IlLUminate Blog Transcript: **Kofi Arhin on AI and the Hiring Process**

Recorded November 1, 2023. Listen to it here.

ANNOUNCER: 00:04	[music] This podcast is brought to you by ilLUminate, the Lehigh Business blog. To learn more, please visit us at business.lehigh.edu/news.
JACK CROFT: 00:15	Welcome. I'm Jack Croft, host of the ilLUminate podcast for Lehigh University's College of Business. Today is November 1st, 2023, and we're talking with Kofi Arhin about the current state of artificial intelligence and his research regarding how AI may help make the hiring process more fair and equitable. Dr. Arhin is an assistant professor in the College of Business Department of Decision and Technology Analytics, known as DATA. His research interests include artificial intelligence design and implementation, information security, ethical issues in information systems, human-computer interaction, and web technologies. Kofi, thanks for being with us today on the ilLUminate podcast.
KOFI ARHIN: 01:07	Thank you for having me, Jack. I'm really honored to be here.
CROFT: 01:11	Now, it's kind of a big topic, so I thought we'd start by setting the stage with kind of an overview of where we're at with AI today. Because since ChatGPT launched almost a year ago now, which introduced a lot of us, I think, to the generative AI for the first time, there has been considerable debate about its ultimate impact on the world. Proponents on one end see it as a bridge to a utopian future in which human suffering is greatly reduced. And critics on the other end see it as nothing less than an extinction threat to the human race. I think most people seem to be somewhere between those two poles. So I'm wondering, what do you see as the current state of AI?
ARHIN: 02:00	That's a very interesting question, Jack. I think to talk about this, we'll probably have to take a step or two back to understand how we got here. And then I can speak to the concerns that people have raised and the future opportunities that AI brings, right?
CROFT: 02:20	Sounds good.
ARHIN: 02:21	So not to bore anyone with history, but AI has in the past provided people, companies, and society the opportunity to automate processes. And so with advancements in technology and production of technological equipment, we have seen the capability of these basic AI tools grow from performing mundane tasks to taking care of very complex activities, right? And so when large language models became popular in 2018, beginning with BERT [Bidirectional Encoder Representations from Transformers], it was not easy to predict that five years down the line we would have tools like ChatGPT. But one of the key things that has fueled this growth is the ability of industries and companies to produce chips and memory cards and all of these computing resources at a very fast-paced rate. And so with these advancements, it becomes very easy-- or we have all the tools we need to build more sophisticated tools.
ARHIN: 03:45	Now, the way AI works is-- one of the key things that is required for good performance or efficiency in any AI system is data. You need data to train your AI

models or systems on so that it can come out with the expected output. And so generally, the more data, the better. And so if you have resources that allows you to store large sums of data, then your AI models can also be trained on these large sums of data. They require higher-speed processers and higher memories to be able to come up with accurate predictions and outputs. And so with these advancements, there are a lot of opportunities for growth, like I said, for businesses and for society as a whole. We are seeing AI penetrate all facets of society and different industries. You're talking about health care, HR [human resources], media, and so on and so forth.

ARHIN: 04:56

For example, in health care, doctors and hospitals are using AI to assist with patient care. Sometimes it takes just a scan of the human body for an AI assistant or tool to diagnose a condition for a patient. And the doctors can then pick on this information and work with. In hiring, a lot of companies are using AI to help pre-screen job applications that they receive, and this makes the work of the hiring managers very easy. So in the past, if you had to manually look through millions of applications, now your AI system can easily go through your application database and make recommendations for you. On the other hand, the concern here is that if you're training AI on data, right-- now we have the capability to train AI on large sums of data. The issue here is that these data sources contain historical human decisions, right? And so if these human decisions were biased or discriminatory, for example, to some extent, then the AI is now equipped with the capability to exacerbate these discriminatory practices or these biases, right? And so the data feeds the AI and the AI learns the patterns of historical human decision-making.

ARHIN: 06:38

That's one of the challenges with AI, especially with generative AI. The concerns are even more serious than that. Now, the way generative AI works is it learns from all of these data sources, again, because we have the tools and resources to give it all the information it needs, and then when you put in a query, it combines all of the knowledge it has to give you the best output. Now, imagine a future where we give everyone access to generative AI to solve their own problems or to interact with this generative AI. It will be very difficult to put safeguards in place to prevent people from using these generative AI resources for harm. For example, bad actors could use it to develop weapons of mass destruction. Recently, you may have heard about deepfakes, right, bad actors using the face or the image of celebrities to create some kind of misinformation or disinformation campaigns online. And so all of these are valid concerns.

ARHIN: 07:57

So I've told you about the good parts and then the bad parts, right? Now, the statement sitting on the fence may suggest that, OK, you're not sure whether it's a good thing or a bad thing, but just like any tool or asset you have, it is possible to be-- it is possible to be excited about the opportunities that it will bring whilst also being mindful of the concerns. And so, for me, I am very hopeful that AI is going to improve a lot of our processes. AI is going to make us smarter. It's going to make us more efficient. People might lose their jobs. That has been one of the concerns that have been raised, but there will be other opportunities to improve human skills. And so that's what I'm excited about, the opportunities that AI and all the advancements in generative AI is going to give us. And for me, it's made some mundane tasks very easy. When I'm working with code and I'm trying to debug my function, I no longer have to spend time on Stack Overflow looking for the exact solution to my problem. I put it in ChatGPT and then it gives me the solution right away. Saves me time.

ARHIN: 09:22	And so there are some tasks that these tools can help us to improve on, but we also need to be mindful of the harms it will bring. And so I was excited that the U.S. is one of the countries leading oversight over these tools or the recent policy that the government just passed. I haven't read it in detail, but I know some of the-- I know about some of the safety guidelines and policies that the government is trying to put in place to prevent a future where, yes, we have this great tool, but it is very difficult to be excited about it because it's so easy for bad actors to use it to harm people and society. That was long. I mean, we can break it down. You can ask me questions about the different things I've touched on, but. Yeah. I got too excited. [laughter]
CROFT: 10:23	We like people who are excited about their work here, so. Yeah. You had talked about some of the safeguards and some of the concerns in hiring, and we'll talk about that in detail in just a few minutes. But I'm wondering what some of the main things we should be paying attention to in particular regarding the potential future of AI. What are some of the things we should be watching out for to make sure that it doesn't go in the wrong direction?
ARHIN: 10:59	That is a very difficult question to answer in the sense that-- like I said, five years ago, I don't think anyone predicted that we'd be here, right? And so some of the basic things we have to look out for: which companies are developing these tools, who is getting access to these, what are the safeguards put in place, how transparent are the companies who are building these in talking about the challenges they face and the potential harm that the tools that they are developing might cause? For example, in the U.S. - I think a law was passed some time ago - when a company faces a security breach, they have to announce it or publicize it. I think this was in 2005. And so the companies are forced to talk about these breaches, and sometimes they have to compensate people who have been affected, right? In the same way, I think that when we start getting to the point where companies are, yes, highlighting opportunities that these tools bring but also talking about the potential harms and the shortfalls they've had or they are encountering now, that will be great. It will open everyone's eyes to the potential dangers ahead of us and even invite people who have solutions to these problems.
ARHIN: 12:37	I think that the fact that there are people who are skeptical about these tools and there are people who are excited is a perfect scenario. I would rather have been more concerned if everyone was excited and jumping on this bandwagon and were all going in the same direction. I think it is good for us to have people coming from different contexts on these tools that we are building. And I think that will help us to get to a consensus or a region where, yes, we're excited, but we're also mindful. And I think that would be the best place to be in. Yeah.
CROFT: 13:21	I think it might be helpful, then, to talk about one of the studies you've done recently - and you've done some other work in this area as well - looking at this question of how AI can be enabled to help make the hiring process more fair and equitable. And you had talked about-- when you feed large amounts of data into AI for something like the hiring practices, then the human biases that went into those hiring practices over the years kind of gets transferred in there with it. So if you could talk a little bit about that history of what the experience has been in human resources or HR with the use of AI, some of the problems that have been highlighted quickly, and then what it is that you're looking at that might be able to make things better. Because I can say, as someone who's been a hiring manager at several places, when you're

screening those initial applications, up to 1,000 people, that is tedious. And there's a huge percentage of those that are just-- they're nowhere close to being qualified. So having a tool that fixes that-- I get that. That would certainly be a boon for efficiency in the job process, so.

ARHIN: 14:55

Yeah, that's great. And like I said, AI is a great tool that's helped companies hire the right people over time. As humans, we all have our biases. I have my biases. You have your biases. And in my short academic career so far, I've found studies, and just in my experience interacting with people, that human biases are very difficult to remove. And I don't mean this in a bad way. Sometimes biases help us make decisions quickly when we have very little information. And so there's nothing wrong with-- while, in my opinion, we are humans and we have our biases - and I don't think anyone should feel bad about it - the issue is providing an intelligent agent access to decisions that we've made in the past and allowing that agent to replicate those decisions in a short amount of time. That's where the problem lies, right?

ARHIN: 16:03

And so the challenge, from literature and from reports, has been that these AI tools are learning from data and these data contain historic human decisions that may be discriminatory. For example, research has shown that the way people look, whether it's the color of your hair, your skin color, even the way you dress sometimes, can impact a hiring manager's decision to employ you or to move you to the next stage of the personnel selection process. Other studies have also looked at how language impacts the interview applicants-- or job applicants. And there are several standardized tests that have also been shown to affect underrepresented group members negatively. And so in the U.S., the Civil Rights Act of 1964 and the Uniform Guidelines provide guidelines on how people who apply for jobs should not be discriminated based on age, race, religion, gender, and so on. So going back to the hiring with AI, when you train AI models on historic human decisions, you might be replicating or automating the biases in those decisions, right? And so the solution to this challenge, especially in hiring, has been identifying the human biases and trying to address them.

ARHIN: 17:44

In my study, I look at it from the perspective of loose coupling, and I make the argument that-- let's assume that a company has identified that their AI system could be discriminatory, right? What they have then to do is to go back to the data and try to take out information that may be highly correlated with a particular subgroup, like race or gender or religion or something, right, so all the features or attributes in your data that are correlated with these different protected classes. Now, the challenge, I argue together with my co-authors, of course, is that when you start manipulating the data, you have to do it for-- so let's talk about race, for example. If you're doing it for Black applicants, you have to do it for Hispanic applicants, you have to do it for Indigenous applicants, and so on. There are so many groups in race alone. Then you go to gender. You have to make sure that people with different gender identifications are being treated fairly. And then you go to religion, and so on and so forth. In the end, you're going to have a data set where you have removed a lot of information about underrepresented group members or you have tried to manipulate the data in such a synthetic way that it is no longer representative of society. When you train an AI model on this data, yes, it might give you, in the short term, fair outcomes, but in the long term, you're going to have an AI system that has a lot of information about majority group members, because you are not taking their information out of the system, and knows very little about underrepresented group members, right? That's one setback.

ARHIN: 19:35

Two, in manipulating your data or training your AI to be more fair, there's a potential that you might be discriminating against the majority group members, right? And so let's look at gender in terms of majority and underrepresented. So let's say we have a job vacancy where male candidates dominate the application pool. They've dominated the successful hiring process over a long period of time, right? I'm assuming two gender types. And so the female-- if they are the underrepresented group, now you have to treat them differently if you want your AI system to address that historical human bias. The male group, which is the dominant group-- in treating the female group fairly, you might be discriminating against the male group.

ARHIN: 20:33

I'm trying to think of a simplified example, but the best that comes to mind is-- and this is a very basic example and it's not practical at all. Let's say there's a test, right? So male applicants, female applicants have to take a test, and the cutoff point is 80% or 80 of 100. You have a lot of male applicants making the 80% mark and female applicants, again, just in the context of this example, not making the 80% mark, right? What you have to do then is to adjust the system to treat female applicants separately from male applicants, right? Now, in doing that, you can run into a potential legal challenge where you are discriminating against the male applicants. Remember, the law applies to all. The law is not only meant to treat a particular group fairly and another group unfairly, right? And so there could be legal challenges.

ARHIN: 21:38

What I propose in my study-- and again, I want to highlight that this example I've given is just a very simplified basic problem just to explain my point. Now, what I do in my study is I argue, "Hey, let's keep all the representation. Don't even touch the data because it represents something about these applicants that are important, right? So keep all the information in the data. Do not remove words. Do not remove attributes or features that you find might be highly correlated with subgroup membership. Let's find other sources of information." Let's assume that we are training the AI on our data, right? We are training the AI to model historic hiring manager decisions. In my study, we argue, "Let's train the AI on other sources of information that can help us identify the best candidates in our pool. So don't just rely on historic hiring manager decisions in the past. Train the AI to learn who a good applicant is from other external sources, right? Now, when you've done that, bring it together and then compare what the AI has learned from outside the organization to what your historical decisions have been. Find the intersection between the outside knowledge and the internal organizational data, and then pick the applicants at the intersection of these two sources."

ARHIN: 23:17

Now, the rationale behind that is training your AI on external sources introduces some objectivity. Well, that's our hope in the study that it introduced some objectivity in the way the AI makes decisions. This is not to say that AI systems are not objective, but whatever decision the AI is making is relative to the data that belongs to the organization, right? So don't just rely on the organization information. Take information from external resources. Find the intersection between that, and then your AI system now is equipped with enough knowledge to know who a good applicant is. So we are using information-- we are still using the historic hiring manager decisions, but we are also using external knowledge, and we are finding the intersection between the two. And we find that when we do that our AI systems are fairer. Now, the way we measure fairness in our data is we use the adverse selection ratio or the adverse impact ratio. Now, this is the ratio of selected underrepresented group members to the ratio of majority group members. And it's given by the civil rights legislation. So when this ratio-- it's given by the Uniform Guidelines. When this ratio is below the 0.8 mark, we can argue that there has been discrimination in the hiring process. It's also called the four-fifths rule, in case anyone is looking it up. And that is our goal, that when we train our AI models on historic decisions, the predictions on who gets hired or who advances to the next stage of the selection process should meet that 0.8 threshold. And that's what we see when we train our AI systems on external sources and the organization's information.

CROFT: 25:16 Now, you had mentioned something called loose coupling as you started here. Can you just explain briefly what loose coupling is and how that fits into this?

ARHIN: 25:28 Okay, great. That's a good point. So the process I described to you of learning from different sources is what we describe as loose coupling. Now, let me talk a little bit about coupling theory. It's an organizational theory with-- it has dimensions and mechanisms. But basically, a tightly-- you can think of your mobile phone as a tightly coupled system in the sense that if your battery dies, it doesn't matter if you have a perfect screen or perfect AirPods, everything else. Nothing works, right? But in a loosely coupled system, your battery could die, but your screen could have an alternate power source, and your AirPods or your earphones could also have an alternate power source. A breakdown of one function does not prevent the other parts of the tool from functioning, right? So we argue that machine learning pipelines are tightly coupled in the sense that there's a focus on one source of data, historic hiring decisions in our context. Now, everything in the machine learning pipeline relies on this data source: what has been the hiring manager decisions in the past? What features, what attributes, what sources contribute to the decisions of the hiring manager? So this is a single, narrow view, right? And that's what we call tight coupling.

ARHIN: 27:02 And so in a tightly coupled system, when there's discrimination in the hiring manager decisions, you can easily see why it will translate into the AI's output because it's a single source of knowledge and that information flow goes through the entire pipeline. But in a loosely coupled system, a breakdown of one part does not impact the other parts, whether immediately or later on, and it gives the other parts also time to recover. And so if there's a breakdown in one part, it does not translate to the other part. And I'm using breakdown here in loosely coupled system to refer to-- for example, if there's discrimination in historic hiring manager decisions, in a tightly coupled system, that discrimination flows through the machine learning pipeline. But in a loosely coupled system, because we are using alternative sources of information, we are decentralizing the process. We are breaking that direct relationship between features and historic hiring manager decisions. So all of these things impact the fairness of the model. And so we show that the loosely coupled systems are fairer. And even if you want to argue that there could be a reduction in accuracy, we show that, in fact, at equal accuracy levels, loosely coupled systems meet the 0.8 threshold of the four-fifths rule.

ARHIN: 28:31 Another advantage of our study is that-- if you've read a lot about fair algorithms, you might have come across the accuracy-fairness trade-offs. That is, if you're making an algorithm fair, what is the guarantee that, in adjusting for fairness, you are picking the right candidates? Because these are candidates that historic hiring managers would have rejected, right? And so you are not given any assurance that these fair decisions that you're making will be the best decisions for the company because you're not sure if fair algorithms mean better quality candidates or you're adjusting your model just for fairness's sake, right? You also want to ensure that there's quality. And so the good thing about loosely coupled predictions is that in learning from other sources, you are actually learning the attributes and features of quality candidates and bringing that in, right? So that concern about the accuracy-fairness trade-off is reduced to some extent because the algorithm already knows what a quality candidate is. In a tightly coupled system, the algorithm relies on the hiring manager to provide information about quality candidates. In a loosely coupled system, the algorithm has already learned from different sources and it's checking with what the hiring manager says and finding an intersection between the two. I hope all of this is not too technical. I really didn't mean to go into technical aspects. Yeah.

CROFT: 30:11

No, I think that's been very helpful as an explanation of the balance you're trying to achieve and what the problem is that you're trying to address and how you're trying to address it. And I'm not sure there's really any way to do that other than to get somewhat into the weeds on it so that people understand what's going into these kinds of decisions. And I think it's also a good example of the complexity of AI in terms of those broader problems we're talking about, why it's not easy to just snap your fingers and say, "OK, we're going to fix this." When there's that much data involved, it becomes complex very quickly. I think that's a good place to wrap up. I want to thank you again for joining us on the ilLUminate podcast today. And it sounds like we'll have plenty more to discuss on the topic of AI in the future.

ARHIN: 31:13

Oh, yeah, definitely. I'm happy to share all of the work that I'm doing, and I'm excited that this is getting the attention of people. And the more people talk about this, the more it helps the community, the more efficient and equitable systems we have in place to serve all of us. So yeah, exciting times ahead.

CROFT: 31:36

We're grateful that you're looking into the area of the ethical and legal and even, to an extent, the moral implications of some of the problems that we do need to face up to with AI. Our guest today has been Kofi Arhin, who has taught courses on artificial intelligence in business, human-computer interaction, and ethical and legal issues in computing. His current research projects aim to proffer solutions to ethical concerns in the design and implementation of AI systems. This podcast is brought to you by ilLUminate, the Lehigh Business blog. To hear more podcasts featuring Lehigh Business thought leaders or to follow us on social media, please visit us at business.lehigh.edu/news. I'm Jack Croft, host of the ilLUminate Podcast. Thanks for listening.