MCIDT Innovations in Practice: Introduction and Application of AI in Infectious Disease

Podcast Transcript

Geoffrey H. Siwo, PhD

November 7, 2024

PUBLIC HEALTH PREPARED
A Program of the Michigan Center for Infectious Disease Threats & Pandemic Preparedness

Public Health Prepared
University of Michigan
School of Public Health
1415 Washington Heights
Ann Arbor, MI 48109-2029

PHPrepared@umich.edu
734-615-0144
PublicHealthPrepared.org

0:00:02.0 Storee Harris: Welcome to Prepared Set Go. A podcast of public health Prepared Public health Prepared is a workforce development branch of the Michigan Center for Infectious Disease Threats and Pandemic Preparedness, or MI Initiative, which is housed that and funded by the University of Michigan. We hope this podcast will better equip the public health workforce to handle ongoing and future health crises. Thank you for tuning into our episode. Today, we'll be talking about AI infectious disease and practice. I'm your host Storee Harris-Stubblefield. Today we have Dr. Geoffrey Siwo. Dr. Siwo is a research assistant professor at University of Michigan Medical School Department of Learning Health Sciences. A research associate at the Michigan Center for Global Health Equity.

0:01:00.9 SH: His research interests are in the development of technologies for accelerating equitable advancements in drug discovery, clinical medicine, and global health. He is a founder of Anza Bio Technologies and has served as co-founder and advisor to several other biotechnology startups, including two with clinical stage products. Previously, he was a research assistant professor at the University of Notre Dame, a genomic specialist at IBM TJ Watson Research Center and Lead researcher at IBM Research Africa Labs. He is the recipient of several awards, including TED Fellow and next Einstein Forum fellow. His work has also been featured in several media, including CNN USA Today Fast Company, and Ozzy among other media. Welcome to the podcast. I'm excited to speak with you today.

0:01:55.3 Geoffrey Siwo: Thank you Storee. It's a pleasure to be here today.

0:01:57.3 SH: Thank you. So getting started, Dr. Siwo, let's start by having you tell our audience a bit about your background.

0:02:06.3 GS: So I think, again, Storee, it's really a pleasure to be here today. I am an research assistant professor from the University of Michigan in the Department of Learning Health Sciences. My background is in biological sciences, so I have, a PhD in biological sciences from the University of Notre Dame. My work involves a lot of computational modeling across, various diseases. So I have worked on infectious disease like malaria, as well as viral diseases like SARS-CoV-2 and my research also involves developing computational tools for genetic, diseases.

0:02:48.5 SH: Thank you for providing that additional background. So I would like to ask you, what drew you to the intersections of AI and infectious disease research?

0:02:58.4 GS: So, I grew up in Kenya where malaria, HIV/AIDS and, other infectious diseases claim millions of lives. So by the time I graduated with a bachelor's degree, and this was in Kenya in 2005, I was really convinced that to tackle some of the biggest challenges in ending these infectious diseases, perhaps even in our lifetime, will require us to spend... Will require us to act, at a speed and a cost that is not feasible using conventional way of research. Conventionally biological research takes place in, laboratories, laboratories that have to be well equipped and laboratories that are not available in many parts, of the world. So when I was invited to join a research program on Malaria in 2005, that's when I finished my bachelor's degree. I already had an idea of how to accelerate the design of vaccine candidates using machine learning and AI. So in a sense, I started working on machine learning and AI way back in 2005 before this field grew into what it is today that is very popular.

0:04:13.9 SH: In terms of AI. Can you talk a little bit more about what it is and more specifically, what is generative AI?

0:04:20.0 GS: So lemme start by AI. So AI represents a group of technologies that enable computers to learn, make predictions, and solve a wide range of problems, typically using data. So when people speak of AI one branch of AI that is very common is called Machine Learning. And Machine Learning is a subfield of AI that involves building computer models that use data to do things like make predictions. So for health, making predictions like who is at risk for a given disease. Generative AI, on the other hand, is relatively new and involves AI technologies that are used to generate new data or new information. So Generative AI can be used to... Can be trained to generate images, we've seen this, with generative AI platforms that can generate social images and other kinds of images today. Yeah, Generative AI can also be, used to generate like sound and and videos. So we'll be seeing that, you can have completely AI generated a video that is very hard to differentiate from real videos, so generative AI basically generates new data that is typically hard to differentiate from, real data.

0:05:45.9 SH: Thank you so much for bringing up those points about creating sound and video. We'll touch on it a little bit later on in the podcast and thinking about some ethical considerations. But wanting to pivot a little bit more to the present. How are you using AI in your viral pandemic in global health research?

0:06:05.4 GS: Yeah, so in the are of pandemic research when SARS-CoV-2 hit the world and, there an added need for thinking of new ways to rapidly develop new molecules, we received funding from the Bill & Melinda Gates Foundation to support the development of generative AI approaches that can be used to quickly design potential antiviral, molecules. So we've, done proof of concept for that work, finding that AI models can, basically from scratch be trained to design new molecules that could potentially have, antiviral activity to viruses like SARS-CoV-2. We are transitioning that work, so we have a collaboration with the Walter Reed US Army Research Institute, where we are exploring how to leverage these Generative AI tools, we've developed and applied to other viruses like Dengue, which is an important public health threat. Another kind of work we are doing on AI in global health.

0:07:10.1 GS: We are trying to tackle two important issues. One is that, one of the big challenges in public health when it comes to AI is that there's lack of big data sets that are required to train AI systems. So we are trying to test to whether AI generated data can be used to supplement the lack of large public health or global health data sites. So the idea is that you can train an AI model how to generate data that can then be used to train other AI models to solve important public health questions. And as an example, we have collaboration with Aga khan University in Kenya, where we are using Generative AI to generate synthetic data of, a health and demographic surveillance system within the country with the idea that this data can then be used to train AI models and at the same time, so by using AI generated synthetic data, we can address some of the challenges of sharing public health data that might have private, or confidential information. So these data sets can be shared more broadly. Also, we are developing new approaches that could enable models like ChatGPT to provide information on public health that is grounded by scientific evidence because currently AI models like ChatGPT haven't been tested to see how well they can perform in answering public health related questions.

0:08:51.0 SH: Thank you so much for touching on the ChatGPT and its relevance to public. I think that's a common conversation being had in these spaces, so thank you so much for bringing that up. Moving a little bit closer to home, in what ways can AI be used at local health departments?

0:09:06.8 GS: Yeah, so that's a good question, thank you. Locally, the emergence of Generative AI tools like ChatGPT I think presents exciting opportunities for public health departments. And I should say, ChatGPT is is Generative AI system, it's a

large language models, so large language models of LLMS are generative AI systems that are trained on, large amounts of, text. So text can come from books, can come from web platforms like Wikipedia. So these LLMS have been trained on that kind of information. So the advantage of these, LLMs like ChatGPT from open AI as well as, so Google is Gemini, Llama is from Meta. The advantage of these LLMs is that you don't need to be a data scientist to use them. Traditionally AI models were only used by people who are computer science, engineers.

0:10:10.4 GS: And what these, LLMs are doing is that they are democratizing who can use or leverage AI models in their work. So for public health departments, LLMs can help in engagement with the public. So engagement can be in terms of, public health campaigns, making those campaigns, more accessible or even translating them into different, languages. So LLMs can be used to translate from English, to Spanish or Spanish to English, and that can make, information more accessible. Also, local public health departments can use LLMs to evaluate what public health initiatives are working and which are not. So public health departments, can do surveys of some of the initiatives they have and get responses from those surveys, these LLMs because they can analyze text, right? They can extract information from text, they can provide a rapid way, of taking surveys from the public and using those surveys to figure out what's working as well as what is not working.

0:11:24.1 GS: I think the other big opportunity I see for public health, department is that public health departments may not currently have the data science talent within them to take advantage of data science that have been collected over numerous years by these departments. And public health departments could leverage technologies like AI generated data or synthetic data to make this data more shareable. So, because public health departments may not want to share this data because of privacy issues, but if this data is used to generate synthetic data that preserve privacy, then actually public health departments could engage data scientists in academia, for instance, through collaborations to use this data to inform public health practices. I think this is a really huge, a huge opportunity because in this way, public health departments can invite external talent, to take advantage of this data.

0:12:25.3 GS: And, to give an example. So in 2019, I organized a global competition that basically took some data on Malaria and made it public and invited anyone around the world to come and use this data to build Machine Learning models that can predict Malaria drug resistance. And when we did this, what that did is that

instead of our research lab only working on that data, we opened this data and we had over 360 individuals across 31 countries coming to build computer models on this data. And from that, we got very good models that were better than what we could ourselves like build. So this is why we would encourage public health departments to look at ways of, sharing data broadly and incentivizing participation across the public in using these data sets.

0:13:17.8 SH: Thank you so much for providing that example of the collaborative nature of public health, but also using these technologies to be collaborative. So thank you so much, that was really interesting. You showed a lot of great opportunities for AI in respects to public health, but just curious, what are some challenges in infectious disease research that AI can help address?

0:13:44.1 GS: Yeah, so in infectious disease research, I think AI have a very big impact on forecasting the spread of infectious diseases. And you know we learned from, the COVID-19 pandemic, developing ways of anticipating and predicting the spread of viral diseases is very important because it can inform the safeguards we need to put in place in order to prevent other spread of this. AI models can also be very important in identifying populations at risk. So for many infectious diseases, the risk profile changes across different, demographics. So for some diseases, the elderly would be more vulnerable. For others, it might be the young, or it might be the immunocompromised and AI models can help us identify those at risks so that our public health interventions can be better informed and we can cost effective practices that target the populations that are most at risk. AI model are also going to be very important in discovering new drugs for infectious diseases, especially antibiotics.

0:14:47.2 SH: As you know, the world is in need of new antibiotics with, the rise of antimicrobial, resistance or AMR, and this is an area that AI is showing, a great promise. In fact, there's really like an antibiotic that was designed by AI that is currently being evaluated. And then of course, one of the most important tools in, public health are vaccines, and here AI models can help in terms of designing new vaccines, but also trying to target these vaccines more effectively in terms of figuring out the vaccine supply, ensuring that vaccines, are available at the right place at the right time is something that AI models would really play a big role in.

0:15:41.6 SH: Thank you for talking about those challenges. With those challenges in mind, what are some limitations of the current AI approaches?

0:15:47.4 GS: Yeah, thank you for that question. So one of the big challenges to AI which is also a foundation of AI, is big data, right? And the fact is that big data is not always available. And so this is really a big bottleneck for the success of AI in fields where there's not a large amount of data. The other problem is that because AI relies on data, which is from the past, right? It means that AI can be very sensitive to shifts in data. So for example, if you trained AI models to predict, infectious disease risk before Covid, and then Covid hit, which is a new disease having like many symptoms that, are similar to other respiratory diseases, right? And so your AI model, if you were trained before Covid, they wouldn't have like, anticipated this.

0:16:48.2 GS: And so the performance of those models will be affected because of this shift, a shift in, one infectious disease, which then can also have an impact on other diseases, right? Because we know COVID had a big impact, not just by affecting patients directly, but also indirectly because many patients missed their doctor's appointments, right? And as a result of that, we see that there has been a shift in early cancer detection, right? Because many patients didn't go for their checkups, which could have detected some of the cancers earlier. So this is a big problem with AI the fact that it relies on data. The other big problem is that even when AI does very well in giving you, good predictions, it's not always explainable why it is doing so. And you can say. Okay, if I'm trying to predict or focus the incidents of an infectious disease, right? And we predict that all the cases of RSV are going to like go up in the next few months, do I need an explanation?

0:17:51.6 GS: Yeah. And the answer is yes, you need an explanation in many cases, because explanations can help you do something about the predictions, right? If you make predictions without being able to explain why the cases of the disease are going up, then it's very hard to know what to do in a very targeted way. This is also a problem, especially with AI because AI is actually good at learning shortcuts so shortcuts meaning that AI can learn how to give you very good predictions by using something that may not make sense at all.

0:18:28.0 GS: So I'll give an example here of using X-ray images to diagnose cancer. But when use an AI model to try and use an X-ray image to differentiate between a cancer case and a healthy case, it has been shown that sometimes AI models, they will learn something about the image that has nothing to do about the cancer. So an example is that there might be some label on the X-ray image that is made by doctors when they see that, oh, this patient does... And that label is placed on the X-

ray image. And the AI model instead learns that label so that even if you take away the... You block the part of the X-ray image that has the cancer, the AI model will still give you the correct answer even though you blocked the image. And the reason is that the AI took a shortcut, it learns something else that has nothing to do actually what you're interested in.

0:19:29.7 GS: And then finally, a big problem, and this is specific to Generative AI and large language models, is the problem of hallucination. So, where AI models can give you an answer that sounds to make a lot of sense, but if you start digging deeper, you find that, oh, this answer actually doesn't make sense at all. A quick example is that LLMs like ChatGPT, people are using them to summarize like documents or even to get medical answers. So it's been shown that ChatGPT can do very well on standard medical examinations. But if you ask the LLM like ChatGPT to give you a citation for the answers it's giving you, many times it will give you a citation that doesn't exist. So it will tell you that, okay, based on information from this scientific paper that was published in this journal, this is the correct answer. But when you try to check the literature databases to find out whether this scientific paper actually exists, many times you'll find that, oh, that paper doesn't actually exist. It's all made up by LLM.

0:20:45.3 GS: So this is a very big problem for these Generative AI systems.

0:20:50.7 SH: Thank you so much for providing those examples. I think the most recent one you shared about creating sources that don't exist, I think that's a very important point that once we get those results from the AI models, is contextualizing them and looking at the why. Why did I get this answer? So always keeping that in mind when using these technologies. Yeah, that was very interesting. Thank you.

0:21:14.7 GS: Thank you. I agree. You definitely need to contextualize the answers yeah.

0:21:20.1 SH: So knowing the challenges and the limitations, a common question that comes up is the ethical considerations when using AI, such as bias and environmental impacts. So what are your thoughts about those?

0:21:33.6 GS: Yes, so this is a very crucial issue, thinking about these AI systems and other emerging technologies. I think there's been a lot of emphasis on a lot of like

data sets that go in into AI models are biased. But I think we need to take a step back and start from much earlier in the development of AI models. So we need to begin from the stage of the teams behind AI models. So by teams here, I mean to develop an AI model, you need an engineering or data science team. And I think we need to ensure that the engineering or data science teams are not biased. So those teams, to begin with, they need to be representative of the human society. 'Cause if they're not, then all these are the biases that occur downstream. So like biases in data, they all begin by who is in the team. 'Cause who is in the team determines what data goes into that like AI model.

0:22:40.0 GS: It determines what assumptions are made. And so I believe that we need to have diverse engineering teams to begin with. And then, of course, secondly, the biases in the data, which comes in many forms. So you can have gender biases, you can have racial biases, but also you can have biases in social economics. So you can have biases between urban and rural populations. You can have biases between like young versus the elderly. Right. That if your data that is used to train the AI model comes only from young individuals, then you cannot actually rely on it when working with the elderly. So I think this is a big problem that we need to address all forms of biases that do occur in these data sets.

0:23:32.1 GS: We also need to think carefully about the human supply chain. So in terms of AI models, like ChatGPT or other LLMs that require humans to label, to train the computer how to perform certain tasks. And a lot of this involves manual labor, where somebody is paid to basically train the AI system by labeling things like text or images moderating content on Facebook. All of that is done by humans in order to train AI models.

0:24:07.5 GS: So we need to be cognizant of this manual labor that the people behind are fairly, they're fairly treated. We need to think of their mental health, where they're doing like content moderation. And then large AI models, especially LLMs, they consume huge amounts of electric power to train them. And we know this has a negative impact because of greenhouse gas emissions from many of our energy sources. And then there's the issue that AI models can be used for misinformation. We've seen the rise of misinformation because now people can... They can create like fake videos that basically try to mimic the identity of somebody. And this is a huge issue that has to be addressed.

0:25:00.0 GS: And then there are also like privacy and security, security risks, 'cause these AI models rely on data. And the people whose data is being used, their privacy rights have to be respected. And these AI systems have to be secured so that we can prevent the impact of hackers getting access to some of the information in AI models that could then be them to the underlying data.

0:25:30.3 SH: Shifting gears just a little bit, what exciting advancements or emerging trends in AI do you think will shape the future of infectious disease research?

0:25:41.9 GS: Yeah, so that's a good question. So when it comes to infectious disease research, in some ways, there are many infectious diseases that are underfunded today. So the research is underfunded. And the impact of that is that we have less drugs that are highly effective and safe for many infectious diseases. So in many parts of the world, neglected tropical diseases have affected like millions of people around the world, especially in low and middle income countries, but that receive less funding. And so AI models can reduce the cost of developing new drugs for these underfunded infectious diseases.

0:26:28.7 GS: I already mentioned a little bit that we are facing an urgent need of new antibiotics because of antimicrobial resistance or AMR. And this is an area that, again, AI can have a big contribution. And also, when it comes to pandemics, addressing pandemics at all levels, so trying to identify pandemics early at their source, predicting those areas of the world where new pandemics might emerge and trying to control and eradicate emerging viruses before they spread around the world would be something that can be achieved using AI and other technologies.

0:27:08.5 GS: And eventually, the world should be able to rapidly develop new vaccines and drugs, perhaps even within 100 days of a new virus to develop a new drug or to repurpose existing drugs for a new virus. And there will be an initiative underway called the 100 Days Mission by the International Pandemic Preparedness Secretariat that is trying to bring together teams around the world, so from large pharmaceutical companies to academic groups working on drug discovery to figure out how to quickly prepare for the next pandemic in a way that new drugs and vaccines could be delivered within 100 days. And I see AI as contributing to that effort.

0:27:58.0 SH: That would be amazing if we could have a vaccine so quickly, so I really hope we do get to that point. So we've covered a lot in regards to AI today,

and I wanted to ask you, if you had to pitch the use of AI, what would you say is the main value add?

0:28:16.6 GS: Yeah, so that's a really good question, especially as these technologies changed really fast. I see that one of the biggest impacts of AI will be in helping to level the playing field across many areas of work and daily life. That AI will accelerate the speed at which we can develop new things using knowledge across several areas even when we are not experts. So I see that AI will enable us to personalize art or music without being an artist. You'll be able to develop an app without being a programmer. You'll be able to contribute to scientific research without having a science degree. And I think by leveling this playing field, it will open new ways of doing things faster and cheaper, because others around the world who don't have, like large resources will be empowered also to participate in the global economy.

0:29:23.5 GS: So I see this leveling of playing field as one of the important outcomes that AI will have on the human society.

0:29:33.8 SH: As we wrap today, what is one thing about your job that keeps you coming back day after day, whether or not it's related to today's topic?

0:29:43.6 GS: Yeah, thank you. So, yeah, so one thing that's always on my mind for for several years now, having grown in Kenya, in a very under-resourced setting, I'm always drawn by the need for the process by which we do science and the process by which we make new discoveries, especially new medicines, how that process can be changed so that we can discover new drugs at a faster speed and lower cost for thousands of diseases, including cancer, viral diseases, rare and common genetic disorders. And especially as now we live in a world that is facing accelerating environmental change. Being able to make scientific discoveries at a faster pace will enable us to tackle some of the grand challenges that we face as a society very rapidly at a lower cost and by inviting as many people around the world as possible to participate in this process.

0:30:57.5 SH: Dr. Siwo, thank you so much for joining us today to share your experiences and provide some insight on this topic. To our listeners, we hope that you've learned more about AI, and we encourage you to check out the transcript and resource in the podcast notes. With that, we'll end here for today. Stay safe and stay prepared.

0:31:16.0 GS: Thank you again for the Storee. It was a real pleasure.