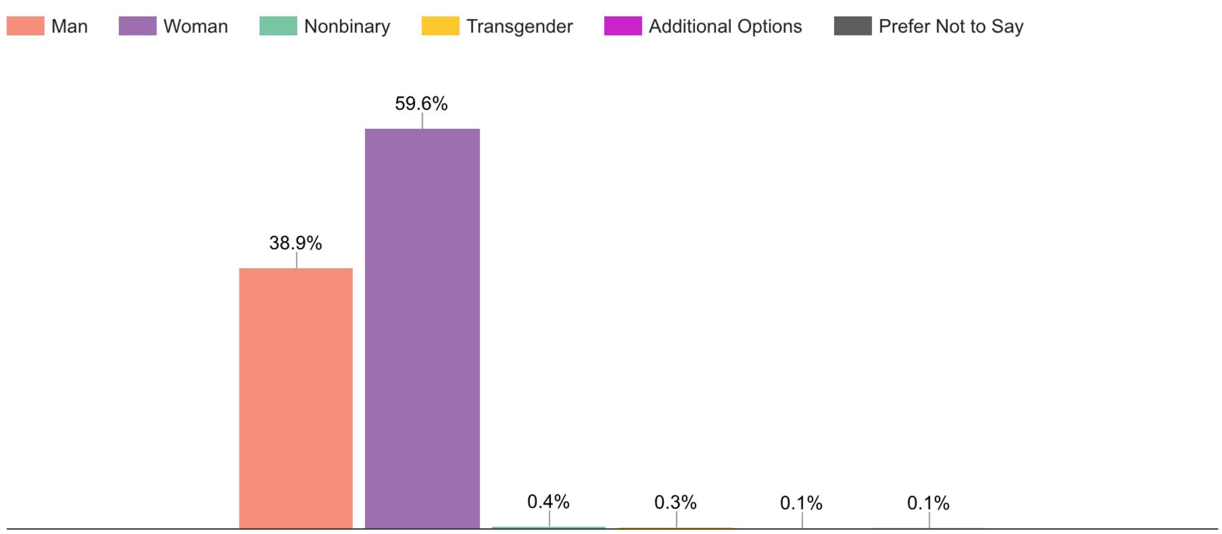


Figure 8 - A bar graph used to represent the self-reported gender identities of participants who have completed the initial steps of the program.

- The graph below represents the age at the time of enrollment of participants who have completed the initial steps of the program. Age ranges are provided to protect participants' privacy.

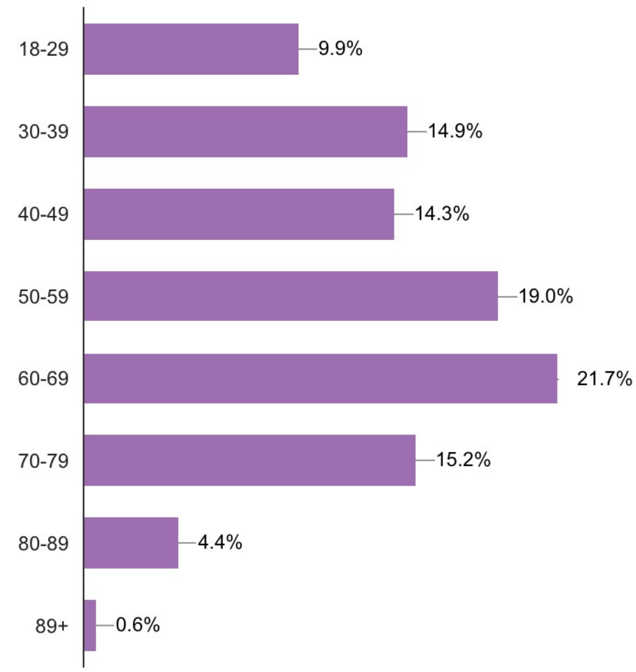


Figure 9 - A bar graph used to represent the age at the time of enrollment of participants who have completed the initial steps of the program.

4.2.3 DATA ACCESS TIERS

The *All of Us* Research Hub uses a tiered-data access model [31]. There are three tiers – public, registered, and controlled.

- The **Public Tier** contains only aggregate data with removed identifiers. It is available to everyone through Data Snapshots and the Data Browser.

- The **Registered Tier** is a curated dataset containing individual-level data. It is only available to approved researchers on the Researcher Workbench (including FAU researchers with a currently valid NIH login). Approved researchers must go through a background check, followed by the *All of Us* provided training, and then pass an exam testing their overall knowledge about the program.
- The **Controlled Tier** contains genomic data in the form of whole genome sequencing (WGS) and genotype arrays, previously suppressed demographic data fields from EHRs, surveys, and unshifted dates of events.

4.2.4 DATA SOURCES

The program aims to enroll one million or more people across the US to help build one of the largest and most diverse health databases in history. *All of Us* is looking to include and encourage people and communities who have been left out of medical research in the past. They also make their data available to a diverse set of researchers in order to improve the overall understanding of health, disease and other medical conditions with the goal of reducing future health disparities. Participants share data from multiple sources including; electronic health records, biosamples [32] and bioassays (used to generate genomic data [33]), surveys [34], physical measurements [35] and mobile health data (wearable devices like Fitbits).

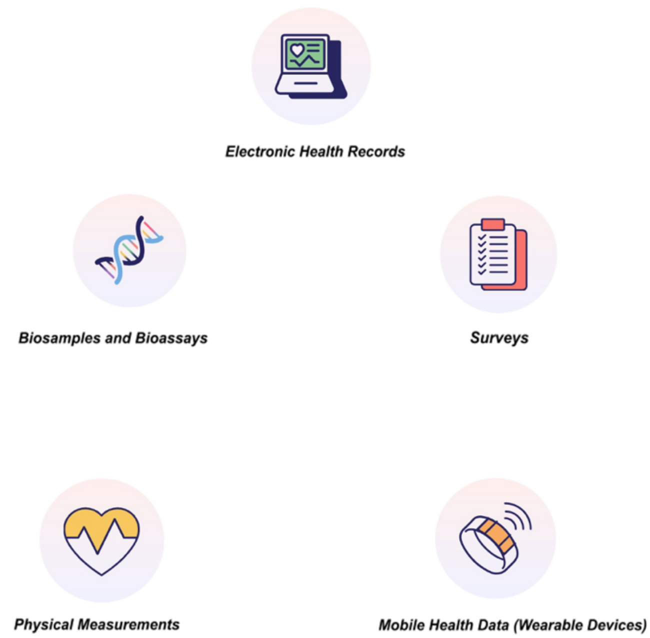


Figure 10 - Visualization of all data sources available on the All of Us research platform.

4.2.5 DATA METHODS

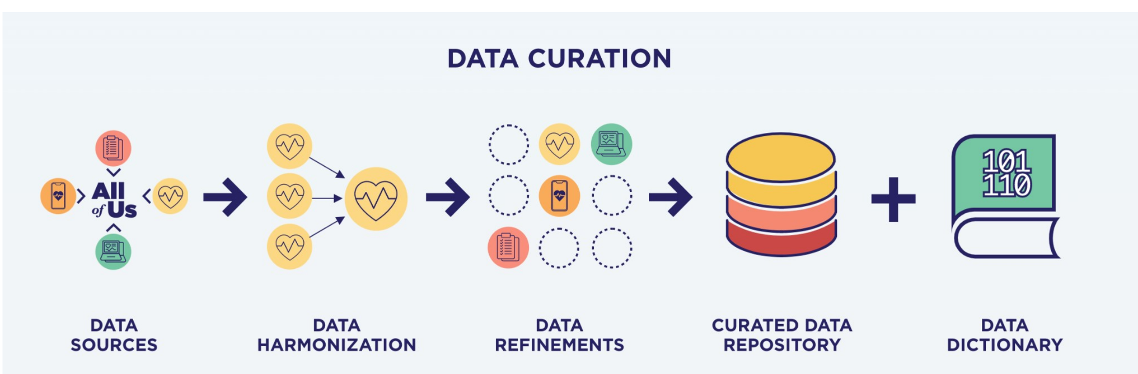


Figure 11 - Visualization of the data curation process.

The Data Curation [36] is a lengthy process. As stated in the previous section, the first step is the Data Sources which include surveys, electronic health records, biosamples, physical measurements and wearables. The *All of Us* Research program utilizes the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) which is used to standardize EHR data for all researchers. Harmonizing the EHR data is crucial as it must meet the set requirements of the OMOP CDM that ensure participants' privacy is protected. During the harmonization, the data is conformed and cleaned in order to provide high quality data to the researchers. The Curated Data Repository is where the data is presented in its final format after the harmonization refinement. There are three different levels of information available on the *All of Us* Research Hub; Public, Registered and the Controlled tier (more information about each tier is provided in the Data Access Tiers section). Finally, the Data Dictionary is used to document the exact data that is available from participants, including the modifications the program makes in order to protect participants' privacy. Within the Data Dictionary, you can see a description for each data field, noting whether it is a standard OMOP field or a custom field. This is created in order to capture data that is unique to the program. It is an extremely useful tool as it provides information on where the data in each field comes from. Researchers can see whether the information provided is from participant health records or individual reports such as surveys. Furthermore, the Data Dictionary clarifies how the *All of Us* Research program cleans the data to improve its quality as well as many of the program's custom concept IDs for easy reference. This includes versioning data, so researchers have access to see exactly what has been changed, added, or removed since the previously curated dataset.

The Observational Medical Outcomes Partnership (OMOP) [37] Common Data Model (CDM) is maintained by an international collaborative called the Observational

Health Data Sciences and Informatics (OHDSI) program. The *All of Us* Research program consolidates the OMOP CDM to empower researchers by using existing, standardized vocabularies and a harmonized data representation. More information focusing on OMOP databases, resources, and the OMOP tables being used is provided on their website (https://ohdsi.github.io/CommonDataModel/).

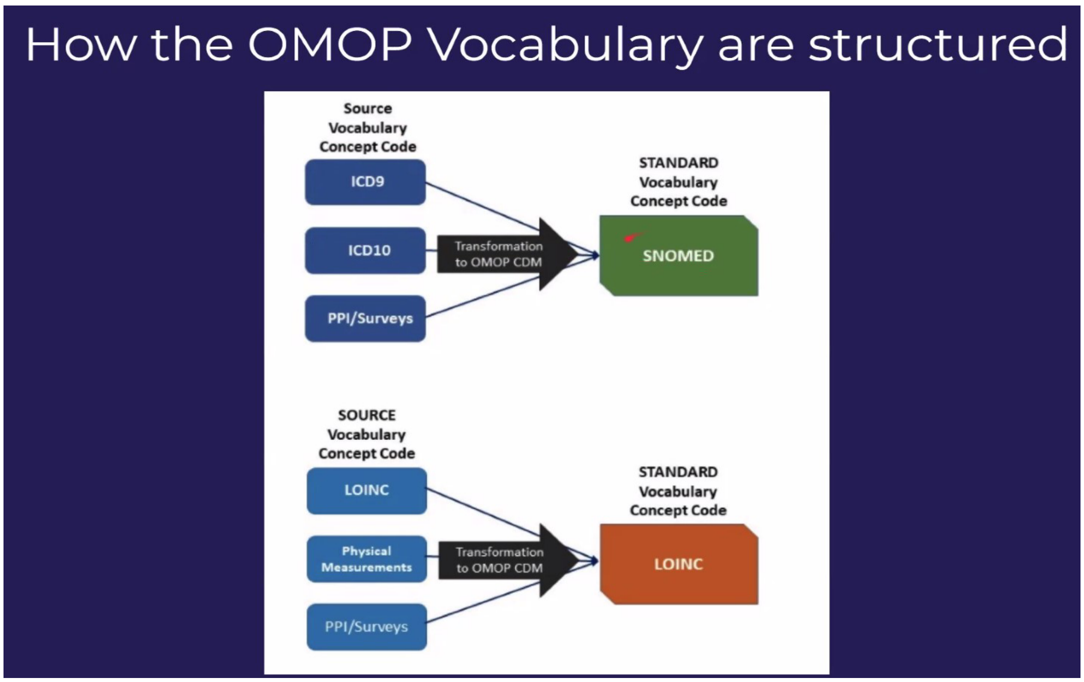


Figure 12 - Breakdown of how the OMOP Vocabulary are structured.

In order to integrate wearables [36] for data collection, the *All of Us* participants who obtain a Fitbit device are encouraged and more than welcome to share their data with the program. They have the freedom to select the type of data they share and are also able to stop sharing at any point. For clarification purposes, a Fitbit is an activity tracker that tracks your activity levels. It reports things like the distance you walk, run, swim, or cycle. It also looks at the number of calories you burn and take in. According to the Fitbit you have, some can also monitor your heart rate and sleep quality. Fitbit data is actually the first of this data type to be included in the Researcher Workbench. It is very promising and has a lot of potential to improve the understanding of researchers as they get a detailed insight into the participant's lifestyle. For example, if a participant has a severe case of anxiety, researchers can use the wearable data to explore whether certain things like lack of sleep, or types of exercise can trigger or worsen their symptoms. The fact that some Fitbits can monitor your heart rate can be very valuable information especially when it comes to anxiety or panic attack-related conditions. Technology is advancing in unimaginable ways which can hopefully result in definitive cures of serious medical conditions.

4.3 SURVEY EXPLORER

Surveys [34] are a valuable research tool in general but they are of particular benefit within the medical field. They efficiently capture information that is vital for a variety of research purposes. In the *All of Us* program, the surveys look at a variety of topics including demographics, healthcare and lifestyle. According to their website, the program has tested each survey for readability and accessibility. They use cognitive interviews and quantitative testing while making sure to include a wide range of educational backgrounds and geographic locations to accurately represent the entire US population. Their core surveys include The Basics, Overall Health and Lifestyle. There is the option to complete other surveys such as Personal Medical History, Health Care Access and Utilization, Family Medical History and COVID-19 Participant Experience. All surveys are offered in both English and Spanish.

4.4 RESEARCHER WORKBENCH

The workbench [38] is a cloud-based platform where registered researchers can access Registered and Controlled Tier data. It has powerful tools that support data analysis and collaboration. The workbench includes **Workspaces** which are a collaborative environment, where registered researchers can access, store, and analyze data for specific research purposes. This allows them to organize research projects and join forces in order to tackle more challenging tasks. **Notebooks** are where researchers with R or Python experience can perform high-powered analysis within the *All of Us* datasets using the integrated, cloud based Jupyter Notebook environment. This is followed by the **Dataset builder** that allows researchers to search and save collections of health information about cohorts called concept sets. It is extremely beneficial when performing pre-populated analysis and dataset reviews. The **Cohort Builder** is a custom point-and-click tool allowing researchers to create, review and annotate groups of participant data or cohorts within the dataset. The primary focus of the cohort builder is ultimately cohort creation. The **Workbench User Support Hub** contains a variety of resources, for example, training materials, that demonstrate how to use the tools and navigate the data within the Researcher Workbench. This is used for learning, support, and guides.

4.5 RESEARCH PROJECT DIRECTORY

The research project directory [39] includes information about all current projects within the Researcher Workbench. This is used to provide transparency about how the workbench is being used. Within the directory, each project specifies whether **Registered Tier** or **Controlled Tier** data is being used. When looking up “Chronic Pain” for example, you see that there are 13 projects with “Chronic Pain” in the title. These projects take into consideration ethnicity, age, and drug dependence. You have to submit a request to review these research papers. In the drop-down bar you have: Scientific questions being studied, Project purpose, Scientific approaches, Anticipated findings, Demographic categories of interest, Data set used and Research team.

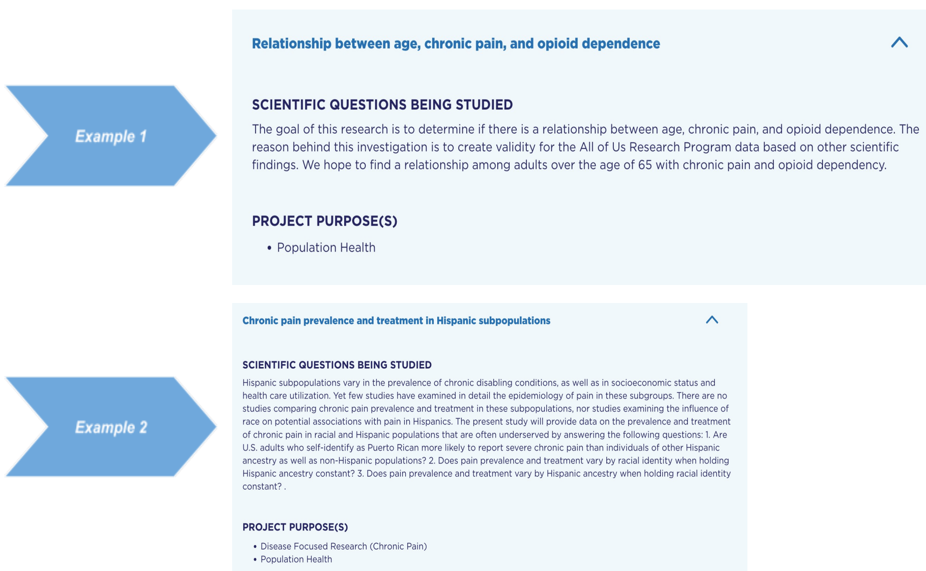


Figure 13 – Visual representation of search output for “chronic pain” within he Research Project Directory.

4.6 SUMMARY

This chapter provided a comprehensive description of the NIH *All of Us* initiative and offered guidance on how to utilize the platform effectively. One of my previous contributions related to working with the *All of Us* research platform consisted of creating a comprehensive document that was used as a guide to provide a detailed breakdown of the workbench tools. This included instructions on creating a workspace, creating cohorts and datasets, and exporting and utilizing the Jupyter Notebook(s). This endeavor involved an in-depth exploration of the *All of Us* platform and proved to be highly beneficial for gaining proficiency in utilizing the available tools. Subsequently, my work was utilized to generate a report, which was followed by the production of a video that was presented to the Fall 2022 class instructed by Dr. Marques in Artificial Intelligence in Medicine and Healthcare. The students were introduced to the *All of Us* initiative and were assigned various projects throughout the semester that revolved around the use of data in the Public Tier. The video served as a useful guide, enabling the students to navigate the platform in a more efficient and effective manner.

Chapter 5 will propose a framework for integrating medical teams and data science teams to solve medically relevant problems using AI and publicly available datasets, such as *All of Us*. It accentuates the need for interdisciplinary teams, breaks down the interactions into a series of steps, offers templates for effective communication, and provides resources to assist with the onboarding process of new researchers. The chapter also highlights an ongoing research project on "Statistical modeling of prevalence and social determinants of chronic pain" and explains how the framework was successfully used to guide the work in that project.

CHAPTER 5

A FRAMEWORK FOR OPTIMIZING THE WORKFLOW OF TRANSDISCIPLINARY RESEARCH TEAMS IN MEDICAL AI

This chapter proposes a framework for integrating MD teams and DS teams to solve medically relevant problems using AI, DS, and a publicly available dataset, in this case *All of Us*. In Section 5.1, we motivate the need for such interdisciplinary teams, break down the interactions into a series of steps, provide templates for documents that can be used for effective communication and workflow, and offer resources to assist with the onboarding process of researchers who are new to the registered tier. In Section 5.2, we show an example of an ongoing transdisciplinary project at FAU in the field of chronic pain, evaluating the transdisciplinary teams working together as well as partial results produced by the DS/CS/AI team.

5.1 BUILDING AND ORGANIZING THE TEAMS

The minimal team for successful research in Medical AI should comprise of a Data Science/CS/AI expert (henceforth referred to as "the data science team") and a medical professional with subject matter expertise (henceforth referred to as "the medical team").

5.1.1 FIRST THING FIRST

The guiding principles behind the proposed protocol were:

- Communication is key.
- Each team's expertise is unique.
- The expertise of the medical team should complement the expertise of the data science team.
- An inevitable consequence of different backgrounds: teams will speak different languages!

To promote effective communication, we adopted a "turn-taking" scheme similar to a 2-player game, (even and odd-numbered steps) reflected in section 5.1.2 – The Protocol. Effective communication between the two teams is key! Having ongoing discussions enable the data science team to extract valuable information efficiently from the datasets that can then be used to answer the medical team's research questions.

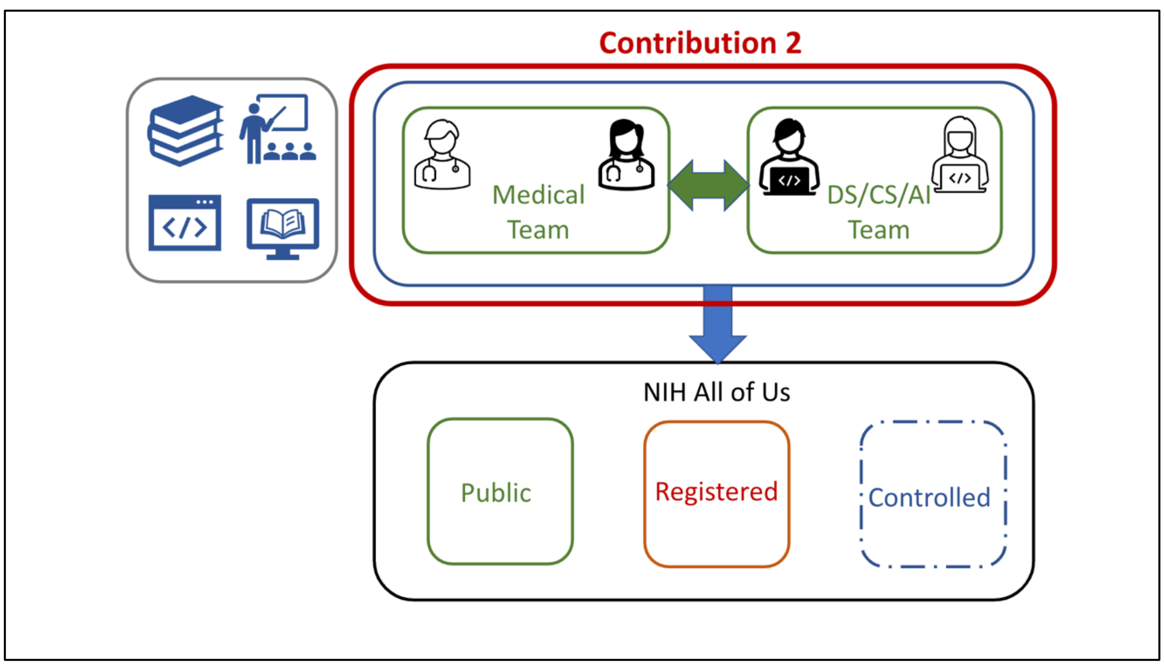


Figure 14 - An overview of the vital and continuous interaction between the Medical Team and the DS/CS/AI Team, who work together and use the NIH All of Us datasets to answer the Medical Team's research questions. Contribution #2 is a step-by-step protocol that allows for effective communication between medical and CS/DS/AI researchers working on data-intensive projects.

5.1.2 THE PROTOCOL

Step number	Team in charge	Scope
1	Medical Team	Provide information to start **Workspace**
2	DS/AI Team	Start a new workspace
3	Medical Team	Preparation for **Cohort** creation
4	DS/AI Team	Cohort creation
5	Medical Team	Preparation for **Dataset** creation
6	DS/AI Team	Dataset creation
7	Medical Team	Provide detailed explanation as to what the research questions are
8	DS/AI	Create **Notebook**(s)

Figure 15 - A breakdown of The Protocol conveying the step number, corresponding team in charge, and the scope at each stage.

1. Medical Team communicates research goals, and needs, to the DS/AI team.
2. DS/AI team follows a detailed workflow to build a study using the *All of Us* Research hub at the Registered Tier.
3. A model + report + notebook are produced.
4. The DS/AI team submits reports/findings/questions to Medical Team.

The cycle repeats until the Medical Team is satisfied.

The loop is closed, and the work is published.

5.1.3 HOW TO USE WORKBENCH TOOLS

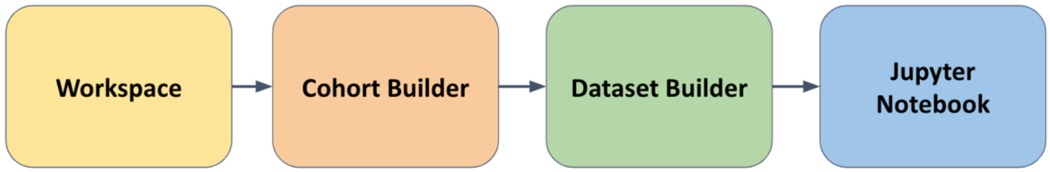


Figure 16 - Breakdown of what the workbench tools consist of and the order in which the steps must be completed.

Prior to creating a workspace, it is the responsibility of the Medical Team to provide all necessary information at the beginning of the process as the DS/CS/AI team will fill out a long questionnaire describing the purposes of the chosen study. Once that is complete, the DS/CS/AI team describes the purposes of the chosen study and submits the information in order to create a workspace. There is a disclaimer stating that answers to certain questions will be displayed publicly which is done to notify the *All of Us* research participants of the intended research purposes.

Once you select the option to create a new workspace, you are redirected to the following:

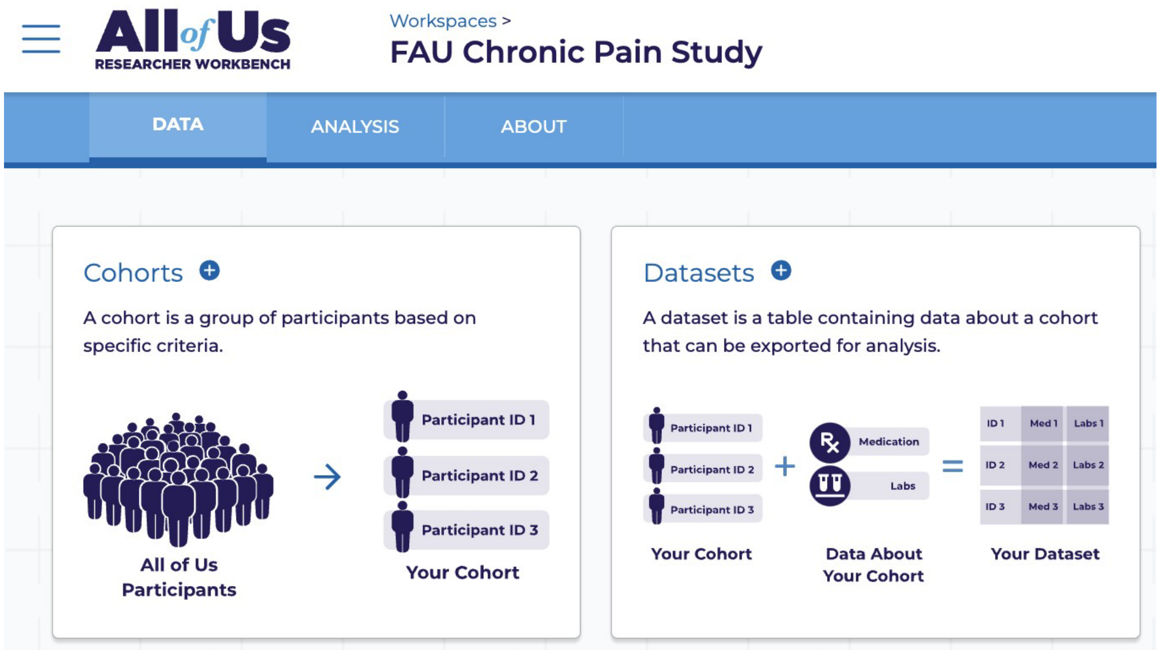


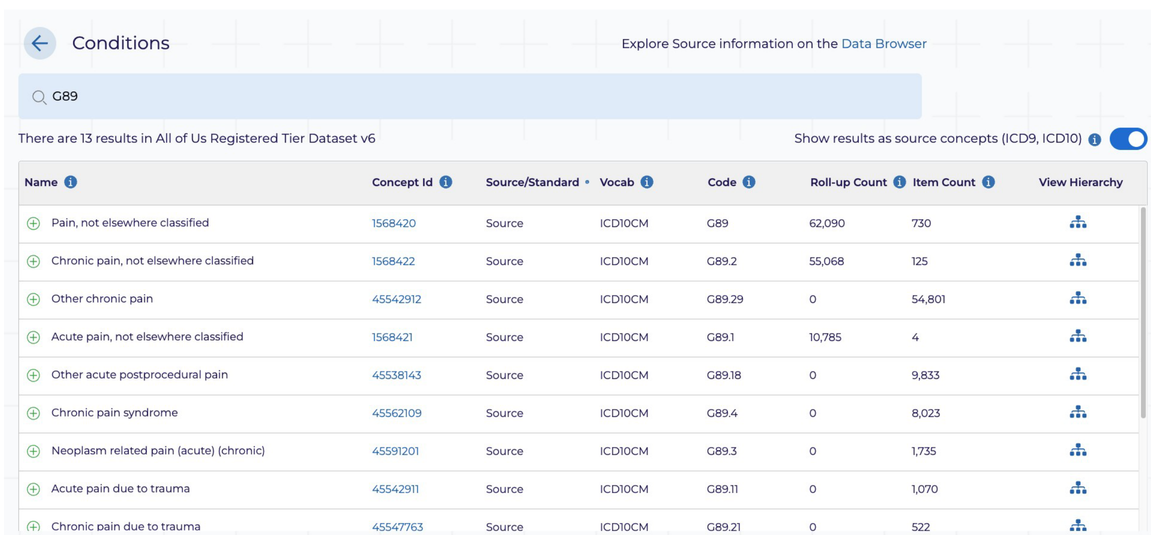
Figure 17 - Screenshot of the Data webpage (Cohort and Dataset creation) in the All of Us Research Workbench.

This is where the Cohort is created. The Medical Team should provide information regarding the inclusion and exclusion criteria, stating how they would like the data to be grouped. The inclusion and exclusion criteria come from multiple different domains, survey questions, and physical measurement categories.

Once the information has been provided and clarified, the DS/CS/AI team can now create a cohort. Cohort selection is a critical step. Understanding the way by which inclusion and exclusion criteria work can save both teams time and avoid complications. It might be too costly to go back to the cohort creation phase, after having invested a substantial amount of time and effort in the subsequent steps. The level of granularity is important. If we start with a very broad concept, we must be prepared to use all relevant exclusion criteria which will lead us to the desired cohort. Starting at a more granular level may save us from accidentally including unwanted data points but with the risk of missing valid data points along the way. The *All of Us* site/portal allows for multiple inclusion and exclusion criteria to be used in the same workspace. Deciding how many criteria and how to group them might help (for example, simulating suppression of selected exclusion criteria).

Let us illustrate these points with a concrete example: **Chronic Pain**

We searched for the condition using the provided code (G89) and got the following results:


Conditions

Explore Source information on the Data Browser

G89

There are 13 results in All of Us Registered Tier Dataset v6

Show results as source concepts (ICD9, ICD10)

Name	Concept Id	Source/Standard	Vocab	Code	Roll-up Count	Item Count	View Hierarchy
Pain, not elsewhere classified	1568420	Source	ICD10CM	G89	62,090	730	
Chronic pain, not elsewhere classified	1568422	Source	ICD10CM	G89.2	55,068	125	
Other chronic pain	45542912	Source	ICD10CM	G89.29	0	54,801	
Acute pain, not elsewhere classified	1568421	Source	ICD10CM	G89.1	10,785	4	
Other acute postprocedural pain	45538143	Source	ICD10CM	G89.18	0	9,833	
Chronic pain syndrome	45562109	Source	ICD10CM	G89.4	0	8,023	
Neoplasm related pain (acute) (chronic)	45591201	Source	ICD10CM	G89.3	0	1,735	
Acute pain due to trauma	45542911	Source	ICD10CM	G89.11	0	1,070	
Chronic pain due to trauma	45547763	Source	ICD10CM	G89.21	0	522	

Figure 18 - Screenshot of the output of conditions generated by searching "G89".

We then went back to the results for G89.4 (demonstrated by Figure 19) and clicked on the "tree" looking icon to view the hierarchy. The 'Finish & Review' button transitions you to the following list:

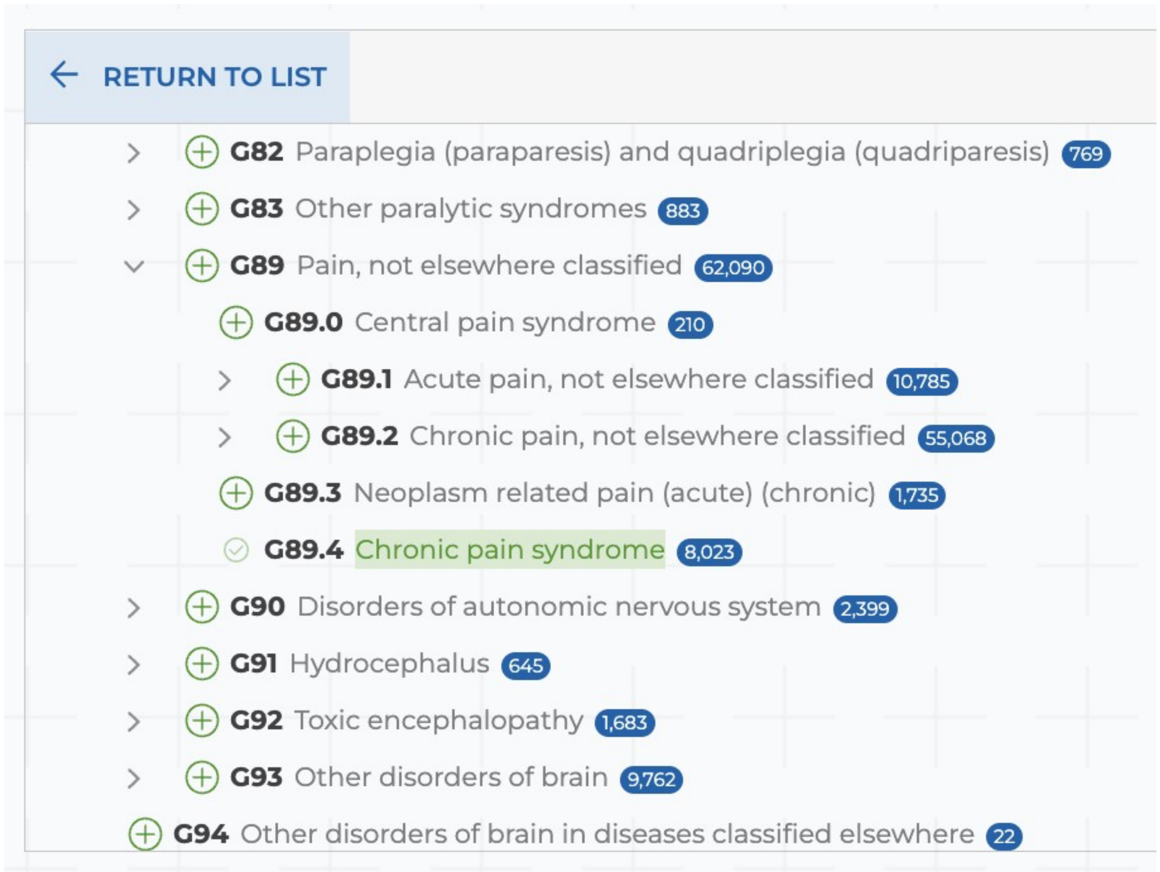


Figure 19 - Screenshot of the hierarchy of medical conditions when looking at code "G89.4".

There are multiple conditions under the G89 code which is another reason why the Medical Team must provide very specific information. This enables the DS/CS/AI team to know exactly which conditions to select as this process is crucial in producing the most accurate results possible. G89 was provided to broaden our search. As demonstrated by Figure 19, G89.0 only relates to **Central pain syndrome** whereas G89 is **Pain, not elsewhere classified**, so it is a much more general condition with a large pool of participants.

This is where the exclusion portion takes place. As you begin the process, you get to see how the number of cases varies. The right-hand side of Figure 20 includes an automatically generated visual representation of the data output which is very useful and creative. It indicates the participants that have the specific medical conditions selected after the inclusion/exclusion criteria has been implemented.

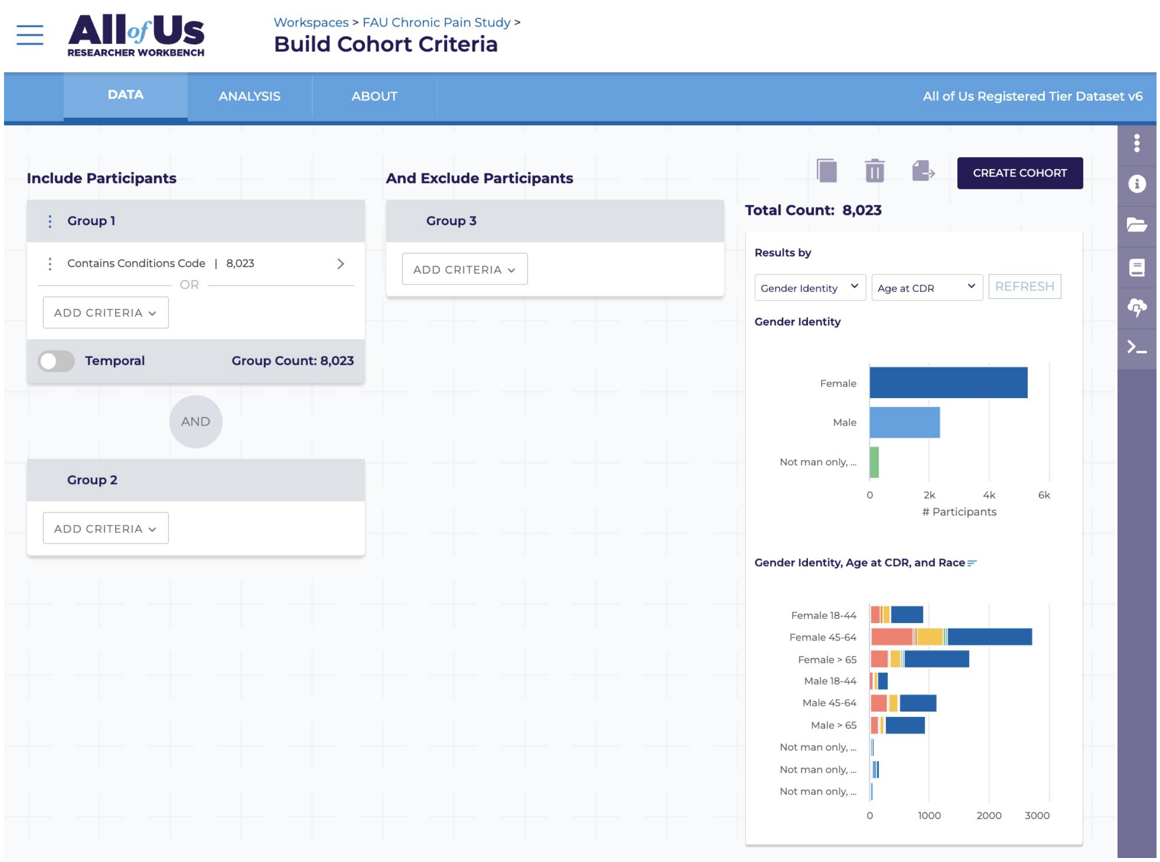


Figure 20 - Screenshot of the webpage conveying inclusion/exclusion criteria and total count of participants when building a Cohort.

When the DS/CS/AI team has completed this process, they create the cohort.

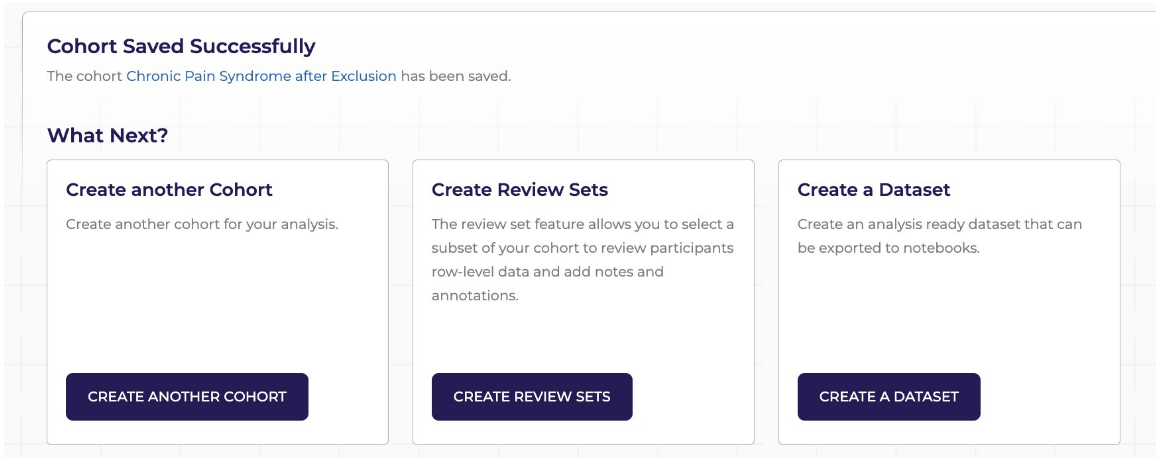


Figure 21 - Screenshot confirming that the Cohort created has been successfully saved and the ability to choose how to proceed.

This is followed by dataset creation. The Medical Team is responsible for providing information relevant to the dataset creation in order to guarantee the upmost accuracy of results.

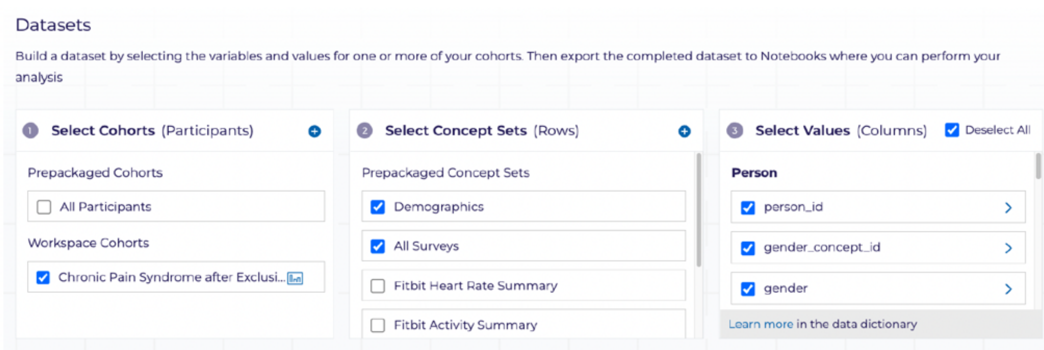


Figure 22 - Screenshot of the Dataset creation stage where researchers can select cohorts, concept sets, and values and preview the visualization of their data table based on that specific selection.

One key piece of information is the fundamentally important difference between concepts used for Cohort creation vs concepts used for Dataset creation. At cohort creation time, the interest is solely on the population whereas the dataset looks at concepts that will potentially answer the research questions.

Once all information has been provided and selected within the dataset creation process, the DS/CS/AI team selects variables/values within the dataset and can now create a notebook. After the dataset is exported, researchers can edit and create an environment that allows the notebook to run. The DS/CS/AI team can write code and retrieve the data of interest in order to answer the research questions/objectives. Notebooks allow the DS/CS/AI team with R or Python experience to perform high-powered analysis within the *All of Us* datasets using the integrated, cloud-based Jupyter Notebook environment. A SQL query preview is generated when exporting the dataset as shown in Figure 23.

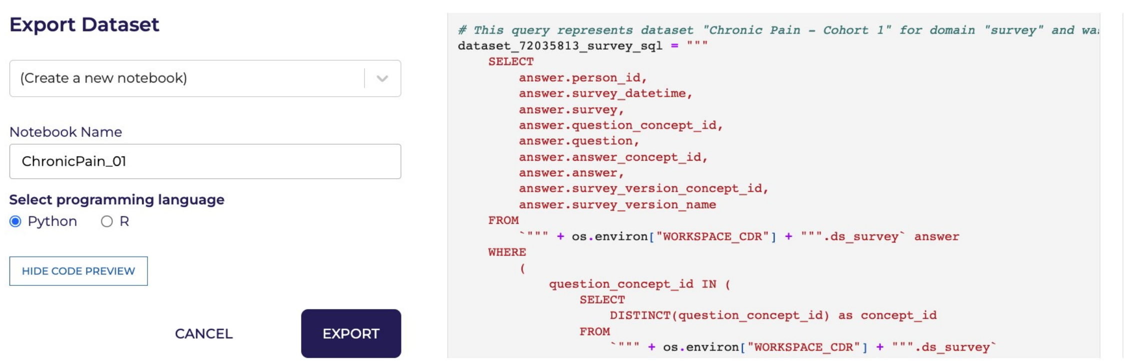


Figure 23 - Screenshot representing a preview of the SQL query prior to exporting the dataset.

5.1.4 ADOPTION OF THE PROTOCOL

To this date, the protocol was adopted for four projects involving three different medical teams. This effort began in the Fall 2022 semester during the Artificial Intelligence in Medicine and Healthcare class taught by Dr. Marques.

Project Number	Topic	Medical Team Lead	Affiliation
1	FAU Chronic Pain Study	Dr. Janet Robishaw (PhD)	Florida Atlantic University
2	Tuberculosis risk factors	Dr. Megan Coffee (MD/PhD)	New York University Grossman School of Medicine
3	Lab-value cutoffs associated with increased risk of intracranial hemorrhage	Dr. Tessa Harland (MD)	Albany Medical College
4	Statistical modeling of prevalence and social determinants of chronic pain	Dr. Tessa Harland (MD)	Albany Medical College

Figure 24 - A breakdown of each project, the topic, and the corresponding medical team lead as well as their affiliation. This conveys how the protocol was adopted throughout the thesis.

- Project 1 – FAU Chronic Pain Study was used as the pilot project which was extremely helpful in enabling us to test the protocol from start to end.
- Project 4 – Statistical modeling of prevalence and social determinants of chronic pain was selected as the main project as it appeared to be the most suitable for our research interests and available resources.

5.2 EXAMPLE: STATISTICAL MODELING OF PREVALENCE AND SOCIAL DETERMINANTS OF CHRONIC PAIN

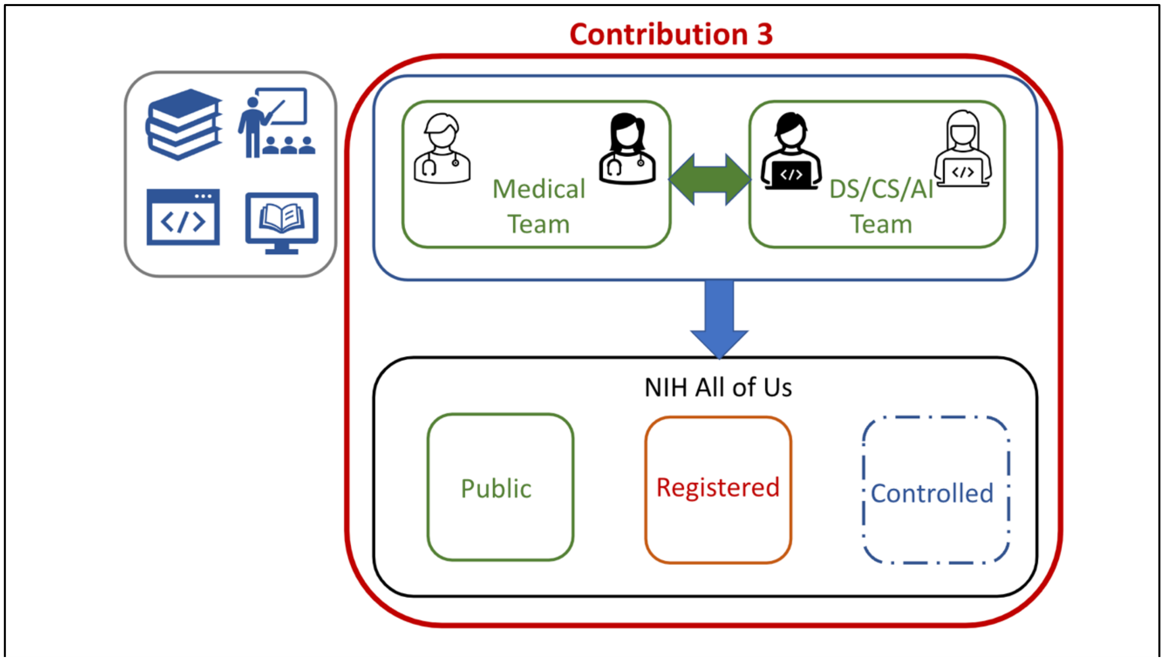


Figure 25 - An overview of the vital and continuous interaction between the Medical Team and the DS/CS/AI Team, who work together and use the NIH All of Us datasets to answer the Medical Team's research questions. Contribution #3 is a pilot project involving transdisciplinary teams using All of Us datasets.

This section discusses a specific ongoing project in which our data science team has teamed up with Dr. Tessa Harland, MD (Neurosurgeon, Albany Medical College) and shows how the framework described in Section 5.1 was successfully used to bring a project to completion.

5.2.1 RESEARCH QUESTION

The research question was formulated as:

- "What are the current demographic features and social determinants of health associated with chronic pain?"

When prompted to select what type of study is this, the medical team chose:

- "Disease-focused research."

5.2.2 ABSTRACT

Chronic pain affects more than 20.5% of US adults, causing significant impact on their daily life and incurring high economic costs [40]. Opioids have been widely used historically to manage chronic pain, contributing to the opioid epidemic. However, there has been a recent shift towards a more holistic biopsychosocial approach to chronic pain treatment [41], [42]. Despite this, chronic pain is still not well-understood, as it is influenced by many unknown factors. Demographic factors such as lower socioeconomic status and lower education level have been previously associated with chronic pain, but the reasons for this are unclear [43]. The National Institutes of Health's *All of Us* Research Program is a prospective data resource that aims to include a diverse range of individuals who have been underrepresented in medical research. This database will be used to provide a demographic overview of patients with chronic pain, with a focus on identifying social determinants of health that contribute to the disease. Multivariate analysis will be conducted to identify predictors of chronic pain that have not been previously described. The study is expected to uncover factors that are more prevalent in

chronic pain patients and can serve as potential targets for interventions that take a holistic approach to addressing chronic pain, potentially changing the way we treat this disease.

5.2.3 COMMUNITY IMPACT

Chronic pain affects more than 20.5% of adults in the US, which has a significant impact on their daily lives and results in high economic costs. The current treatment paradigm has been linked to the opioid epidemic [40], [42]. However, there has been a recent shift toward a more holistic approach to treatment, but there is little empirical evidence to guide this management modality. The *All of Us* database is well-positioned to provide a deeper understanding of social determinants that contribute to chronic pain and may predict it more accurately than any previous dataset. This database includes groups that have historically been underrepresented in research, allowing for the identification of new factors that have not been previously examined. This knowledge could guide the optimization of biopsychosocial treatment for chronic pain, which would have a significant impact on the communities it affects.

5.2.4 PROJECT DESCRIPTION

The project/workspace description can be found in the NIH Research Projects Directory (https://rb.gy/xxul3i), searching for “Statistical modeling of prevalence and social determinants of chronic pain”.

This project utilizes two datasets:

1. A dataset consisting of all chronic pain patients.

2. A dataset consisting of all patients who completed the social determinants of health survey, sorted into non-chronic pain and chronic pain cohorts.

Dataset 1 will provide a demographic cross-sectional overview of patients diagnosed with chronic pain. This will help characterize the disease's prevalence and provide information on groups that are typically underrepresented in medical research.

Dataset 2 will allow for a deeper analysis of the social determinants of chronic pain. By comparing the responses of those with chronic pain to those without, we will determine if certain factors are more prevalent in the chronic pain group. Multivariate analysis will be used to identify predictors of chronic pain that can be gleaned from the survey data.

5.2.5 EVALUATION OF THE TRANSDISCIPLINARY TEAM WORKING PROTOCOL

The protocol works! Steps 1-2 revolved around the creation of a workspace. All information is publicly available on the NIH site, in particular, when utilizing the project directory and looking up "Statistical modeling of prevalence and social determinants of chronic pain". All of the published information is technically accurate. Furthermore, Steps 3-4 allowed us to create cohorts following medical supervision. Cohort 1 has approximately 11,000 records (interest group) and Cohort 2 has approximately 45,000 records (control group). In addition, Steps 5-6 allowed us to create datasets based on Demographics and Survey questions as established by the medical team and their research objectives and Steps 7-8 revolving around the Notebook creation are still in progress.

5.2.6 PARTIAL RESULTS

At the time of this writing, the data science team[1] has produced Python code for exploratory data analysis (EDA) which includes useful plots such as the ones in Figures 26 and 27. Figure 26 shows the age distribution for both control and interest groups. It conveys the age distribution by gender which seems identical, and they are all left-skewed. The mean value of the control group is slightly lower than the interest group (around 4-5 units).

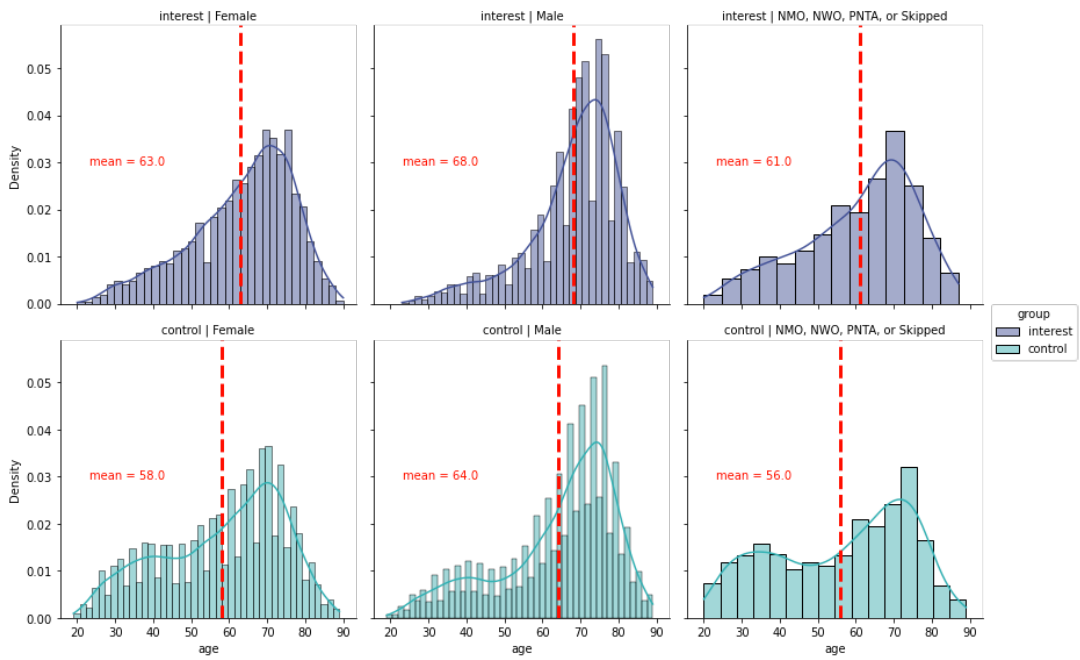


Figure 26 - Age distribution for both control and interest groups.

Figure 27 contains a box plot that compares age versus ethnicity for both control and interest groups. For the interest group, the median age is slightly higher than the control group, with a variation of 1-4 units, except for the Hispanic or Latino population and the Prefer Not To Answer group, which have a difference of 10 units. The

---

[1] Richard Acs, Nha Tran, and Christian Garbin

interquartile interval for the interest group, where 50% of the data is located, ranges from approximately 50 to 75 years old, while for the control group it ranges from 35 to 74 years old. The ethnicity population of the interest group appears to have a smaller range compared to the control group, and the minimum age for the interest group is 25 and above, whereas for the control group it is 20 and above. The control group has a longer box, indicating more dispersed data, and its distribution is slightly positively skewed or approximately symmetric. Additionally, some outliers appear at the bottom (age less than 30) for the Skip and Not Hispanic or Latino population in the interest group. We will take a closer look at these outliers in the next section to determine if they are significant enough to be considered.

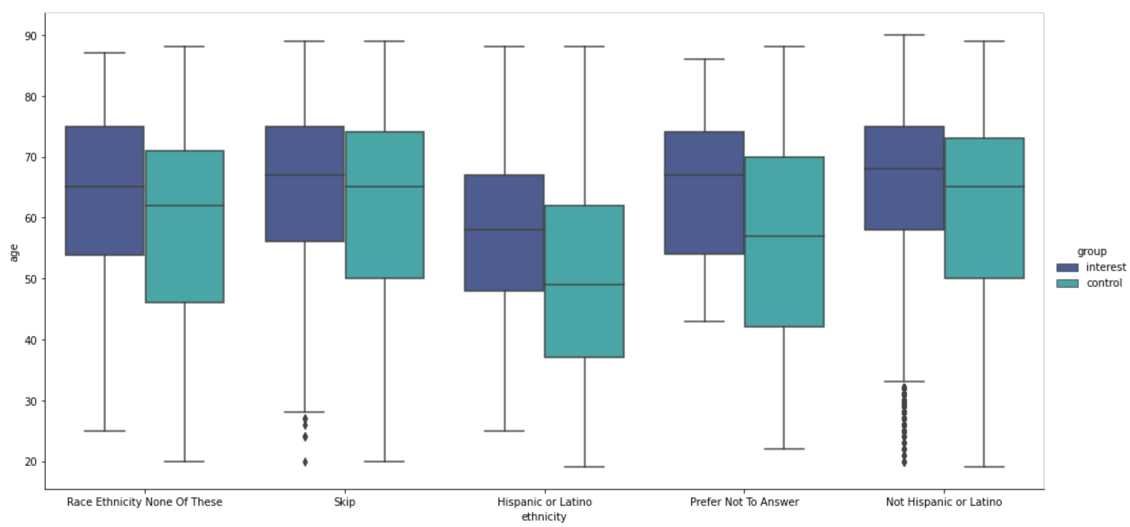


Figure 27 - Box plot of age versus ethnicity for both control and interest groups.

Work in progress includes writing Python code to build and test a predictive model that can identify the most promising social determinants of pain in a retrospective fashion.

CHAPTER 6

LESSONS LEARNED AND CONCLUDING REMARKS

6.1 LESSONS LEARNED

The protocol is not flawless! Due to unforeseen circumstances or busy schedules, the Medical Team might take a long time to return the requested information. This can potentially slow down the overall process. As the process advances, steps become more likely to require frequent fine-tuning and adjustments. For example, in step 3 (preparation for Cohort creation), not even the most competent medical team will be able to anticipate the aspects and steps involved in creating cohorts and selecting associated parameters. Moreover, the lack of clarity can lead to redundant concept selection. The DS/CS/AI team could incorrectly interpret the inclusion/exclusion criteria which can affect the accuracy of results. In addition, there is a limitation of ICD-9/10 codes. The lack of knowledge from the DS/CS/AI team, in terms of medical terminology (ICD codes) could potentially impact the selection of data and the overall accuracy of results. For example, ICD code G89.0 only relates to **Central pain syndrome** whereas G89 is **Pain, not elsewhere classified**, so it is a much more general condition with a large pool of participants. Therefore, clear and effective communication between both teams is vital in terms of the accuracy of results and answers to research questions.

6.2 CONCLUDING REMARKS

There has been much debate on the question of whether artificial intelligence can outperform humans in various fields. Although there has been considerable advancement in AI technology, it remains challenging to conclude that AI surpasses medical professionals in all aspects. Having said that, it is true that AI has the ability to utilize big data and analyze it to detect patterns or trends. Healthcare professionals are still very much needed when it comes to gathering and providing insightful feedback from these algorithms. Artificial intelligence cannot dominate on its own without human involvement. This is because AI is still limited by the data it is fed and the algorithms it uses. In contrast, healthcare professionals are able to use their intuition, creativity, and empathy to solve complex problems which highlights their unique skill set. This thesis implied that currently, a physician with both medical and technical skills will be very valued and in high demand. It is important to recognize that every aspect has strengths and weaknesses. Healthcare professionals go through medical school to become the most knowledgeable and qualified physicians who can provide valuable care and have a positive impact on patients' lives. On the other hand, the digital transformation of the healthcare industry can potentially revolutionize the way care is delivered and experienced. Medical treatments become more efficient and effective and with continued research, innovation, and collaboration between healthcare providers, technological companies, and policymakers, the digital transformation of healthcare using artificial intelligence can ultimately flourish and bring about many exciting opportunities.

APPENDICES

APPENDIX A

ALL OF US RESEARCH PROGRAM

YOUTUBE PAGE

The following table includes all videos on the *All of Us* Researcher Workbench YouTube page as of May 2022. These videos are both useful and informative, covering a wide range of topics that can be especially helpful for beginners or anyone trying to navigate the platform. While some tutorials delve into more technical aspects of the program, guest speakers are also included to share their knowledge and experience on how the *All of Us* Research Hub has helped enhance their research goals and findings. The videos are organized in chronological order. The table allows researchers to have easy access to videos that cover specific queries regarding the All of Us research program. This table was used by the data science team who managed to look up key words and find the corresponding tutorial in the most efficient way possible.

YouTube ID: to access the video, append the ID at the end of the URL (https://youtu.be/)

Video title:	**YouTube ID:**	**Tutorial covers:**
How to use Google Buckets (In Notebooks using R)	_eQ7ItlUJMs	• How to save files in buckets. • How to list files in buckets. • How to retrieve files from a bucket into notebooks.
All of Us onboarding tutorial video	Y92Fa5L4SC0	• Step-by-step registration process. • Requirements to become a member.

Video title:	YouTube ID:	Tutorial covers:
Office Hours (07/17/2020) Information about All of Us Workbench	2e9WBgvmn4E	• In-depth review of the Workbench and how it can be used for research purposes.
Office Hours (07/31/2020) Building a Cohort in the Workbench	hIut3DUs60o	• An in-depth explanation on how to build a cohort. • Step-by-step tutorial. • Why cohorts are beneficial.
How to use Google Buckets using Python	NAiA9oG770Q	• How to save files in buckets. • How to list files in buckets. • How to retrieve files from a bucket into notebooks.
Office Hours (09/04/2020) Workspace Management Tips and Tricks	7YWlAWMYudA	• Sharing workspaces. • Duplicating workspaces. • Deleting workspaces. • Saving and backing up your work.
Office Hours (10/02/2020) All of Us Survey Development and Data	INxPex5bnRQ	• Survey development. • Resources for exploring survey data in the Registered Tier. • Future plans.
Office Hours (10/16/2020) OMOP, All of Us Data Structure and Standardization	jK11qAus8Q8	• *All of Us* use of OMOP for CDR. • OMOP CDM Principles. • OMOP Vocabulary Model. • How it links to *All of Us*.
Office Hours (10/30/2020) User Questions	LKc9yvmIvwQ	• Concept set selection. • Workbench updates. • Drug selection. • Feedback about support materials.
Office Hours (11/13/2020) Data Statistics Dissemination Policy	dU4zfirC1P0	• How the dissemination policy protects participant privacy. • How researchers comply with this policy using the Workbench.
Introduction to Researcher Workbench	NTLJtwLcavo	• Updated tutorial for the Researcher Workbench.
User Support Hub	Ni4PEJVbmSk	• Tutorial video explaining how to make the most out of the User Support Hub.

Video title:	YouTube ID:	Tutorial covers:
Cohort Builder and Cohort Review	G6_GG2CJ9mA	• Tutorial video explaining how to use the Cohort Builder and get the most out of the cohort tool.
Accessing Researcher Workbench via login.gov	MehbfLKvGcs	• Researcher authentication service. • Step-by-step tutorial on how to enroll using login.gov.
Office Hours (12/18/2020) Open Q&A Session	R4nClIM2UqI	• Questions about: Survey information, suggestions for surveys, creating cohorts based on seg/age/ethnicity and control cohorts, basic information about the Workbench, sharing workspaces, new curated data repository, wearable data, connecting with other researchers on the Workbench.
Office Hours (02/05/2021) Popular Questions from the Help Desk	JefXwJ6iMTQ	• Billing questions. • Information about cloud environments. • Costs of tutorial workspaces. • New versions of the curated data repository (CDR). • Data and dissemination policy. • Information about new user orientations.
Dataset Builder and Concept Sets	cUuDKUxjQoo	• Tutorial on how to use Dataset Builder and Concept Sets.
Notebooks and Code Snippets	NvMWBlVyyUU	• Examples of notebooks and code snippets in the *All of Us* research hub. • How to use them.
Office Hours (08/14/2020) Using the Concept Set Selector in the Workbench	vrRbfC1p4qw	• A demonstration about using the Workbench. • Focused on using the Concept Set Selector Tool.
Office Hours (12/04/2020) Erwin curated data repository (CDR) release	NKZMae-ztkg	• Information about the upcoming release of the CDR. • Includes Fitbit and COPE survey.

Video title:	**YouTube ID:**	**Tutorial covers:**
The All of Us Data Browser Tutorial Video	VnmMGKffvcQ	• Introduction to the Data Browser. • Electronic Health Records. • Surveys. • Physical Measurements and Fitbit. • Applied example. • Asking for help.
Office Hours (03/05/2021) Wearable Device Dataset	6TmVyNHNWQ8	• Wearable device data on Workbench. • Higher level view about what might be possible using wearable data from Vik Kheterpal.
Office Hours (04/16/2021) Popular Topics from Help Desk	5KIf9Rd2Y2w	• Information about featured tutorial workspaces in the Workbench. • Accessing the publication reporting form. • Downloading Plotly images. • Requesting more credits.
Office Hours (04/30/2021) Working with COPE Survey Data	aoGd32DWJ-A	• Information about using COVID-19 Participant Experience (COPE) survey data in the Workbench.
Office Hours (05/28/2021) Optimizing Workflow in Workbench	cDMi2gDjJis	• Tips about cohort selection. • Querying data. • Computational environment. • Saving data. • Snippets. • Creating HTML snapchats. • Some user questions.
Office Hours (06/11/2021) All of Us Data Tiers 101	7c96LRhRqHU	• Information about the different Tiers (Public and Registered Tiers). • How participant privacy is protected. • What information is available and suppressed in these Tiers.
Office Hours (06/25/2021) Implementing Medications in All of Us	Ik4PZiQN8_s	• Lina Sulieman from VUMC walks through medication organization in the *All of Us* database and how to extract this data in the Workbench.
Office Hours (07/23/2021) Tips for Wrangling All of Us Data	COrwr7JeP8k	• Tips for analyzing data in the Workbench. • Merging data from different data types.

Video title:	**YouTube ID:**	**Tutorial covers:**
Office Hours (08/20/2021) Registered Tier Refresh	jPc5eR63vgY	• Information about new available data in the Registered Tier. • Update released in September.
Office Hours (10/15/2021) Using R in the Workbench	qvT4YB97LOw	• The Research Support Team provides information and tips about using R in the Researcher Workbench. • Answers to user questions.
Office Hours (11/12/2021) Using All of Us Data for Cardiovascular Research	-e2bad06ZyA	• Julian Acosta and Audrey Leasure from Yale School of Medicine discuss their work utilizing the *All of Us* dataset to analyze the epidemiology of cardiovascular diseases.
Office Hours (12/17/2021) New User Orientation	gVOE-GamKXk	• Information about the *All of Us* Research Program. • How participant data is curated and organized. • Overview of how to use the Researcher Workbench. • Information about the User Support Hub.
Office Hours (03/18/2022) Controlled Tier Introduction	W3QD481VPJs	• Overview of the release of the Controlled Tier dataset on *All of Us* Researcher Workbench containing genomic data. • Includes information about the dataset and how to access it. • Support materials available for users.
Selecting Variants in the All of Us Workbench	N8w3XXFUycE	• Manipulate Hail MatrixTable. • How to select specific variants from the Hail MatrixTable. • Getting started with genomic data.
Office hours (04/15/2022) Accelerating Cancer Research using All of Us data as a supplementary resource	aBLiHw75I8Y	• Dr. Jay Ronquillo from the National Cancer Institute discusses how the *All of Us* Research Program is being used in conjunction with the NCI Cancer Research Data Commons to accelerate cancer research.

Video title:	**YouTube ID:**	**Tutorial covers:**
Office Hours (04/29/2022) Writing a Meaningful Workspace Description	3ij-wBizh5A	• The NIH policy office details the importance of writing a meaningful workspace description.
Office Hours (05/06/2022) New Researcher Orientation	StJYQiJnXfA	• The *All of Us* Research Support team walks through how to use the Workbench. • Updated video that now includes Controlled Tier.

BIBLIOGRAPHY

[1] "How health information systems are driving digital transformation in healthcare," *World Economic Forum*, Sep. 09, 2022. https://www.weforum.org/agenda/2022/09/health-information-system-digital-transformation-healthcare/ (accessed Mar. 15, 2023).

[2] "Digital Transformation in Healthcare: Trends, Challenges & Solutions," *Peerbits*. https://www.peerbits.com/blog/digital-transformation-in-healthcare.html (accessed Mar. 15, 2023).

[3] L. Veljanovski, "Apple Watch Saves Man's Life After He Falls From Electric Bike," *Newsweek*, Feb. 02, 2022. https://www.newsweek.com/apple-watch-saves-mans-life-falls-electric-bike-california-tech-1675335 (accessed Apr. 06, 2023).

[4] "Patient portal is primary engagement tool used by health systems," *The Center for Connected Medicine*, Nov. 20, 2019. https://connectedmed.com/resources/patient-portal-the-primary-engagement-tool-employed-by-health-systems/ (accessed Mar. 15, 2023).

[5] "Call to Action to the Tech Community on New Machine Readable COVID-19 Dataset – The White House." https://trumpwhitehouse.archives.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/ (accessed Mar. 15, 2023).

[6] National Institute of Allergy and Infectious Diseases (NIAID), "A Phase I Study to Assess the Safety, Reactogenicity and Immunogenicity of Two Quadrivalent Seasonal Influenza Vaccines (Fluzone(R) or Flublok(R)) With or Without One of Two Adjuvants (AF03 or Advax-CpG55.2) in Healthy Adults 18-45 Years of Age," clinicaltrials.gov, Clinical trial registration NCT03945825, Feb. 2023. Accessed: Mar. 14, 2023. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03945825

[7] A. Park, "AI develops 'turbocharged' flu vaccine," Jul. 02, 2019. https://www.beckershospitalreview.com/innovation/ai-develops-turbocharged-flu-vaccine.html (accessed Mar. 15, 2023).

[8] B. Meskó, G. Hetényi, and Z. Győrffy, “Will artificial intelligence solve the human resource crisis in healthcare?,” *BMC Health Services Research*, vol. 18, no. 1, p. 545, Jul. 2018, doi: 10.1186/s12913-018-3359-4.

[9] B. Meskó and M. Görög, “A short guide for medical professionals in the era of artificial intelligence,” *npj Digit. Med.*, vol. 3, no. 1, p. 126, Sep. 2020, doi: 10.1038/s41746-020-00333-z.

[10] G. Briganti and O. Le Moine, “Artificial Intelligence in Medicine: Today and Tomorrow,” *Frontiers in Medicine*, vol. 7, 2020, Accessed: Mar. 15, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmed.2020.00027

[11] “AIMI Invited Lecture: From the Moon and Formula One to Beethoven’s Ninth Symphony: Musings of a Clinician-Data Scientist - Anthony Chang, MD, MBA, MPH, MS | Center for Artificial Intelligence in Medicine & Imaging,” Mar. 15, 2023. https://aimi.stanford.edu/events/aimi-invited-lecture-moon-and-formula-one-beethovens-ninth-symphony-musings-clinician-data (accessed Mar. 15, 2023).

[12] V. B. Kolachalama and P. S. Garg, “Machine learning and medical education,” *npj Digital Med*, vol. 1, no. 1, Art. no. 1, Sep. 2018, doi: 10.1038/s41746-018-0061-1.

[13] D. Li, K. Kulasegaram, and B. D. Hodges, “Why We Needn’t Fear the Machines: Opportunities for Medicine in a Machine Learning World,” *Academic Medicine*, vol. 94, no. 5, p. 623, May 2019, doi: 10.1097/ACM.0000000000002661.

[14] S. A. Wartman and C. D. Combs, “Reimagining Medical Education in the Age of AI,” *AMA Journal of Ethics*, vol. 21, no. 2, pp. 146–152, Feb. 2019, doi: 10.1001/amajethics.2019.146.

[15] C. Guan, J. Mou, and Z. Jiang, “Artificial intelligence innovation in education: A twenty-year data-driven historical analysis,” *International Journal of Innovation Studies*, vol. 4, no. 4, pp. 134–147, Dec. 2020, doi: 10.1016/j.ijis.2020.09.001.

[16] B. Gallix and J. Chong, “Artificial intelligence in radiology: who’s afraid of the big bad wolf?,” *Eur Radiol*, vol. 29, no. 4, pp. 1637–1639, Apr. 2019, doi: 10.1007/s00330-018-5995-9.

[17] J. Grunhut, A. T. Wyatt, and O. Marques, “Educating Future Physicians in Artificial Intelligence (AI): An Integrative Review and Proposed Changes,” *J Med Educ Curric Dev*, vol. 8, p. 23821205211036836, Sep. 2021, doi: 10.1177/23821205211036836.

[18] J. Grunhut, O. Marques, and A. T. M. Wyatt, “Needs, Challenges, and Applications of Artificial Intelligence in Medical Education Curriculum,” *JMIR Med Educ*, vol. 8, no. 2, p. e35587, Jun. 2022, doi: 10.2196/35587.

[19] E. J. Topol, *Deep medicine: how artificial intelligence can make healthcare human again*, First edition. New York: Basic Books, 2019.

[20] D. B. Meskó, "A GUIDE TO ARTIFICIAL INTELLIGENCE I N H E A LT H C A R E".

[21] A. Bohr and K. Memarzadeh, Eds., *Artificial intelligence in healthcare*. London ; San Diego, CA: Academic Press, imprint of Elsevier, 2020.

[22] M. Matheny, S. Thadaney Israni, M. Ahmed, and D. Whicher, Eds., *Artificial intelligence in health care: the hope, the hype, the promise, the peril*. Washington: National Academy of Medicine, 2020.

[23] K. L. Holley and S. Becker, *AI-first healthcare: AI applications in the business and clinical management of health*, First edition. Sebastopol, CA: O'Reilly Media, 2021.

[24] Anshik, *AI for healthcare with Keras and Tensorflow 2.0: design, develop, and deploy machine learning models using healthcare data*. Berkeley: Apress, 2021.

[25] R. Shimonski, *Ai In Healthcare:
*. Indianapolis: John Wiley and Sons, 2021.

[26] A. Nguyen, *Hands-On Healthcare Data: taming the complexity of real-world data$nAndrew Nguyen*, First edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly, 2022.

[27] "All of Us Research Program | National Institutes of Health (NIH)," *All of Us Research Program | NIH*, Jun. 01, 2020. https://allofus.nih.gov/future-health-begins-all-us (accessed Mar. 15, 2023).

[28] "All of Us Research Hub." https://www.researchallofus.org/ (accessed Mar. 15, 2023).

[29] "Data Snapshots – All of Us Research Hub." https://www.researchallofus.org/data-tools/data-snapshots/ (accessed Mar. 15, 2023).

[30] "The Basics Survey – All of Us Research Hub." https://www.researchallofus.org/data-tools/survey-explorer/the-basics-survey/ (accessed Mar. 15, 2023).

[31] "Data Access Tiers – All of Us Research Hub." https://www.researchallofus.org/data-tools/data-access/ (accessed Mar. 15, 2023).

[32] "Biobank," *All of Us Research Program | NIH*, Jun. 24, 2020. https://allofus.nih.gov/funding-and-program-partners/biobank (accessed Mar. 15, 2023).

[33] “Genomics Partners,” *All of Us Research Program | NIH*, Jun. 24, 2020. https://allofus.nih.gov/funding-and-program-partners/genomics-partners (accessed Mar. 15, 2023).

[34] “Survey Explorer – All of Us Research Hub.” https://www.researchallofus.org/data-tools/survey-explorer/ (accessed Mar. 15, 2023).

[35] “Health Care Provider Organizations,” *All of Us Research Program | NIH*, Sep. 10, 2020. https://allofus.nih.gov/funding-and-program-partners/health-care-provider-organizations (accessed Mar. 15, 2023).

[36] “Data Methods – All of Us Research Hub.” https://www.researchallofus.org/data-tools/methods/ (accessed Mar. 15, 2023).

[37] “Exploring concepts and concept relationships within Observational Medical Outcomes Partnership (OMOP),” *User Support*. https://support.researchallofus.org/hc/en-us/articles/360039585491-Exploring-concepts-and-concept-relationships-within-Observational-Medical-Outcomes-Partnership-OMOP- (accessed Mar. 15, 2023).

[38] “Researcher Workbench – All of Us Research Hub.” https://www.researchallofus.org/data-tools/workbench/ (accessed Mar. 15, 2023).

[39] “Research Projects Directory – All of Us Research Hub.” https://www.researchallofus.org/research-projects-directory/ (accessed Mar. 15, 2023).

[40] R. J. Yong, P. M. Mullins, and N. Bhattacharyya, “Prevalence of chronic pain among adults in the United States,” *Pain*, vol. 163, no. 2, pp. e328–e332, Feb. 2022, doi: 10.1097/j.pain.0000000000002291.

[41] N. Hylands-White, R. V. Duarte, and J. H. Raphael, “An overview of treatment approaches for chronic pain management,” *Rheumatol Int*, vol. 37, no. 1, pp. 29–42, Jan. 2017, doi: 10.1007/s00296-016-3481-8.

[42] B. Marshall, M. K. Bland, R. Hulla, and R. J. Gatchel, “Considerations in addressing the opioid epidemic and chronic pain within the USA,” *Pain Manag*, vol. 9, no. 2, pp. 131–138, Mar. 2019, doi: 10.2217/pmt-2018-0070.

[43] A. Shmagel, R. Foley, and H. Ibrahim, “Epidemiology of Chronic Low Back Pain in US Adults: Data From the 2009-2010 National Health and Nutrition Examination Survey,” *Arthritis Care Res (Hoboken)*, vol. 68, no. 11, pp. 1688–1694, Nov. 2016, doi: 10.1002/acr.22890.