

Transcript of Conversation with Sophia Johnson (Program for Economic Research, Columbia University), and Chengkai Li (Professor and Associate Chair, Department of Computer Science and Engineering, The University of Texas at Arlington)

Wednesday, February 24, 2021

Sophia Johnson:

Good afternoon and welcome to the Department of Economics’ Program for Economic Research virtual live series. I’m Sophia Johnson with the Program for Economic Research, one of the world’s leading programs for identifying opportunities and strategies for enhancing economic research. Thanks for being with us today.

This event will be livestreamed by the Economic Department's YouTube channel the conversation will be recorded and closed captions will be provided in the days following. By attending, all conference participants agree to abide by the event's Code of Conduct, which is posted on our website, [econ.columbia.edu/per](http://econ.columbia.edu/per). As always, the Program for Economic Research in the Department of Economics at Columbia University takes no institutional position on matters of policy.

A little housekeeping before we begin- each presenter will speak for 10 to 12 minutes. In this case, we have a keynote speaker. The presenter has agreed to take all your questions live or in the chat at the end of his presentation. And you may unmute yourself and ask the question or simply type your question in the chat box if you're joining us on the livestream, the live YouTube feed, you can post your questions or you can also add them on social media and we will incorporate that into today's discussion.

Lauren Close, the Program Manager here at PER, is joining us as well. She will be adding updates on social media as well as monitoring and posting comments in the chat box during our conversation today. We invite you to follow us on Facebook, Twitter, and LinkedIn for this event. We're using the hashtag #PERLiveSeries.

So, Artificial Intelligence: The Implications for Business Strategy. We start this hour defining the many different entry points for AI in the global economy and exploring how far we have come on the path toward achieving this vision of future productivity through AI and the ways organizations can improve their odds of success.

Our keynote speaker is Dr. Chengkai Li, Professor and Associate Chair in the Department of Computer Science and Engineering at the University of Texas at Arlington. Dr. Tulia Plumettaz of Wayfair will not be joining us today. Dr. Li is the Director of the Innovative Data Intelligence Research Lab at UT Arlington, where the focus is on building impactful interdisciplinary research in several areas related to big data intelligence and data science. His research interests include data management, data mining, natural and language processing, machine learning, and their implications in computational journalism. Dr. Li pioneered the field of computational fact checking and has produced a significant body of research on topics related to knowledge graphs. He has also led a [inaudible] organization and multidisciplinary team of researchers and partners from academia, government, and industry to conduct NSF convergence accelerator projects on credible open

knowledge network, which really aims to ensure the credibility of decision making software powered by knowledge graphs. Professor Li, welcome to the Econ Live Series.

Now, to begin, research on the economics of Artificial Intelligence almost exclusively focuses on the potentially transformative economic purposes of AI. This thesis is that it may significantly lower the cost of prediction. However, only 10% of companies by the story that they can obtain significant financial benefits from artificial intelligence technologies, why so few?

Chengkai Li:

Hi, Sophia. First of all, thank you for having me here it's my honor and great pleasure to have this opportunity to speak to you all. So how does this go? Should I address your question first or... ?

Sophia:

Sure, you can certainly start with the question of why there are so few and then move right into your discussion.

Chengkai:

Sure. I think the term 'artificial intelligence,' of course, nowadays, it is a focal point of our conversation in the society. Everyone is excited about it, everyone is talking about it, and, you know, this term can refer to very broad, you know, subjects and scenes. It can pretty much refer to everything related to computer science.

But in the narrower sense, you know, what makes people most excited about artificial intelligence is really the development of - and advancement of - machine learning, deep learning technologies and their applications in the past about 10 years and really breakthroughs there. So you know I think there are several reasons. Artificial intelligence isn't nothing new at all, right? It actually started as a discipline fifty, actually longer than that - seventy years ago, pretty much at the same time that computer science started to form as a discipline. And for a long time, people really didn't make important breakthroughs, but it started to take off really quickly in recent years for multiple reasons. One is that nowadays we have data everywhere. Our capacity of generating data and collecting data and analyzing data is much greater than before. And so that's number one: because data is everywhere. And number two: it's because of the advancement of, you know, computing infrastructure and hardware, including CPU, GPU, larger memory and storage cloud computing technologies and so on.

And then number three: of course, you know there, there were algorithmic advancements from computer science, you know deep learning particularly. So it's natural that every business feels the urgency of adopting AI and to be prepared. To, you know, full impact created by a advancement. So I, you know, I do not have first hand experience with regard to the statistics, you mentioned, you know 10% or 15% - I'm pretty sure those of you in the audience know that number in a deeper sense than I do, but my guess is that you know, there could be multiple reasons. Number one is that it really requires resources and talent in order to tap into the capacity of artificial intelligence, right? And there also needs to be realistic need. If you are a small say, you know, personal business, and you don't really have large amount of data from which you can gain insights, then you know, the

chances are there's less you can leverage by using artificial intelligence and you do not really have the resources to carry out whatever insight and decisions you can make based on artificial intelligence.

Yeah, but I think for corporates and large businesses, it's not really that, you know, a desire it's really a must to have investment in AI technologies in order to stay cutting edge and competitive in their businesses. I hope, to some extent, that addresses your question.

Sophia:

Thank you. So I'll turn the floor over to you now to begin your presentation and we'll take questions at the end.

Chengkai:

Cool. So I prepared some slides to basically summarize some of the projects that I've been doing in the past, say, decade. I understand, this is a very brief presentation, so will, I will not go into technical details, so you may hear quite some jargon so bear me with that. And I understand that there was another originally there was another panelist. They couldn't participate today so, in theory, I may have a little more time, so I'll indulge myself to go beyond the originally assigned 10 minutes to me. And I have a few slides, and after that I can share a video with you about a project that I recently worked on.

Sophia:

Excellent Thank you.

Chengkai:

Let me share the slides with you. Just one moment. All right, are you able to see the slides?

Sophia:

Yes, perfectly.

Chengkai:

Right, so a lot of the focus that my research group has - and our research is on, you know, misinformation and fake news, and most specifically, developing computing tools to help fact checkers and the reporters, and the public in tackling misinformation. And I don't need to explain what is misinformation / fake news. You know, I'm pretty sure all of you have heard of it, and you know, this has become an increasingly important challenge to our society, right? And here are just some statistics and to just remind you about the impact misinformation has on our society and economy as well, right? And if you're not familiar with what people are doing in tackling misinformation, I just want to mention a few organizations, you know the fact checkers - Washington Post, Politifact, the New York Times, and factcheck.org, and so on, there are many more. And so they are working diligently in basically vetting factual claims made by people and organizations and inform the public about the truthfulness of those factual claims. However, they are not able to keep up with the large amount of information that is spread, you know, spreading online. For instance, we have learned that typically a fact checker will need to spend between several hours and a day or even longer to really investigate a piece of misinformation and write about it and

publish it. So, you know, you can compare this with the huge number of you know, information online, including a large portion of the social media posts on Twitter and so on, that you know, that that are spreading misinformation, right? So that's a huge challenge you know, in terms of scale for fact checkers to deal with.

And here's just a more concrete example to explain of why fact checking is something that is not trivial and takes time right? Of course I'm only using this as a example that's available from Politifact. There is no, there is no personal political stance here. This is a factual claim made by Mitt Romney a few years ago. He said our navy is smaller than it's been since 1917. So it seems that he was referring to statistics about the number of, you know, battleships in the United States Navy, since its creation in 1916. So, if you look at the numbers, indeed, the number of battleships in 2009, in 2012 - that was when the claim was made - the number of battleships at that time was close to a historical low, right? So in that sense, if you literally interpret this chart and his claim, you would say his claim is largely true. However, his claim was rated by this fact checking website called Politifact. Rated by it as Pants on Fire. And the reason is the claim essentially was comparing the battleships 100 years ago with the modern aircraft carriers, right? And so you cannot just look at the number of battleships in measuring the strengths of the Navy. So this is an example that demonstrate the subtleness and, you know, complexity in fact checking. Oftentimes it's not about, you know, looking at, you know, data from somewhere and evaluating whether it checks out and then you actually need to have deeper understanding of the context of it.

So we have been working on this project called called Claim Buster for the last seven or eight years, and our goal is to automate the process of fact checking. And we say this is toward the holy grail of automated fact checking. This is just to say that this problem itself is to some extent overwhelming and daunting, right, and we are not really close to our truly automated, you know, fact checking system that that can tell you immediately whether something is true or false. I mean, we can do that, to some to a very limited extent on certain types of claims, but largely, this requires, this is an example where artificial intelligence systems and human workers need to collaborate together in order to achieve certain goals. But nevertheless, we have made progress in some of the directions, where automation is more possible and more effective and particularly we looked at how we may help fact checkers in deciding what to fact check. So we call this problem, the claim spotting problem. So this is a screenshot of a file that Politifact called a buffet of factual claims basically they had in terms collecting factual claims made by people and appear on various you know TV programs, news outlets, and so on, and then they highlight those factual claims and then they decide which ones to further investigate. So our tool aims to develop a machine learning algorithm that is able to rank factual claims. So that fact checkers can look at that rank the list and focus on top rank the factual claims. Because the top ranked ones are more likely to be the ones that deserve their attention and they have limited bandwidth and so it's really important for them to be selective in focusing on what to fact check.

So, you know, this might be the most technical slide in my presentation, so I hope that's acceptable. This is a fairly typical supervised learning task. So basically we looked at, you know transcripts from all the past January election presidential debates, and then we look at all the sentences, we use human experts to annotate those sentences in terms of whether a sentence is a factual claim that is worth checking out or not. So by doing that we we annotated 23,000 sentences. And that's the

collection of ground truths that we have, and so we have various machine learning models learn from those ground truths. And then, this is our claim buster claim spotting model. Now, you can apply this model on a live event or apply that on a web page or news article and then this model will be able to rank sentences in that piece of text and recommend sentences that - recommend the factual claims that are worth checking to the fact checkers. So, for instance, we, well, I'll talk about that later.

Just a little you know slide to give you examples of you know, what are factual claims - what a factual claims that are important and the worst checking. So, you know, for instance, "I was in Iowa yesterday." This is a factual claim, but most likely, this is something that is not important or worth checking. "I ate a burger yesterday," you know, and so on. And, and there are factual - there's, you know - there are statements that are not factual claims. They could be opinions or questions and so on, right. Something like "I will be tough on crime." And that's really a pledge but that's not a factual claim. Or "Seven 7% of unemployment is too high," and you know that's an opinion that's not a factual claim. And on the other hand, the top three sentences in this slide - these are considered important factual claims by our human annotators. So the machine learning algorithms look at these annotated sentences and figure out the signals from these sentences and then build a model, so that when the model is applied on future sentences, it can make a prediction with regard to whether that sentence should be fact checked or not.

So we have built a public API. Basically, this is a code base where programmers can tap into and they build their fact checking programs. That can call our API if they have a statement, they can call our API, and our API will tell their program whether that statement is worth checking or not. So they can use this for various purposes. And you know our code base is also publicly available, and yeah we welcome people to work together with us in developing this and contributing to this. So if you go to our project website, you will find out more details.

This is just to show pictures that I took, I think, six years ago, when we worked on using this tool to fact check presidential debates. This was in 2016. We have some device connected to our TV and we take closed captions from the device and then run claim buster models on the closed captions. And then Claim Buster will figure out what you know sentences are worth checking. And so you can imagine fact checkers can look into this and decide what to fact check at that moment. And we have been applying Claim Buster on all past presidential debates, and you can find the details about this on project website. This has been used by fact checkers, you know, there there's a news article from Washington Post about how they pick up a lead recommended by Claim Buster and so you know the fact checker said it would have been lost to history if it had not been for Claim Buster. And so this is a screenshot of a newsletter or alert email that is produced by reporters. They apply Claim Buster on transcripts of TV programs, social media, and even Congressional records, and then they find that those highly important factual claims and then they compile those claims and send those to fact checkers. And so, to some extent you see this, this is a match of the Politifact buffet, that I showed earlier, and this shows that Claim Buster, to some extent, can be used to do the work that intern was doing. You know, for Politifact. We are also using Claim Buster to monitor factual claims people made on social media, particularly in Twitter. So if you go to this project website (<https://idir.uta.edu/claimportal/>), you will find the factual claims made by all major politicians, you know House representatives, Senators, the President, and so on. And you look at the claims they

made, you can look at the scores that Claim Buster gave on their claims in terms of whether they are worth checking out.

There are some other tools- for instance there there's a browser plugin that can flag health misinformation related to COVID-19 and so on. We also built a dashboard for for tracking COVID-related misinformation. So if you click at one place, it will show tweets from the government official, but it will also show what kind of factual statements and misinformation has been spread by you know people from from that area.

Well, I have some more slides that I can use, maybe, in responding to some of your questions, but I'll stop the running of the slides at this moment. Do we still have time for showing that video, Sophia?

Sophia:

Sure, we certainly have time for the video and then maybe we could talk a little bit about the economic... its reaches - the reaches of the work you're doing in the economic space. You know, hypothetically. And then we can open it up for questions.

Chengkai:

Sure, so let me share, let me share that video with you.

Video Audio:

Credible Open Knowledge Network, COKN, is a project in the National Science Foundation's Convergence Accelerator Program.

Good data enables good decisions. The open knowledge networks created by the Convergence Accelerator Projects are a public knowledge infrastructure that will benefit us but will the network's be vulnerable to inaccurate information? This is a real concern, as evidenced by many poor and disastrous decisions caused by bad data and misinformation. COKN is a suite of frameworks and tools. It helps software developers and domain experts build credible decision making software powered by open knowledge networks.

Let's consider an important use case: vaccine misinformation. Exposure to misinformation made many parents delay or refuse vaccines for their children. This led to completely preventable disease outbreaks, such as a measles resurgence. One in ten infants were unvaccinated in 2016. Vaccine hesitancy is a top 10 threat to global health. Today, parents are given brochures called Vaccine Information Statements from the CDC. These brochures are accurate, but in practice they help very little. They often do not address the parents' specific concerns. A parent may be worried about autism, but the brochure doesn't even mention it. Its language and style maybe unappealing or confusing. And for folks that don't trust the government, a statement from the CDC often won't help. So what should we do? Our insight is that the brochure must be contextualized for it to have a better chance of being perceived as relevant or credible. We are building an app to help healthcare workers. Imagine you are a nurse speaking with a hesitant parent. Quickly, you can enter the parent's concerns you'll get back the key points you can make based on the parent's electronic medical records the app will recommend contextualized interventions. For example, if a parent is worried about autism and doesn't trust mainstream media, the app could recommend a video featuring a pro-

vaccine mother of a child, with autism. But if a parent has read specific research connecting vaccines and autism the app would suggest articles exposing bad research and retracted publications. For a second application, consider cyber security and how software and hardware vulnerabilities are currently addressed. The standard practice is to examine a ranked list of vulnerabilities reported by various sources. Given limited resources, the analyst within an organization would fix top rank vulnerabilities first. But the risk scores ignore the organization's particular context, For instance, a vulnerability may pose high risk because attackers can exploit it over the internet. But in a power grid a device with this vulnerability may not even be connected online. Without this context, poor choices can lead to severe financial and even human loss, such as an unnecessary shutdown of a power grid. To solve the problem we are building a tool to identify and explain credible threat based on each system deployment profile. The COKN and framework and tools draw insights from these applications. They will be important across domains. Software developers and domain experts will use COKN to improve objective credibility, which is about fact accuracy. Tools will help fix bad data and explain query and analytics results. But credibility is more than data accuracy. COKN also helps developers improve the subjective credibility of their software because user perceptions are critical to making information convincing. The framework includes strategies to contextualize results from queries and analytics to match decision makers' needs and profiles. COKN captures not only facts but also inaccuracies. And it models user profiles. It uses taxonomies to semantically connect the knowledge graphs and user profiles to achieve contextualization. We've already developed some taxonomies and knowledge graph for vaccine misinformation and security vulnerabilities. Our interdisciplinary team includes computer scientists, social scientists, and application domain experts from Academia, industry and government. To create real world impact we'll work with our partners and other teams to apply, test, and deploy our technologies. Credibility is key to sustaining open knowledge networks and thus empowering the American people. Without COKN, we risk going back to square one, with low quality data and poor insights to drive decisions.

Sophia:

This is great, thank you. So, tell us, Professor Li. Thank you so much for your presentation and your contribution, your insights into, really, this remarkable space. It, you know Claim Buster, I feel like it's somewhere in the future, this might be something that we can all download on our computer, for example. For use at home, at school, so it really is a testament to sort of where we are going with the global economy.

I wanted to talk a little bit about - to the speak to the questions because the questions started coming in, on social media. One of, you know, it seems we have a lot of economists in the space - what part of the of the business would benefit most from applying some of the work that you're doing? What is it - are we talking about the supply chain, the logistics, the financial sector? How should students think about a collaboration in this space? Like, what should that relationship look like?

Chengkai:

That's a - that's a very good question, I guess, there are multiple questions so, then you try to address some of these. So, in terms of you know, the research that I've been working on and what this project COKN as well as Claim Busters - pretty much, you can say that misinformation now is

everywhere. In every business sector. And I don't know if you were able to see clearly the several examples in one of the frames in that video, in which we can see examples of misinformation across everywhere. So it's a, well, just one moment so I'm finding that slide so I can refer to that. You know there's misinformation related to vaccination, there's misinformation related to misleading [inaudible]. There's misinformation about rumors with regard to particular products. Misinformation in data about safety of highway bridges, and so on. And I, you know, I just stumbled upon a article about- I don't know if you have heard of this, you know Dasani this bottled water.

Sophia:

Yeah.

Chengkai:

From Coca-Cola. They donated a large number of bottles to Waco, Texas. And you all probably have heard of the ordeal we went through last week. So, so this is the, this is a you know good act, right, you know it was good intention but it's surprisingly, it was not received well. And it became even a public relation crisis, so this is the example that can show - and part of it, I think it has something to do with misinformation. People you know, claiming a strange flavor of the water and so on. And so really this impacts every business sector and every corner of our lives. Another example I like to refer to is GameStop. You may have heard of the strange stock price increase by maybe ten times just in this year. Because of investors self-organized themselves through social media and to act against Wall Street. And so, you know, I heard that there's misinformation there spreading in the Reddit channels of, how do they call themselves, Wall Street Banks. That's the Reddit forum where the investors are organized, yeah. So I don't know if this directly corresponds to the question you asked at the beginning, and you know if you could remind me. You know I would be happy to go further.

Sophia:

I think the students want to - I think what we're hearing is that well how can.. One what can businesses do? And two, as I'm thinking about my post-academia experience, what can I do to ready myself or what area of business, might I consider? What are the openings for the next generation of scholars sort of going into- considering working in the private sector? How can they help to find the types of solutions that you're working on? And two, what are businesses doing about these challenges? You know, what has been the response to businesses? Are they reaching out to collaborate with you? What are some of the hypothetical solutions here to the problems that we're seeing and the impact it's having on markets, for example, for businesses?

Chengkai:

So, with regard to my own project- we have been collaborating with reporters. Through them, we get connected to the fact checking community, so the impact there so far is still largely on journalistic organizations in the sense that we have tools that can help fact checkers in doing their work and mostly their work focuses on political fact checking. It's about vetting whether something said by a politician is true or false, right? So it's less directly related to, you know, private businesses and so on, but it does have - it's indeed a highly related to some of the well known examples of misinformation that impact our life, you know. One is this vaccination related misinformation and

you can extend that to COVID-related misinformation as well. So for these we're just working with the fact checkers at this moment, and not really with private sectors. But the point is that misinformation is a very broad concept, it goes beyond the political fact checking from there all types of misinformation that is there and, in that we feel we have provided an example where a cyber security business will need technologies in tackling misinformation.

Sophia:

Very good. I would like to open up the panel to questions from our viewers online, watching and joining us on YouTube but also participants on Zoom you may type your question in the chat box and also you can also unmute yourself and reveal your video, and then we could take your questions one by one, from Professor Li. A question here: "Good evening. My question would be: how has the system been back tested or vetted so that we know that the issues/questions to fact check are indeed the most relevant ones?" Professor Li?

Chengkai:

Yeah I was actually hoping to include in this presentation, a chart that shows the correlation between Claim Buster scores, and the factual claims indeed vetted by professional fact checkers. So basically, we looked at, you know things that the fact checkers decided to work on and the scores received by these statements. And we find a strong correlation there. I'll show that to you.

Sophia:

Ok.

Chengkai:

This is, this is in a paper- research paper that we recently submitted. So just bear with me for a moment, sorry about that.

So you know, the way we make sure that the tool is producing sensible and effective results is by collecting feedback from our partners at Duke Reporters Lab. And from then the fact checker community. So that that's one approach and the other approach is what I said: we conducted the investigation to see correlations between our tool and what the fact checkers decided to work on and then of course we are seeing more and more people started to use our API. You know, they sign up to get an access token in order to use our API, so we are getting more and more people requesting that token and some of them even contributed to our code base. So now, I have the paper. Let me show that chart to you.

So this is a paper that we recently submitted, let me enlarge it. Yeah, so if you see this chart and the distribution here to the left hand side you know the blue one you know, this is the distribution of scores that Claim Buster gave to those sentences that are not chosen by fact checkers to fact check. So we are looking at, let's say all the sentences from a speech by a politician that's a State of the Union speech by the President. So Claim Buster gave each sentence in that speech a score. And some of those sentences were fact checked by fact checkers somewhere. So the blue distribution, the distribution of Claim Buster scores on those sentences not factual and then these other three distributions in orange green and, well what's the name, brown color. So these are the distributions of Claim Buster's scores on sentences. Politifact, Washington Post, New York Times. So you can

clearly see, you know, the separation between these two distributions that shows that Claim Buster is able to separate sentences that are worth checking from those that are not checking. So keep in mind that not everything in this blue distribution is not worth checking. It could be that they should be fact checked but you know fact checker, they do not have resources to fact check everything.

Sophia:

Thanks for your response, Professor Li, and thank you for your question, Eric. We have another question coming in from Liz.

And this question: how does your, Liz Johnson wants to know, how does your fact checker different from Twitter's fact checker that was introduced during the 2020 presidential campaign?

Chengkai:

So Twitter, of course, they work with the international fact checking network and other organizations. They largely rely on, you know, human monitors to decide on whether an account is spreading harmful and untruthful information. Of course, they may have you know proprietary tools or algorithms that we are not aware of. Or you know, for you know the details are not made public, so they could have tools some similar to ours. Facts that say, that have factual claims, or even factual claims that could be controversy. So that's an area where potentially you know organizations such as Twitter can apply a tool like Claim Buster. So, in fact, if you remember, I showed a screenshot of our own internal project called Claim Portal and the goal of Claim Portal is to use Claim Buster to monitor and highlight factual claims made by politicians that are untrue.

Sophia:

Thank you. Also, any plans to go further to directly match the statements to machine learning driven systems which process them and present a conclusion as to their veracity?

Chengkai:

That's a - that's very - that's a very good question. So we actually had some preliminary efforts in this direction, overall, this is a much harder problem than flacking, you know, important factual claims. So weeding the truthfulness of a claim is much harder than spotting involving factual claim. So that you know, there are several directions, one can take. So you know you can think about data that could be available, you know from various resources, including the knowledge graphs I mentioned, you know things like, say, information from IMDB, from Wikipedia, from [inaudible] all these sensors you know- economics-related statistics. You have all those data. You also have non-numerical data such as you know, statements made by people- comments made by people, and you know, when you are looking at a factual claim, you can figure out, you know, try to understand that claim and then figure out how you decompose it and then leverage these various data sources to understand whether it's true or false. And, of course, this this is much harder, we had a preliminary tool that can you know handle basic factual claims. So you know if someone says "The capital city of Texas is Dallas" and that's relatively easy to figure out that that's not true right and but you know if it is something like "the US Navy is the weakest since 1912", and, as I explained at the beginning. And even if the data checks out, you will - you may make the wrong conclusion. Or you will make

the conclusion that is different from what the fact checkers will say that they may say it's actually Pants on Fire that requires subtle understanding it's very complex.

Sophia:

Thank you, we have three questions coming in, on social media. I'm also keeping an eye on the time. First question for you, Dr. Li, do you think businesses will continue to adopt AI at a rapid pace? What barriers might there be to adopting AI?

Chengkai:

I think that's very good question. I think there's no doubt that businesses will continue to adopt AI and you know these big corporates you know, like a nation, they really have large amount of data, large number of partners and customers. And they can leverage AI in looking into the data for many, many different purposes and I'm pretty sure you have encountered the stories about this right, you know Amazon using recommendation systems to figure out what customers may like to buy. You know, Netflix and recommended videos to you. Even you know the pandemic creates opportunities, you know you, you may have you may heard of at AT&T / Google working on, you know, tracking people's movement stop spreading of virus and so on. Yeah, so there's you know to me there's no doubt that there will be still a large amount of investment and development in this arena. And oftentimes it's not a choice, even if you don't do anything - you need to be prepared. For challenges related to this, you know the examples I gave related to GameStop and Dasani, right? So, for instance, a company they need to have a capacity to monitor social media - to understand the public opinions about their products and their brand. Whether it's positive or negative and whether there is misinformation rumors spreading about their brand online. And if they can, you know, spot that earlier, they can take action earlier to counteract.

Sophia:

Yeah, I think we certainly talked a little bit about that, I think you know the impetus is on companies to sort of stay ahead of the social media posts and to be proactive in how they engage with what's happening in the in the social media space, positive and negative. That's a good point. What are some of the current trends in AI that you are really excited about? That's a question coming in from social media.

Chengkai:

Yeah, well, I think all these questions are great and this reminds me, you know the the last question, there, there was also the second component about obstacles or limitations. So let me answer both together, maybe about you know the pitfalls or challenges, first. And, you know, most, I believe many of you have heard of you know pitfalls related to buyers and so on. And you know one example, I can use related to again my own research projects, you know if you think about what to recommend to the fact checkers. There could be bias if you're not careful in building our data model- building our machine learning model. The bias, can you know come from various different sources, including the bias of the human annotator so when one annotates a sentence one is deciding what's an important factor. It's also related to, you know, the conversations - the political discourse that people have online, you know what's - what appears more often what appears less frequently. So we had some preliminary experiments that shows the Claim Spotting model, if not

carefully designed, it could be vulnerable to buyers as well. One example is, we have to factual claims which are otherwise identical they only differ by reference to two different ethnic groups. You know, one may say say the Black community, the other men say Hispanic community. And they are rather identical, the rest of the two sentence. And the model could potentially give one sentence high scores and the other sentence low scores and suggest a very different worst instance of a check. And that made, you know, if you totally rely on have a media tool for what to fact check, then you may kind of focus on fact checking since particularly related to one community by ignoring the other community.

Sophia:

Good point. We have one final question, I think we have time for one more question. What are some things people transitioning from jobs in Academia to private industry should know? What, in particular, should PhDs and MAs and postdocs in AI, data science, know thinking ahead on the future global economy? What are your suggestions, what are your predictions? What are you encouraging our students to do?

Chengkai:

So you know these are, these are all fascinating questions and very big ones. You know I'm not a- you know any authority, so let me just provide my opinion. So data science by nature is interdisciplinary discipline. So it's you know, it's supposed to be that you know, everyone, you know, regardless of your background, you need to have various level of data - they call it the data acumen. You know the you know the capability of wrestling with data, analyzing data, you're getting insights from data, and you know, making decisions based on that. Being able to recognize the potential pitfalls of those decisions, you know, such as bias, and being able to even interpret you know machine learning and AI algorithm results. Being able to communicate, you know in data science it's also important that you have a way of describing the results in a story in the context. Keep in mind the if you are data worker, you need to communicate data science-based decisions to decision makers. To your boss. And that communication is very important as well.

So my sense is that this is a really important skill that everyone should have. You know, if you're not - if you're not a computer science student, you know you don't necessarily need to be a machine learning guru know the mathematics behind it. But you need to know the basics, you need to know of what tools are available there, and you know nowadays there are a lot of open source tools, tutorials online and they are made a really accessible to everyone with you know basic college level education and you can build a very useful tools quickly by following some of these tutorials online. So data science, or you know, AI I think it's exciting fighting and it's accessible to everyone and it's something that cannot be a dodged anyone.

Sophia Johnson:

Thank you very much, on behalf of the Program for Economic Research and the Department of Economics at Columbia University, thank you, Professor Li for joining us this week. The conversation will continue online. This video will be posted in the next day or so. You can follow our keynote speaker, Professor Li, on Twitter with the handle @chengkai\_li. You can also follow Columbia Economics on social media, our handle is @Columbia\_econ.

Thank you so much, everyone, for watching, thank you for attending. We encourage you to visit our website, [www.econ.columbia.edu](http://www.econ.columbia.edu), for more information about our upcoming public events. Thank you, Professor Li.

Chengkai:

Thank you very much.

Sophia:

Bye bye.

Chengkai:

Bye. Take care.