



Meeting of October 13, 2016 - Big Data in the Workplace: Examining Implications for Equal Employment Opportunity Law - Transcript (<https://www.eeoc.gov/meetings/meeting-october-13-2016-big-data-workplace-examining-implications-equal-employment>)

PRESENT:

JENNY R. YANG, Chair

CONSTANCE S. BARKER, Commissioner

CHAI R. FELDBLUM, Commissioner

VICTORIA A. LIPNIC, Commissioner

CHARLOTTE A. BURROWS, Commissioner

ALSO PRESENT:

D. PATRICK LOPEZ, General Counsel

PEGGY R. MASTROIANNI, Legal Counsel

BERNADETTE B. WILSON, Acting Executive Officer

TABLE OF CONTENTS

AGENDA ITEM PAGE

1. Announcement of Notation Votes

Period 6/17/2016 through 10/7/2016
Approval Litigation
Approved amicus
Approved 7 Subpoena determinations
Approved Enforcement Guidance
Approved Contracts
Approved 1 & Disapproved 6 pilot projects
Approved Alternative FY 2016 funds
Approved 2016 Regulatory Agenda
Approval Final Rule for Submission to OMB
Approved Paperwork Reduction Act
Approved Strategic Enforcement Plan
Approved Resolutions

2. Big Data in the Workplace

Kelly Trindel

Eric Dunleavy

Michael Housman

Marko Mrkonich

Michal Kosinski

Ifeoma Ajunwa

Kathleen Lundquist

PROCEEDINGS

{1:03 p.m.}

CHAIR YANG: Good afternoon everyone. The meeting will now come to order. Thank you all for being here. In accordance with the Sunshine Act today, today's meeting is open to public observation of the Commission's deliberation and voting, and at this

time, I will ask Bernadette Wilson to announce any notation votes that have taken place since the last Commission meeting, Ms. Wilson.

MS. WILSON: Good afternoon, and before I begin, is there anyone in need of sign language interpreter services?

Okay, thank you. Good afternoon again Madam Chair, Commissioners, General Counsel, Legal Counsel. I'm Bernadette Wilson. I'm sorry, we do have someone in need of sign language interpreters.

Okay, good afternoon again, Madam Chair, Commissioners, General Counsel, Legal Counsel. I'm Bernadette Wilson from the Executive Secretariat. We'd like to remind our audience that questions and comments from the audience are not permitted during the meeting, and we ask that you carry on any conversations outside the meeting room, departing and re-entering as quietly as possible. Also, please take this opportunity to turn your cell phones off or to vibrate mode.

I would also like to remind the audience that in case of emergency, there are exit doors to the right and left as you exit this room. Additionally, the restrooms are down the hall to the right and left of the elevators.

During the period June 17th, 2016 through October 7th, 2016, the Commission acted on 42 items by notation vote:

Approved litigation in five cases and disapproved litigation in one case;

Approved Amicus participation in nine cases;

Approved seven subpoena determinations;

Approved an Enforcement Guidance on Retaliation and Related Issues;

Approved the following contracts:

Interagency agreement with federal occupational health for headquarters health unit services; OIT fourth quarter acquisitions; a comprehensive communications cloud-based email subscription management system; a skill soft E-learning system; and the technical assistance services EEO-1 survey;

Approved one and disapproved six pilot project requests submitted by federal agencies;

Approved alternative use of end-of-year state and local FY2016 funds;

Approved the 2016 fall regulatory agenda;

Approved a final rule for submission to the Office of Management and Budget:
Affirmative action for individuals with disabilities and federal employment;

Approved the Paperwork Reduction Act, 30 day notice proposing revisions to the
EEO-1 survey;

Approved the strategic enforcement plan for fiscal years 2017 through 2021; and

Approved resolutions honoring Valerie Baxter and Lisa M. Williams on their
retirement. Madam Chair.

CHAIR YANG: Thank you Ms. Wilson, and thank you all for joining us today for this
important meeting on big data in the workplace, examining implications for equal
employment opportunity law.

Before we get started, I wanted to take a brief moment to thank our General
Counsel, David Lopez for his tremendous service to the EEOC. As some of you may
know, he has recently announced that he will be leaving the EEOC in December and
we just wanted to thank you David, for all your incredible leadership. {Applause}

And I would like to thank all of our witnesses for the time that you took into
preparing your testimony and coming today. We are expecting our -- Commissioner
Barker is on her way, and I know we will -- are expecting our last witness, as well,
but we will get started and I'd like -- to acknowledge Commissioner Lipnic, whose
office helped organize today's meeting, and in fact, today's meeting digs deeper
into a question that Commissioner Lipnic asked, at a Commission meeting for our
50th anniversary in July 2015, and she asked one of our witnesses to explain how a
friend of hers may have come to receive a rejection notice for a job he applied for
online, a mere 28 minutes after he hit submit, and this was at midnight.

So, at that Commission meeting, we looked back at how far we have come over the
past 50 years, and we also looked ahead at the emerging issues for the 21st century
workplace, and we had a witness testify about the increasing use of big data in
selection procedures and today, we are looking forward to taking a closer look at
the kinds of data and technology that employers have been exploring and using to
improve their selection procedures.

We are interested in learning more about how vendors and employers are taking the vast amounts of data that are being generated every day, and converting that data into analytics to make employment related decisions. In today's meeting, we're calling this concept big data; but it is also closely associated with concepts such as people analytics and artificial intelligence. Today's meeting will explore the ways that this data is generated; sometimes it's generated through work, through programs that track a worker's movements, productivity and even interactions with colleagues, or an automated system that can code video interviews with a candidate. Other data is generated through public records or through our role as consumers, by social media activity, online browsing and mobile apps, for example.

So, with all this data now available, there are entire industries emerging that seek to aggregate and translate this data to assist them in predicting how people might perform. Employers are understandably eager to learn how data may assist them in making better employment decisions in hiring, in performance evaluations, in promotions. And our witnesses today will explain how some employers have begun to use big data, and the important equal employment opportunity considerations that these methods can raise.

Big data has the potential to drive innovations that reduce bias in employment decisions, and help employers utilize untapped talent. At the same time, it is critical that these tools are designed to promote fairness in opportunity, so that relying on these expanding sources of data do not create new barriers to opportunity. At the EEOC we know that big data is fundamentally changing the way selection decisions are made. As we consider how to best apply our anti-discrimination protections to this reality, we are mindful of the value in promoting innovation, while at the same time, recognizing that it is critical to ensure that reliance on these vast sources of data do not create new barriers to opportunity.

I am very interested in hearing from our witnesses today, about the ways in which these algorithms rely on correlations and the risks this can create of screening for qualities that may not be accurate predictors of job performance, and because the data and algorithms rely on an extraordinary array of data points that can be quite opaque, it can be difficult to identify and correct these problems.

Today's meeting also has important connections with the recent Commission meeting we had on diversity in the tech sector. In that meeting, witnesses highlighted how hiring practices at tech firms and funding decisions at venture capital firms had a tendency to simply replicate the existing model of a successful

tech entrepreneur, based on people's ideas of what a successful tech entrepreneur looked like. These tendencies can continue patterns of hiring that limit the diversity of talent employed at these firms, and that is related to today's meeting in two ways:

Big data can similarly have the tendency to simply replicate the existing model of success, and if the people designing the algorithms aren't aware of the different possible outcomes by race, gender, age and other groups; the algorithms themselves may get certain biases built in. I am optimistic that today's meeting will provide helpful recommendations on how big data can be used, in a way that advances equal employment opportunity, and I look forward to hearing from you all about your testimony, and in addition to all of our witnesses, I'd like to thank a number of EEOC staff, including Bernadette Wilson, Yolanda Bond, Peach Soltis, Donald McIntosh, Jim Paretti and Aaron Konopasky for their contributions to today's meeting.

I would now like to invite my fellow Commissioners to make any opening statements or comments, and I would like to turn this over now to Commissioner Lipnic.

COMMISSIONER LIPNIC: Thank you Madam Chair. Good afternoon everyone. First, I want to extend my appreciation to you, Madam Chair and your staff, for working so closely with my office to organize this meeting on this important topic, and in particular, I want to thank Peach Soltis from your office, who put so much time and effort into this. We have an excellent panel of witnesses today, and we have someone phoning in. Do we? Are we good?

Welcome to each of you and thank you, as the Chair said, for all the time and effort you've put into your testimony, and for being with us today. If you believe recent headlines, it's now a big data world, and perhaps one of the things we will resolve at this meeting is whether it is data or data. We're just living in it. Talk of big data seems to be everywhere, and there is seemingly big data in everything we do, advertising, marketing, sports and entertainment, the medical and healthcare worlds, politics, finance, insurance, housing, education, security, policing and yes, in the workplace and employment. You name it, and data is, we are told, being collected and analyzed like never before, and while we may only be in the first generation of big data, or maybe we're already at Version 2.0, whichever it is, the growth has been break-neck, especially in the business world.

Earlier this year, the Wall Street Journal noted a study showing that the number of Fortune 1000 U.S. companies, using some form of big data, had jumped 58 percentage points over the past three years, to 63 percent. Seventy percent of the companies surveyed said that big data is important to their firms, up from 21 percent in just -- in just a short time ago as 2012. So the era of big data is clearly upon us, and yet for most right now, this is all happening behind closed doors. We may see it in action from time to time, spotting an ad in Google or maybe a product recommendation in Amazon, that's clearly targeted at us; but it seems strange to me that at this stage, there is very little real understanding of where and how big data is being used for the layperson. Chair Yang mentioned the question I had asked at a hearing a while ago, about my friend receiving a rejection, 20 minutes after applying at midnight, and another friend of mine, as he has applied online to various jobs and has actually gotten -- moved onto a next round, when it has resulted in that, refers to it that -- refers to that as, "I made it through the commbomulator," whatever the commbomulator is. So, who knows how big data is -- how it's being used and will be hopefully, defining and discussing all of that today.

Whatever the case, I think that story illustrates that for the most part; there is an opaqueness at the heart of big data, which is where I hope we will spend a good bit of our time this afternoon. I hope we can start to lift the veil on some of this as a practical matter, and begin the process of understanding exactly where and how it is being used in the workplace, whether in hiring, retention, performance evaluations and so on. I also hope we can talk about how big data tools are being developed and marketed by third party vendors.

Newer technologies, such as data analytics can pose significant challenges to an enforcement agency like ours. We tackled similar issues at our meeting on social media, although certainly those technologies are more apparent and accessible to all of us. There is the constant race to chase and develop a technical institution-wide understanding of the next big thing.

On top of that, there is determining whether, when and how our laws may apply in our increasingly technology-driven workplaces. Sometimes this is relatively simple, in the case of long-standing issues merely addressed in new or updated packaging. Other times it can get more challenging, when new technologies are laid on top of laws that may have been devised against, as in our case, the backdrop of the 20th century workplace. Challenging, yes, but as I see it, that is the core of our responsibilities; ensuring that our understanding of today's workplaces and that our

interpretation and administration of the law are as current and fully informed as possible. It's for that reason; I think meetings like today and the discussion they spur, are critical. So thank you again Madam Chair, for the opportunity to work on this. I look forward to the testimony and our discussion to follow. Thank you.

CHAIR YANG: Thank you Commissioner Lipnic. Commissioner Feldblum.

COMMISSIONER FELDBLUM: Thank you. So again, thank you Madam Chair for convening this very important meeting and to the various staff you mentioned, including Peach Soltis, who is sitting right behind me, those of you who want to know, and to Commissioner Lipnic and her staff, Donald McIntosh and Jim Paretto, who is sitting all the way to your left. Commissioner Lipnic has been talking to me about her questions about big data, and I'm coming out as a data versus data person, for six years. So, I am very glad that we're able to have this meeting.

I also want to add my thanks and kudos to our General Counsel, David Lopez, who has maintained a truly incredible level of energy over the past six and a half years, and as you all know, we enforce civil rights laws on behalf of people who have been discriminated based on race, religion, national origin, sex, including pregnancy, sexual orientation and gender identity, age and disability, and General Counsel Lopez has made an impact for each of those areas.

I also want to extend my thanks and kudos to Director Pat Chu, who is here, and she can raise her hand, if she wants. She is the -- has been the Director of Office of Federal Contract Compliance Programs at Department of Labor and one of our really important partners, and she is also leaving, and I just want to extend my personal thanks and kudos to all her work.

I think it's clear from the testimony that the data we are dealing with now is not only not our parents' data, it's not even our data from our lifetime. I think for many of us, certainly for me, I think about data as demographic data, and that's very important for civil rights work. There's demographic data that we collect through our EEO-1 forms. Every employer with more than 100 employees is telling us their demographic data, based on race, sex. We collect very important information from the federal agencies, not just in those categories, but also on disability, based on our Management Directive 715. This is important data to understand what the baseline is, and then how are we doing in terms of diversity.

But big data is clearly different because it is combining forms of data from lots of different places, to then as one person's testimony said, to really - well all really, predict behavior, and that's often behavior in the workplace. I think there are some people who would love to be able to predict my behavior, but I actually don't think that that's necessarily something that can go that individualized, or maybe it can, and that's something we will hear from you.

So, I want to say, I want to thank all the witnesses, not only for the time you put together in putting together really, just excellent testimony from all of you, but also the time you're going to take to engage with us in conversation to help educate us. So, thank you.

CHAIR YANG: Thank you. Commissioner Burrows.

COMMISSIONER BURROWS: Good afternoon. I join my colleagues in welcoming our distinguished panel. I am very excited to hear from all of you. I also commend the Chair for convening this important hearing, and thank her staff and Commissioner Lipnic's staff for all the work that they have clearly put into this.

I want to note, I guess, that as others have said, the collection and use of data have become increasingly commonplace in so many areas of our lives, including employment. Today, companies have access to data bases that contain a wide variety of information on applicants and employees, including data on their educational background, credit and criminal background history, internet browsing history and purchasing habits. Companies may also view an individual's personal information on social media platforms, such as Facebook and Twitter. As the Washington Post reported just yesterday, so-called personalized learning computer programs developed by private companies are increasingly being used in schools for children as young as five; that collect a vast amount of personal data about these children from grades, their academic performance, in some instances, discipline, disabilities, family relationships and socio-economic status; and in many cases, how this educational information is safeguarded, how it will be retained and for how long, and who might eventually have access to it is largely unclear. And not surprisingly in recent years, there has been growing interest in harnessing the potential of big data in a host of different ways, including to help companies make better employment decisions, from recruiting and retaining top talent, to training and employee engagement. While it's possible that analyzing large data sets using complex algorithms may reduce discrimination and increase workplace diversity, by eliminating some of the subjectivity in hiring; big data may also have the opposite

effect, if the algorithms rely heavily on correlations that reflect cultural similarities, rather than differences in skills or abilities.

I'm especially concerned that inconsistencies or mistakes in the data, such as errors in credit reports or flawed assumptions that are built into algorithms used to analyze data may simply serve to compound discrimination and lack of diversity in a company or an industry. In short, the algorithms used to assess big data are only as good as the assumptions that underlie them. Because the use of big data in employment is just beginning, this is an ideal time for the Commission to take a close look at this issue and see whether there are concrete suggestions that can help employers and employees benefit from this new technology and the potential of it, while still advancing equal employment opportunity. In the half century since it was established, this Commission has played a vital role in promoting equal opportunity and fairness for everyone in the workplace, and I am pleased that the Commission is continuing to do so by examining emerging issues for the 21st century workplace.

I look forward again, to hearing from our distinguished witnesses about some of the many challenges and opportunities presented by big data, and the lessons we can draw from your experiences. Thank you. I yield back my time.

CHAIR YANG: Thank you, and now, I'll turn it over to Commissioner Barker.

COMMISSIONER BARKER: Thank you very much. I want to thank the panel in advance. I've read your testimony and I'll have to say, I learned a great deal from it. This is such a new issue for all of us, and I know that we'll all keep an open mind, as to whether or not the issues that the collection and the use of big data presents is something that this Commission needs to weigh in on at this point, or whether we know enough to determine whether or not this is something that raises genuine discrimination issues, such that we should get involved. Thank you.

CHAIR YANG: Thank you for those opening remarks, and now I am pleased to introduce our distinguished panel today.

We have first Dr. Kelly Trindel. She is Chief Analyst in the Office of Research, Information and Planning at the EEOC, where she has been providing analytic support for systemic investigations in case development since 2010. We have Dr. Eric Dunleavy, who is the Director of Personnel Selection and Litigation Support Services at DCI Consulting, and is appearing today on behalf of the Society of

Human Resources Management. Dr. Michael Housman is the workforce scientist in residence at hiQ Labs. He received his PhD in applied economics and managerial science at the Whorten School and his AB from Harvard University. Joining us today via VTC is Dr. Michal Kosinski, an Assistant Professor of Organizational Behavior at Stanford Graduate School of Business. He received his PhD in psychology from the University of Cambridge. Marko Mrkonich is a shareholder at Littler Mendelson PC, where he focuses his practice on representing management in labor and employment matters, and is one of the leaders of Littler's Big Data Initiative. Dr. Ifeoma Ajunwa holds a PhD in sociology from Columbia University and is a fellow at the Berkman Klein Center at Harvard and a Law Professor at the University of the District of Columbia Law School. Dr. Kathleen Lundquist is a nationally recognized organizational psychologist who testifies frequently as an expert witness in employment discrimination class action lawsuits, and is President and CEO of ABT Metrics, Incorporated. Thank you all.

Today's panel will consist of one panel and then we will open the floor for questions and comments from members of the Commission. Each panelist will have six minutes to make oral presentations today, but your complete written statements will be available on our website www.eeoc.gov and placed in the meeting record. Please note that we are using the timing lights at the center of the console in front of me. The yellow light will appear when you have one minute remaining for your testimony, and the red light will appear when your allotted time expires.

Commissioners' questions and comments will begin after all of you have completed your opening statements. Again, we are very pleased to have such an exceptional panel of witnesses today. Thank you for being here, and we will begin with Dr. Trindel.

DR. TRINDEL: Thank you Chair Yang. Chair Yang and distinguished Commissioners, thank you for the opportunity to speak with you today on the use of big data in employment. I'm really honored and excited to be here to talk about this.

I think it's important to first define our terms. In the employment context, I would define big data as the combination of non-traditional and traditional employment data, with technology enabled analytics to create processes for identifying, recruiting, segmenting and scoring job candidates and employees. This field is sometimes referred to as people analytics. So, non-traditional employment data comes from places like financial and consumer data systems, public records, social media activity logs, sensors, internet browsing history, mobile devices and

communications meta data systems. And this list is by no means complete. Every day it grows. In fact, the very words that I'm speaking today could potentially be thought of as data. As the Chair mentioned, they will be published at eeoc.gov and thus, made public. These words can then be scraped from the website, tagged, coded, classified and organized into a matrix, which will then be available for analysis.

When this type of non-traditional data is quantified and merged with traditional employment data; it can be used to look backwards in time at -- from worker outcomes to a broad swath of data points representing previous patterns and behaviors.

Generally speaking, big data algorithm is given a training data set, containing information about a group of people, typically current or former employees, from which it uncovers characteristics that can be correlated with some measure of job success. The successful profile can then be used in a number of ways, including seeking out passive candidates, screening active job applicants, or allocating training resources or incentives for current employees. Job success can be operationally defined in multiple ways, such as job performance ratings, quantified worker output, tenure, culture fit, turnover or churn, absenteeism or pre-disposition to safety incidents. People analytics seems to have naturally developed from other areas of the business that have undergone their own analytics revolution, such as marketing. It was somewhat inevitable that this type of spillover would occur, particularly when human resources departments have access to so much information about workers, and can use it to calculate return on investment metrics.

Marketing products to the right people at the right time has morphed into marketing job opportunities to the right people at the right time, for example. Why not optimize hiring and talent management in the same way that we've optimized these other areas of the business? Employers must bear in mind however, that these algorithms are based solely on the training data that's fed to them and the definition of success that's input by the programmer. For example, if the training phase for a big data algorithm happened to identify a greater pattern of absences for a group of people with disabilities, it might cluster the relevant people together in what's called a high absenteeism risk profile. This profile need not be tagged as disability; rather, it might appear to be based upon some common group of financial, consumer or social media behaviors. It may not be obvious to the

employer, or even to the data scientist who created the algorithm, that subsequent employment decisions based on this model could discriminate against people with disabilities.

Similarly, if most previously successful employees at a firm happen to be young, white or Asian American men, then the model will codify success in this way. If married women of a particular age are more likely to churn, then the model will codify this in proxies and subsequently predict lower success rates for similar women. All of this can happen without informing the model of disability status, age, race or gender, and all while giving the appearance that the machine is working just as it should, to increase worker ROI.

I hope that the issues raised in today's meeting will serve as an important reminder to vendors and employees that -- sorry, vendors and employers, that they must be mindful about adopting these strategies, especially given that many of the people who build and maintain these algorithms may not be familiar with equal employment opportunity law. Data scientists transitioning from marketing, for example, may lack the regulatory and legal background to properly consider EEO compliance, and perhaps more importantly, these data science may -- scientists may limit -- may be limited in their availability. I'm sorry, they may limit their available labor pools unnecessarily by using irrelevant criteria. Employers must be warned not to simply trust the math in this case, as the math may very well lead them astray. Thank you again for inviting me to speak today.

CHAIR YANG: Thank you Dr. Trindel, and now, Dr. Dunleavy.

DR. DUNLEAVY: Good afternoon. I'd like to thank the Chair, the Commissioners and the Society for Human Resource Management, for providing me the opportunity to briefly discuss the current state of big data tools in HR practice, with particular focus on employment. I present my perspective today as a SHRM member, as an I/O psychologist, and as a scientist/practitioner who has conducted applied research and consulted for a wide range of employers across a variety of industries. I have no vested interest in whether employers use big data tools in employment. My purpose in testifying today is to provide some contemporary perspective on what we currently know and don't about the use of big data tools in employment.

As other panelists have already shared, this is a complicated topic, and there is still little research on big data tools in the context of employment or EEO. I organize the next five minutes around a few key questions. First, what do we mean by big data in

employment? To simplify, big data tools related to employment typically pair large and involving organizational data sets from various sources of information with machine-learning methods, which are really approaches to combining that information via evaluating existing data to identify rules that maximize the prediction of an outcome, maybe it's performance, maybe it's turnover or absenteeism or to better conceptualize that outcome. So, there is the content and then there is the means to combine that content to make decisions and potentially achieve organizational goals.

What questions are employers and HR practitioners trying to answer with big data? Many questions employers are trying to answer via big data are not new. Employers and HR practitioners have long been asking questions related to whether the future behavior of applicants and employees can be predicted from available information on hand. For decades we've evaluated factors like bio-data, which refer to a wide variety of work and life history data points to predict outcomes like performance, absenteeism, employee attitudes and turnover.

So how are the data being used? In my experience with clients and colleagues, I am most familiar with these tools in the context of recruitment and hiring. Online recruiting in the form of external job boards, social media and internal career platforms may be very useful strategies for identifying talent. Applicants are easier to reach today, and can apply to many jobs from anywhere. In part because of this situation, automated steps at the front end of a hiring process may be particularly useful, given the large size of internet-based applicant pools and the human capital effort required to evaluate those applicants on eligibility and qualifications. The tools themselves may take on a variety of forms. Some may look and feel like application blanks or resume reviews. Others may look and feel like broader bio-data evaluations, and still others may look like interviews. Some may not fit neatly into a traditional category. The sources of applicant data that may be evaluated can vary from narrow self-report information submitted as part of a resume, to much broader personal data that is not self-volunteered and instead is mined from social media sites or other public information.

In other scenarios, video game play may be used to collect content on applicant information, or video interviews may be recorded and evaluated for verbal and non-verbal information. Data metrics on employees may also be available in organizational systems and leveraged to evaluate performance, which could then be used to make decisions related to promotion, placement or compensation. I suggest

that readers review the SHRM Foundation report entitled, "Use of Workforce Analytics for Competitive Advantage". There are a number of interesting examples there.

So what do we know about contemporary use of big data? At the current time, the answer is not much. SHRM recently conducted a survey of 279 members, to better understand this issue, and reported that about one-third used big data to support HR. Those in larger organizations were much more likely to do this. As such, it will be interesting to monitor the results of similar surveys in recent years -- in future years.

So what are some EEO issues worth thinking about? SHRM and the broader employer community look forward to what legal scholars have to say on this issue. A few practical matters conclude my testimony. With regard to intentional discrimination, big data concerns may focus on whether direct protected group status membership is incorporated in some way in the algorithm used to evaluate content. If group membership is not included in the algorithm, and human judgment decisions on the front end did not directly consider protected group status in deciding what content to evaluate, what outcomes to prioritize or what algorithms to use, then intentional discrimination, as a function of the algorithm may not be an issue of concern.

With regard to disparate impact, big data tools seem to fit with the traditional Title VII scenario and a uniform guideline style analysis to scrutiny. Is there substantial adverse impact? If yes, is there evidence of job-relatedness? If yes, were there less adverse and equally job-related alternatives? A number of additional examples -- complexities may need to be considered, for example, ambiguity of what's being measured. The scope of data inputs and the complexity of the machine-learning may make it hard to understand what's being measured and how decisions are made. Algorithms may be constantly changing as well, which may make isolating a specific procedure a challenge. There may be ethical and security issues, such as those involving confidentiality and privacy, and some content inputs may, on their face, violate particular laws.

In summary, big data methods are being used in the employment setting and come with potential risks and rewards. The prevalence of these methods is still unclear, and I reiterate that I've seen little empirical research related to the validity and adverse impact with these tools. It may be prudent for big data tool users to

consider the purpose of the tool, what's being measured, what research is available to support validation and use, and whether sub-group differences exist. Thank you.

CHAIR YANG: Thank you Dr. Dunleavy, and now, Dr. Housman.

DR. HOUSMAN: Good afternoon Chair Yang and members of the Commission. Thank you for the opportunity to testify. My name is Michael Housman. I'm trained as an economist. I am the Workforce Scientist in Residence at hiQ Labs. At hiQ Labs, I translate publicly-available data into insights that allow large employers to identify employees that are potential flight risks, and take actions to help retain them and improve overall workforce performance. Prior to hiQ, I worked for a company called Evolv that used predictive analytics to help large employers make better hiring decisions. So I've seen how data and algorithms are being used at each end of the employee life cycle.

The recent trend towards the use of big data in our everyday lives has become its own narrative in the news, and even the White House has weighed in. The Obama Administration has released a series of reports on big data and its implications in education, health, advertising, criminal justice and the economy. The latest report included a particular focus on the civil rights opportunities and challenges of the technology's potential use in the employment context, particularly in recruitment. Yes, it's true that large employers are turning increasingly towards computer algorithms to determine who is and is not a good fit for the job. Although the results consistently suggest that these "robots recruiters" are effective at helping hiring managers to choose employees that stay longer on the job and perform better; there is still some skepticism as to whether computers can replace human judgment, when it comes to evaluating talent and its potential to discriminate.

What I think is important to recognize first is that the current system isn't perfect and recruiters aren't unbiased. In fact, there's a long line of research that documents empirically the existence of a 'like me' bias that leads recruiters to hire applicants like themselves. This may benefit job applicants who happened to have gone to the same school as the interviewer, but unfortunately, it tends to hurt anyone who didn't. The inevitable outcome of this bias is that the most talented or skillful individual may not automatically get selected for the job, but rather, the applicant that the recruiter likes the most. Not only that, but there's the possibility that hiring like-minded individuals tends to reduce diversity in the workforce. Contrast this with the algorithms that have been built to select the best applicants. These algorithms are designed to make assessment decisions based on the factors

that actually matter, and have been correlated statistically with on-the-job performance and outcomes. When they've been calibrated by data science teams that are monitoring the right metrics, they can be engineered to ensure that they do not have an adverse impact on groups protected by gender, race and age.

At Evolv, there was a slate of tests that we would apply to any of our scoring algorithms before we deployed them, knowing that one's performance on the assessment was basically uncorrelated with their gender, race or age. In doing so, we trained the algorithms to select the most qualified applicants, and to ignore the fact that he or she went to Harvard and played squash. The data support this claim. In fact, a white paper was published in November 2015 by the National Bureau of Economic Research, by researchers at the University of Toronto, Yale and Northwestern, that analyzes hundreds of thousands of hires, and finds that the adoption of job testing is associated with a 20 percent reduction in quitting behavior. If anything, I believe it's more likely that online assessments reduce bias in the hiring process.

Consider the fact that recruiters typically spend approximately seven seconds screening each resume, and what do they look for? Among other things they look for are previous work history, job-relevant experience. At Evolv, where I served as chief analytics officer, we released studies demonstrating conclusively that job hoppers and long-term unemployed stay just as long and perform just as well as individuals with a more typical work history. In fact, the White House released a report in October 2014 about the long-term unemployed, in which they cited Evolv's work with companies like eBay, AT&T and Xerox as helping them get the long-term unemployed back to work. These are factors that shouldn't play a role in the screening process, yet two to six percent of all job applicants are dismissed immediately because of a less than ideal work history. Pre-hire screening reduces personal biases by allowing job hoppers and the long-term unemployed to be considered on the basis of their true knowledge, skills and abilities.

The fact that computers are playing a bigger role in the hiring process causes some trepidation, understandably, but it's important to realize that these algorithms aren't meant to replace recruiters. They're simply meant to arm recruiters with more information, with which they can use to make a more informed decision. It's an exciting era, not only because the technology is capable of issuing recommendations around something as complicated as hiring, but also because

this capability is going to give a fair shot to millions of job applicants who wouldn't have been considered previously. Thank you very much.

CHAIR YANG: Thank you Dr. Housman, and now Dr. Mrkonich. Did I say doctor, as well? We have an unusually high number of doctors. I think you might be our only one not on the panel. So, thank you for being the lawyer, the JD doctor.

MR. MRKONICH : Well, and have you ever felt like the world was a tuxedo and you're a pair of brown shoes? I thank you Commissioner Yang, and the other Commissioners, staff, counsel, and frankly, the interest shown by the audience sitting behind me. It is a fascinating topic. It's very interesting and we at Littler appreciate being included on the panel, and having a chance to speak from the trenches, so to speak, working with employers who spend millions of dollars each year to comply with EEO laws, achieve a diverse workforce that is both responsible and effective.

Big data has revolutionized everything we do. We've heard that from my fellow panelists. I think we'll hear it from everyone. It's unrealistic to think it's going to affect everything we do, except our jobs, given how important our jobs are to all of us. And from the trenches, I can tell you it's starting to make what I think of as a positive difference already in some of the ways mentioned by Dr. Housman, by expanding the applicant pool, in some cases, beyond those who even apply. It's eliminating the application barrier to permit broader consideration.

Now, I've submitted some written remarks to the Commission, and they're intended to shed some light on how a practitioner views this, both from a counseling and a litigation perspective, on behalf of management, and I'm fully prepared to respond to questions about those things.

As I go forward with my few minutes today though, what I'd like to do is just highlight a couple of the concerns and a couple of the opportunities I think, that make this a key crossroads time for employers, for the Commission and for those concerned about fairness in the workplace and achieving an ideal workplace. First, I'd like to reiterate, as everyone has, that big data analytics work, and that's why we're here. They're not going away. Employers that more effectively use the information they currently possess are able to out-perform their competitors. The chief knowledge officer at Hewlett Packard was once quoted as saying, "If Hewlett Packard knew what Hewlett Packard knows, it could make some money," and I think that's where employers have been. On one hand, they have a pile of resumes and

applications with much information. On the other hand, they have a workforce that they've watched perform and do well, and if they could match those two up, they could improve their hiring process in a way that made everyone more effective; and of course, we now diagnose flu epidemics not based on how often people go to doctors' offices and tabulating those reports, we diagnose our flu epidemics by measuring Google searches by zip code, and when certain searches become prevalent in a zip code, that means that's where the flu has arrived.

So, big data is not going to go away. It's not going to leave us out of the workforce. Second, big data analytics are not well understood by many employers, employees or regulators or courts yet, because frankly, the field is emerging and evolving. It's the combination of massive amounts of data, new methods of analytics, but it's not data alone and it's not algorithms alone. It's also people who oversee some of those things. Now, people aren't sitting back, looking at a data field and determining what the algorithm looks like. I have a son who is a PhD student in economics, who will be a doctor someday, I hope, and when he hits 'enter' on his computer and the algorithm starts to run, it's four hours later, before a formula emerges. If he spent his own brain trying to do that, he would spend the rest of his life, and he wouldn't finish what the computer did in three to four hours.

So, I think of it as a combination of technology and people. The tests have suggested right now that if you have medical symptoms and you're diagnosed by a doctor, you achieve a certain level of success. If you're diagnosed by a computer, it's more successful. But if you're diagnosed by a doctor working with a computer, it's most successful. I think that's the model that applies in the employment sector, as well.

Third, I think big data used correctly eliminates, reduces, if not eliminates discrimination in many of its most egregious forms. It eliminates bias; it reduces bias; it opens opportunity for people. For example, people will often suggest, employers will say, "Why isn't your applicant pool more broad?" Well, if you're using some big data tools, you are able to gain access to the entire work potential workforce, not just those who figure out that you're looking to hire on that particular day. So, it's a way to reach out. For other employers, it's a way to eliminate the subjective, express, implicit, whichever language we choose to use on any given day; reduce that bias in the decision making and the process by which people are selected, promoted and managed.

A fourth concern sometimes raised is the private interest, the privacy interest and those sorts of things. Two things come into play, I think here in a major way, that we're seeing. One, it's the anonymization. That's a word that's sometimes hard for me to say. The anonymization of the data makes it very safe and secure to use. So, what big data looks for isn't what this particular person is going to be doing or where they're walking on a particular day. That tends to be how we think about life. Big data says, what happens when people work together? One of Dr. Housman's formulas showed for example, I believe that people who eat lunch in a big group are more productive in the afternoon, as opposed to people who ate lunch alone or with one other person. So, the company redesigned its lunch room and had a more productive and successful operation.

From the perspective of practitioner, the current discrimination models don't work in a world based on correlation instead of causation. They don't work the same way. I don't think they always work at all. So, the concept of disparate impact, it grew up in a world where we had job descriptions with criteria we could measure and look at; and then we said with those skills, what are the measurement tools that we're going to use to see if someone has them? That -- you throw that out the window when you hit 'send' or 'enter' on the computer, and that decides what the algorithm looks like. We need new tools in the industry and the courts, and all of us need a dialogue. We need to shed some light on this, before we get down to the dialogue of determining what that standard might be. The challenge for all of us is to permit this development to take place, without artificially interfering. It's a good thing and it's here. Thank you very much.

CHAIR YANG: Thank you Mr. Mrkonich, and now, I'd like to turn to our VTC, where we have Dr. Kosinski.

DR. KOSINSKI: Madam Chair and Commissioners. Thank you for the opportunity to testify. I believe that if used properly, big data and modern computational techniques can not only reduce the discrimination in the workplace, but also improve person-job fit, boost the competitiveness and productivity of our economy and the well-being of the employees. Importantly, recruitment, career planning and performance appraisals, based on big data and computational models will disproportionately, I believe, benefit groups that are currently the most discriminated against, such as women, members of racial minorities and those who lack access to education, mentors and role models.

A quickly growing body of research shows that such digital-footprints can be used to accurately predict future behavior. Real-life outcomes and psychological traits, such as career choices, intelligence, personality, happiness and job performance. Such models are remarkably accurate in almost all cases that I worked with, much more accurate than human judges and way less biased. I believe that the computational models can help in reducing discrimination. To do so however, they need to be used properly. Fortunately, we have decades of experience in doing so, that we gained from using traditional tests and questionnaires. Computational models differ from traditional questionnaires or interview-based assessment tools, but they should be subject to the same principles of fair, valid and accurate assessment. One of such principles often overlooked in the discussion of computational models states that only the factors causally linked with performance in a given job can be employed in the ranking of the candidates, while hiring software engineers for instance, one can rank them based on their conscientiousness or cognitive abilities, but not based on the gender. The same principle has to be applied to computational models. As such, models are usually so complex that it's impossible to fully comprehend their functioning; special steps have to be taken to ascertain that they do not discriminate against candidates based on factors that are unrelated to job performance. Models that directly predict performance in a given job are prone to such biases, because they can recreate those that are present in the current workplaces, and they will get it from training data.

To avoid this issue and retain full control over the factors used in ranking the candidates, the model should not be aimed directly at job performance, but at the well-defined factors causally linked with job performance, such as personality, ability or skills.

Let me now move on to discuss the areas where big data and computational tools can offer great advantages, so, first recruitment. Bias in recruitment, as it was mentioned today before, is one of the main sources of discrimination in the workplace. Traditional methods using recruitment such as non-structured interviews and rating resumes or CVs were shown over and over again to be poor predictors of candidates' performance and are heavily affected by number of biases. Many of such biases could be avoided if unreliable and bias tools, such as non-structured interviews were replaced with objective psychometric tools, such as tests or structured interviews. This would not only reduce the discrimination, but also improve the efficiency of people in the workplace. However, objective psychometric tools such as structured interviews and standardized ability tests require expertise

and are expensive to purchase, develop and time-consuming to administer. Thus, they are rarely used when recruiting for entry-level jobs, and are virtually absent from small organizations. As a result, a large part of the workforce is deprived of the benefits of objective assessment. Computational models provide an inexpensive and powerful alternative to traditional occupational assessment methods. While the cost required to develop a traditional psychometric measure can reach millions of dollars; computational models can be developed quickly at a fraction of these costs. Once developed, the marginal cost of assessing an individual is comparable with the marginal cost of conducting a Google search and is as fast. By decreasing the cost of assessment, computational models can enable any candidate to be judged based on their objectively measured potential and not on a subjective and avoidably -- unavoidably biased opinion of a recruiter.

The same really applies to performance appraisals. They are also subject to the same biases as the recruitment process is, and big data assessment can again, bring us great benefit here. The analysis of digital footprints generated in occupational context provides great alternative to, you know, peer rating or manager rating. As I mentioned before, such models should not be trained on past year based ratings, as they will carry on and recreate the biases, but should be based on objective measures that are important for a given job, objective performance indicators such as the quality of the written text or the amount and quality of interactions with other team members. Those factors were really difficult to observe in the past. Now, as we live in so much digital footprint, machines can be used to track such behaviors and measure them. Also, one other important advantage of computational models is that it allows us to identify talent, rather than skills.

Now, factors such as individual socio-economic status, geographical location and so on create large inequalities in access to education and vocational training. Thus, skills and knowledge are unevenly distributed across genders, geographical areas and ethnic groups translating into self-perpetuating barriers to jobs and careers. Digital footprint models can be used to reveal late and psychological traits and predict future behavior. This is similar to what can be obtained using traditional ability and personality tests, that can also be used to measure talent as opposed to skills or a potential as opposed to skills, but it can be achieved much more inexpensively. In other words, such models can measure an individual's ability and potential rather than skills and knowledge, and it can do it at a much lower price. Also, culture norm stereotypes and the lack of role models discourage members of some groups from seeking employment or training opportunities in certain

professions in the first place, even if they do not lack in talent and ability. This problem could be confronted by using behavioral and demographic-targeting tools that are available across many online platforms such as Facebook or Google. Employers and educational institutions can use such platforms to reach out to under-represented groups with offers of training or employment. Moreover, this approach could be expanded by employing the passive assessment approach to make such offers more specific. Passive assessment employs the computational models to identify digital footprints that best predict a given characteristic like conscientiousness, for instance. Knowing what digital footprints are typical of a person likely to obtain a particular characteristic can be used to target them in the digital space. For example, if visiting a particular set of websites has been shown to be linked with the talent for computer programming, one can target the visitors of these websites with adverts encouraging them to sign up for a programming course or to apply for a given position.

Finally, meaning big data of digital footprints can provide policy makers with broad and instant information about the state of the job market, the distribution of the skills and potential sources of outcomes and outcomes of discrimination. For example, participation in online courses, visited websites and questions asked and answered in online forums can all be tracked live and then can be used to track the distribution of skills and knowledge across demographic groups. This cannot only inform policy maker, but can also be used to measure dynamically, the outcomes of interventions or additional training provided to certain groups.

So, to summarize, I believe that the digital footprint models offer the potential to reduce discrimination in the workplace. However, like many other new technologies, such models are not without risks. If used improperly or without -- or with malicious intent, they can perpetuate and amplify discrimination, rather than reduce it. Additionally, such models pose particular severe risks to privacy, as they could also be used to infer intimate traits, such as political views, religion or even sexual orientation. I would like to encourage the Commission to facilitate the development and publication of a set of guidelines for the development of computational models that are accurate, valid and free from bias. This would encourage the development and adoption of computational models and could have a strong and positive impact on reducing discrimination in the workplace. Thank you.

CHAIR YANG: Thank you very much. Now, Dr. Ajunwa.

DR. AJUNWA: Good afternoon Chair Yang and members of the Commission. I would like to thank the Commission for inviting me to this meeting on big data in the workplace. My testimony today is drawn from several papers that I have co-authored or authored regarding worker privacy and genetic discrimination. Going beyond the applicant experience that my co-panelists have presented, today I will summarize a number of practices that employers have begun to deploy, to collect information about employees, and my concerns that such information collection could invite employment discrimination and the invasion of worker privacy.

Absent careful safeguards, the data collection practices I detail could allow for demographic information and sensitive health and genetic information to be incorporated in big data analytics, which in turn, influence employment decisions thereby, challenging the spirit of anti-discrimination laws such as Title VII, the Americans with Disabilities Act and the Genetic Information Non-Discrimination Act. And furthermore, we should not envision a workplace where workers are called upon to trade all privacy rights for the benefit of employment. The monitoring of workers is not a new phenomenon. Workplace surveillance has become a fact of life for the American worker. What is of new legal concern is that with technological advancements and lowered costs, employee surveillance now occurs both inside and outside the workplace, impacting the equal access to employment for all Americans.

Two novel users of technology in the workplace provide particular cause for concern: The first is a new dependence on productivity applications and the second is the use of wearable fitness trackers, as part of workplace wellness programs. Productivity apps, which now represent an \$11 billion industry, have been touted as technology that will lead to greater efficiency in the workplace. However, given recent cases such as the California case in which a productivity app continued to track a worker in her off-hours; there remains the issue of whether there is a power-asymmetry about such apps that negate consent and whether the invasive nature of such apps could permanently erode worker privacy and provide opportunities for worker discrimination. The prominence of productivity apps countered that the apps promote efficiency, innovation and fairness in the workplace. Such discourse, while failing to consider privacy implications also neglects to take into account the sociological experience of productivity apps by workers themselves, particularly with regard to the effects of such apps on worker morale and turnover.

In addition to the collection of data through workers typical workplace activities, some workers -- some workplaces employ wearable fitness trackers as part of their workplace wellness programs. As previous research has shown, the data resulting from fitness trackers is oftentimes irregular and unreliable. Furthermore, the data from such devices requires interpretation. Data analytic companies, which interpret the data and now defining -- they are now defining the standards that measure a worker's health status and health risks. One problem, as scholars have noted, is that medical and health research is rapidly advancing, such that current standards as to what is healthy are not the same as they were in the past. Yet, companies that interpret the data from wearables lawfully operate as black boxes, revealing nothing about their data sets and the algorithms used for interpretation.

In the age of big data, there is legitimate concern that the behaviors, activities, demographic and genetic information of workers collected at the workplace, even if not used directly by the worker's -- by the employers collecting it, could become part of data sets that are ultimately used to make employment-related decisions. As so much of the activity behind internet scraping and machine-learning is not known or knowable to the people designing the programs, it is not always foreseeable that a certain algorithm will access and deploy prohibited information in its decision making process.

To summarize, while big data could have positive functions for the workplace, we must remain alert to the potential for big data to be used in ways that -- in ways that essentially violates the spirit of the laws of anti-discrimination and in ways that would permanently erode worker privacy.

My recommendation is that employers should take seriously the worker privacy and employment discrimination concerns raised by big data and worker surveillance, and commit to concrete steps that will help minimize the risks posed to their workers. I suggest that listeners review the article 'Health in Big Data' available on SSRN, which offers some practical solutions for employers to consider. Thank you very much for listening to my testimony today, and I remain open to questions. Thank you.

CHAIR YANG: Thank you very much. Dr. Lundquist.

DR. LUNDQUIST: Thank you Chair Yang and distinguished Commissioners. I really appreciate the opportunity to share with you, my thoughts about the challenges and the opportunities of big data, and specifically today, I want to focus on

validation issues, and what some of the best practices are relative to big data. There really is no doubt that big data is the future of HR. We've heard it from the other panelists, but it presents a future that's both promising and scary. Big data is a tool and it's only going to be as good as the way it is implemented. So we need to look not just at its potential for reducing subjectivity in the workplace or outreach. We also need to look at its potential, in terms of unreliability or errors in those data sets or models, as well as the absence of validity evidence relative to the models. I'd like to begin by talking a bit about the algorithms themselves and machine-learning. As we've talked about, the algorithms iterate on the data until they maximize the prediction for a particular target group, often times, best performing employees, and what's done is a match between the characteristics of those best performing employees, such that that can then be applied, for example, to recruits outside.

In my own experience, I've seen algorithms developed that have hundreds or even thousands of data points, maybe even more than thousands of data points, that are combined in an algorithm. But for any individual candidate, not all of those data points will be populated. Because you're reaching out to the internet, there is going to be missing data for some of those. So, for example, in one algorithm I reviewed, over 100,000 individual data points were potentially scoreable. But for any individual candidate, because of missing data and the nature of the characteristics that were scored, only about 500 data points were scored. So, from person to person, those 500 data points were not necessarily the same. In that situation, you have two candidates who are being compared on -- or who are being evaluated on different pieces of data. Effectively, this is a test with different questions for different people. Often times where those algorithms are using publicly available databases, the problems of accuracy and incompleteness of the data just generates more of the missing data that we've talked about. In some cases, the algorithms themselves, because they're built to maximize prediction, may look at variables in a counter-intuitive way or a non-logical way.

So, for instance, you may see that some GPAs that are lower are given more weight in the algorithm than GPAs which are higher, because they're looking to maximize the data. In some cases, and this has been discussed previously, the algorithms are trained to predict outcomes which are themselves, the result of previous discrimination. So, this high-performing group that on which the algorithm is trained, could be a non-diverse group. It could be a very diverse group. It could be built in different ways, but it's important to look at what is the group on which the algorithm is being trained, and if it's a non-diverse group, for example, the

algorithm is being trained to maximize the characteristics of that group. So, it may reflect the demographics of that group more than the skills and abilities that are needed to perform the job. My concern as an organizational psychologist is that the algorithms often times appear to be matching people characteristics rather than the skills and abilities needed to perform the job.

By its nature, the development of these algorithms are A-theoretical. So, they are a black box, very often, driven by the data and not necessarily based on the theory of how you would expect something to predict performance. So, for example, as I talked about before, if you're going to score thousands of different data points, maybe 100,000 different levels within those data points, it would be important to know what those are and what is the intent of them. So, where you see organizations that are scoring by grams, combinations of words from somebody's resume, how does that predict performance and what is it that we're getting beyond the correlation? So, in order to get greater stability in the predictions, as well as greater interpretability and understanding, I would argue that we need to have some understanding of the relationship between the variables that are being measured, and job performance, not just the simple correlations. The correlations are the result of the method.

Where an algorithm is not fixed, and some employers are using algorithms that are not fixed for a given period of time, so for instance, I run my study and this is the way I'm going to score everything. If they continue to iterate the data, that means that you have different tests each time there's a new iteration. So, I would argue that there needs to be more -- separate validation evidence for each of those decision criteria that are being used, because there, the test is changing repeatedly.

The lack of complete data that I talked about before is something that affects the reliability of the test, and without reliability you can't have validity. So, if a test does not consistently measure, then in fact, it will erode its validity in the end. I think that's something that really needs to be carefully looked at.

Finally, I'd like to end with some best practices that I think ought to be considered, in terms of big data. First, that the valid -- there be validation of the predictive model's accuracy over time, and with different employee segments. So, if you build an algorithm based on current employees, will that affect applicants in a different way from current employees? Will there be different levels of experience that necessitate different predictive models, and does that change over time? If you

change the diversity of the high-performing group, does that then change the model?

Second, conducting a job analysis to -- because I always have to say, you have to conduct a job analysis, that's part of my role in life. But it's important to know what knowledge, skills and abilities are required to perform the job and to track that back to what's actually being measured in the black box.

Looking at the representativeness, a third, looking at the representativeness of the populations and the variables that are included in the development of the algorithm to ensure that they are reasonably inclusive, and then a couple of things that often times do not get measured.

Fourth, I would like to see managers be trained, so that they can interpret the information that's being used and that candidates finally, five, candidates be informed of the use of the information to avoid some of the privacy concerns previously discussed. Thank you very much.

CHAIR YANG: Thank you Dr. Lundquist, and to all our panelists. Now the Commission will have an opportunity to direct questions to you and each Commissioner will have six minutes. We will have first round of questions, then we'll take a 10 minute break, and then we'll come back for a second round of questions, and I will just get us started.

Dr. Lundquist, you had mentioned that the big -- the way these big data screens are created, it can lead to people actually being evaluated based on different types of tests or different questions, because the information isn't complete, that you may not have the same information on each candidate, and what do you see as the validity implications for that kind of discrepancy, and then how do you address that? What are some of the strategies to try to structure algorithms to avoid those problems?

DR. LUNDQUIST: I think it -- I think it really depends on how you're going to be using the algorithm.

So, if you're using something for recruitment, I see it as a different kind of implication, than if you're using it for selection, but if -- because in the recruitment situation, you're doing outreach and you're trying to find different types of folks who might all be able to do the job.

Where I'd become more concerned is if it becomes a screen that prevents somebody from moving forward, and in those situations, I would like to see the algorithms based on categories or features of categories that are fewer than hundreds of thousands, and perhaps at least -- it might be that you have to sacrifice some level of predictability in order to get something that is meaningful and interpretable where there is more consistency from person to person.

CHAIR YANG: Thank you. That's helpful. Dr. Kosinski. You had testified about the ability of big data to identify talent, as opposed to skills that have been acquired, and what kinds of factors are these big data algorithms using to identify talent, and what are ways for employers to use this concept?

DR. KOSINSKI: Well, so, as has been mentioned a few times before, big data base methods are very accurate and we can aim them at predicting many different outcomes. In the legal context, for instance, they are being used to predict how likely someone is to re-offend, and this information is used to make decisions about the state parole, granting parole and so on.

In the context of employment, we can aim those methods at predicting job outcomes like job performance, but we can also go one step below that and say, let's put job performance on the side for a second, because here we are prone to all of the biases that Dr. Kathleen mentioned before, and let's focus on dimensions that we know we have a theory and we have a lot of evidence from previous studies, that we know that those factors predict performance, like cognitive abilities, personality traits and particular skills, and now, digital footprint based models can be based to - - can be trained to very accurately predict those and again, as has been discussed before, there is a huge number of methodological issues and those things are not simple or obvious. But there are people doing their PhDs in and studying how to deliver those tests in a valid and reliable way, and we have a lot of evidence, that are more valid and more reliable than traditional pen and paper questionnaires, and also they are way more valid and way less biased than decisions made by humans, who are the most unreliable judges of character and also most prone to biases.

And I just want to also quickly refer to the issue that was discussed before, of algorithms using different types of data for different people. This also happens in traditional assessments. So, a computerized adaptive test like GRE or GMAT are doing exactly the same thing, and there are computer algorithms, they are making sure that even if you're using these different pieces of data for different people, the assessment still remains accurate and valid.

CHAIR YANG: Thank you. I wanted to follow up on a point Dr. Lundquist mentioned, that perhaps we should consider using fewer than thousands of different data points and focusing on some, you know, specific areas that we know to be more closely correlated with job performance.

Dr. Housman, do you have an opinion on that or have you seen employers use that in different ways, in terms of how many different factors they may be considering?

DR. HOUSMAN: Yes, I will say that some of the assessments that Dr. Lundquist was describing are very different than what I've been exposed to, and I think it's -- it underscores the point that there is a million different ways to build a model, and so -- and it -- and there is a lot of variety out there.

I will say that the models that I've worked with, or the few companies that I've worked with and/or advised have been theoretic -- you know, we generate based on job relevant characteristics. We're looking for specific traits, like conscientiousness and technical abilities. We're not throwing thousands or millions of data -- I just haven't seen that done, but it could certainly be done, and I'm just not aware of it, and we're not validating against an existing population. We're using it by testing it against new applicants.

So, I will say it makes sense, I think, including a million factors into a model of which the majority are missing data points is not best practice, but that's also not something I've seen in my time doing this work.

CHAIR YANG: Thank you. Does anyone else want to speak to that question?

DR. DUNLEAVY: I think the IO psychology literature does have some related information, and this goes back to that concept of bio-data that I mentioned.

This has been around for a long time and when bio-data research began, I think in the 1940s, a dust-ball empiricism approach was used and we -- you know, researchers looked to see what correlated with performance and moved on from there, and eventually we were able to compare that approach to more rational approaches to scoring bio-data items and found that the empirical approach did a little bit of a better job.

But I think the contemporary thought today is that a hybrid approach is really preferred and can maximize multiple goals in using bio-data for selection, and so, that -- that hybrid approach is to leverage empirical keying as a starting point, but

then really provide some rational thought on what it is that you're looking at, what variable relations you're seeing, and whether or not they are real, and on top of that, organizations can then take a values-based perspective and say, hey, there may be a correlation here, but my values may lead me to choose not to use it in the algorithm for various other reasons, including diversity potentially.

CHAIR YANG: Thank you. I see that I'm out of time, but I'm sure we'll have time to get more testimony on that. Commissioner Barker?

COMMISSIONER BARKER: Well, I want to start by apologizing in advance for my questions, because this is so beyond me that I know all my questions are going to sound totally stupid, but I, you know, ask for your patience in allowing me to do that.

So, I'm trying to understand how this works, and I have two thoughts on this, and I'll just throw them out to whoever would like to answer it.

The first question is, so, we look at all the different -- a number of different ways -- a number of different -- well, protected groups that we are trying to make sure that any process like this does not adversely affect.

So, let's look at age discrimination. So, I've got, let's say, applicants for a law firm, and I've got a -- and you can't discriminate against an older applicant. So, I've got applicants for an international law firm, that gets hundreds of thousands of applicants a day, and I've got applicants who are in their 20s, and their 30s, and their 40s, their 50s, their 60s.

How are you able to equate the -- the data, particularly the non-traditional data that you get from a 20-something with a 60-something, particularly if that 20-something has a huge digital footprint and that 60-something has a very small digital footprint? That's my first question.

And the second question is, what's to prevent a whole cottage industry from developing by data scientists who now -- who have been in the business of developing these models for employers? What's to prevent them from going into business, designing and manufacturing artificial digital footprints for applicants? So, anybody want to answer my two questions?

DR. HOUSMAN: I'll provide some insight into how we do what we do, right, which is we would engage with that company, gather outcomes data. We would -- again, our

exposure with psychometric data. When they -- we have -- sorry, we go through and construct a job profile, based on what we think were job relevant characteristics for someone to succeed in that job, then we would deploy an assessment and we'd gather enough data from applicants to see whether the questions that we ask them correlated with their outcomes on the job, and from that, a scoring model was built that would score those applicants and provide, in our case it was a red, yellow, green score, based on how they responded to those questions.

So, it's a little independent of some of the examples that Dr. Lundquist was giving around social media, right? Scraping the internet, I don't know of a lot of companies that are doing that. I'm sure it's being done. That wasn't quite how we approached the problem, right? We really thought about job relevant characteristics, and in our mind, if we found that conscientiousness and technical abilities and aptitude for learning or legal -- you know, understanding the law, were correlated with outcomes, we thought that would be independent of age, sex, race, gender, things like that, and so, our view is that if you can deploy a good scoring model that is based in some sort of theory, we're not -- and we can always run tests after the fact, to see if there was any sort of discrimination. But that was the approach. Hopefully, that helps answer the first question.

MR. MRKONICH: And Commissioner Barker, as to the second question, I think whatever is done, creates a cottage industry. Right now, there is a cottage industry based around validation studies, right? There are people out there who earn their living, quite legitimately and appropriately, conducting validation studies of tests and other devices used in the hiring process.

There are people out there who earn their living preparing people to take the SAT and other graduate school tests. So, I think you would expect to see advisors and counselors and consultants grow up that help employers do this correctly, and you'd also expect to see outcomes based on employers learning more and more about doing this.

With any new device, there is an opportunity for abuse, just as there is with the current system. So, what I expect to see, counterfeiters who are trying to create artificial devices? Perhaps, there is at least an anecdotal story that one application screening application screened for certain elite schools in the resume. Well, after that happened, people who did not attend those schools learned to put a watermark on the piece of paper on which their resume was printed that included

those schools' names, so that the scanning device would pick them up and they would trigger that as if it were a positive result.

Those things happen. But abuses will happen whether we stand still or move ahead, and I think that the real opportunity here is, those kinds of -- those kinds of consultants that can help employers do it right will help all of us.

COMMISSIONER BARKER: No, my concern is, you've got somebody who has made a point of staying off the internet.

MR. MRKONICH: Right.

COMMISSIONER BARKER: And you've got somebody who doesn't have access to the internet. If you've got a company that is, as you say, scraping the internet, you know, how do you give equal opportunity to those -- a candidate who doesn't have access to the internet, and one who has built a huge digital -- carefully manufactured a digital footprint on the internet? Dr. Anjuwa?

DR. ANJUWA: Yes, I think that's a very important question that you raise, and which goes to the heart of, you know, the future of big data, whether it will privilege people who already have access and leave people who don't have access further in the margins, and I think that --

COMMISSIONER BARKER: But Dr. Housman, very quickly, I only have a few seconds. Do you have a response to that? Is that a legitimate concern or not?

DR. HOUSMAN: It's -- it is a legitimate concern. I will say with employers we worked with, they will gather resumes from job fairs and enter it into the system. I mean, there are ways around it that are a little bit more manual, but you know, I understand the concern, and I do think that access to the internet is a predicate for taking one of our job assessments and for --

COMMISSIONER BARKER: I'm out of time. Thank you very much.

CHAIR YANG: Thank you. Commissioner Feldblum.

COMMISSIONER FELDBLUM: Thank you. So, certainly one thing I think we can say from all of you is that big data is here to stay. So I take -- I derive two things from that.

One is, we have a responsibility to understand it as best as we can, to get ahead of the curve, to the extent we can, and I think the second is, what can we do as an agency to enhance the positive diversity effects and minimize the potential negative ones?

So, I'll go to -- Dr. Kosinski described it and it captures what many of you did. Would doing objectively measurable data to accurately determine employment success? I think that sort of captures the positive diversity piece.

So, I'd like to go to the -- just talk about some categories of people and see could there be -- what might be the problems, right, because as Dr. Trindel said, the experts who are developing the algorithms are not thinking necessarily ahead of time about certain populations.

So, my question is both, do you see a problem and do you see that there is some way of thinking ahead about this -- these groups to -- to stop things from the beginning?

So, Dr. Trindel mentioned two, people with disabilities that might have a number of absences. Women of certain age group that might have more breaks in employment. I want to add four more: people on the autism spectrum. This is specifically in terms of personality questions. People with autism often read words very literally and don't get what seems like a basic social understanding.

People who come from relatively poor families and/or parochial settings, you know, very religious or particular cultural. I say that personally, as someone who grew up as an Orthodox Jew in a poor family, and lots of activities I never would have done or things I would have known about, which I now have a sense of because I've spent 30 years in what I call the secular world, and everyone calls it the world, and with more income. So, I can see the difference.

The last category was the people who are not on social media, which you heard already, but again, some of this uses social media, some not.

So, I guess my question is, do you think there is a way for the experts who are developing these algorithms initially to be educated in some way about these groups, such that some potentially discriminatory outcomes might not occur, and I actually want to go to Dr. Anjuwa, Dr. Trindel first, because of having thought about it and then for other folks. I mean, you've all thought about it, but having raised it.

DR. ANJUWA: I think that's an excellent question, and I think it -- you know, your question goes to the heart of something that happens with big data, which is sort of this -- this sort of like, neglect to realize that correlation is not always causation, and therefore, factors such as absences are linked to success when necessarily they might not be.

I'd like to add one other group: Veterans, who, you know, have long periods of absence from the workplace, and if that's not highlighted for the algorithm, the algorithm would mark that as unemployment, for example.

So, I think issues like that are really important to be addressed in the forefront, in terms of how the algorithm is coded and in terms of the variables that go into determining a particular characteristics, and just to keep in mind that, you know, correlation is not causation.

So, the fact that we have certain factors, as the first panelist mentioned, you know, being white, being middle-class, being linked to success; that is not necessarily a causation for success. So, I think that's a really important question that you raise.

COMMISSIONER FELDBLUM: Other folks?

DR. TRINDEL: Well, I think a certain amount of outreach and training for the types of data scientists that are creating these algorithms is called for.

As I mentioned, the way I see a lot of these things developing is a movement from marketing to employment. I see a lot of that, a lot of crossover happening there, and so, as I say in marketing, I'm not an expert in that area, but it seems like it's much less of a regulated area than employment.

So there's really nothing to stop you in marketing from the type of profile that you've described. I mean, I've seen some -- I've seen profiles that are like, I'm trying to think of the exact wording, but "about to retire broke". There is a part -- there is a person. Let's target them for like, certain -- you know, like maybe reverse mortgages or something like that.

And so, these profiles have been developed based on our behavior online, and to the extent that they're already moving into employment, I'm sort of thinking about that they could be, and that it seems like they're likely to, and so, it's important for us to, I think as you said, stay a step ahead of it and be there to point these things

out and maybe educate the types of persons -- the types of people who are developing these algorithms, and I just want to mention one other thing.

Something much more basic, like recruiting from LinkedIn. This is something that's definitely going on. So, passive recruiting, reaching out to people, this looks like somebody I want. If you look at the demographics of LinkedIn, they're not the same as the demographics of the U.S. population, or even when you look at the demographics of people on LinkedIn, in particular locations. They don't line up with the demographics of people in those same locations.

So, right there, you're limiting your labor pool and you would have to be mindful about that as an employer, and have an idea for why you're doing that.

DR KOSINSKI: So, if I could chip in here for a second, because this discussion is very close to my research interests and I develop big data models myself, predicting traits like that, and we always have to keep in mind, what are the alternatives to the big data, and the alternatives are going to LinkedIn and looking for people who are white and middle-class and match our imagination of what the right stock broker should look like, right?

The alternatives are personality tests, and it's totally true, if you have an autistic or a disabled person, they are going to fail the traditional personality test and they are going to fail an interview with a white middle-aged recruiter who is deciding whether to hire them or not.

Digital footprint based models are not going to change the lives of people who are already well prepared for a job market. If you have evidence that is not -- gathered from the social media, if you have evidence in the form of strong CV and degree from Stanford University, you don't need people looking at the digital footprint. You are basically hired.

Now, digital footprint lets people who do not have that, Veterans who do not have employment history, people with autism who cannot excel at tests, people with lack of access to education. Now, you can go beyond the interview and the biases of people, or again, searching through LinkedIn, which is just a traditional way of recruitment. We can go into trying to find ways of distinguishing between people who were unemployed for their whole life, and some of them were unemployed because maybe they are not motivated or they're not skilled. Some of them were

unemployed because they had a disability or they have the wrong color of the skin, or they were in the army for many years.

Now, digital footprints give us a chance to basically go beyond the first impressions and biases in recruitment.

COMMISSIONER FELDBLUM: Thanks. My time is up. We all have it -- but I'm sure you all have other things to say and hopefully, get woven into other answers.

CHAIR YANG: Thank you. Commissioner Lipnic.

COMMISSIONER LIPNIC: Thank you Madam Chair. So my question is just to go back a step. Let's say a vendor comes to an employer, and is making a pitch to the employer about, we have these great data analytics and we want you to buy this package from us. What is it that they are pitching to that employer? So, I'll start with Marko and Mr. Housman and Dr. Dunleavy, if you want --

MR. MRKONICH: Thank you very much Commissioner Lipnic. I think what they're really pitching is the ability to create a more fair and effective decision making model for employment decision making. It is ultimately the way to escape the 'me too' bias that exists in subjective decision making models. It's a way to focus on and capture candidates you otherwise might miss. It's a way to focus on -- on factors that correlate, at least, to job success.

I do think with big data, you have to accept at some level, there is a correlation versus causation change, but that's true in every aspect of our life right now. It's from the weather forecast to how we buy -- how we buy anything we buy.

So I think what they're selling, is success, so that -- that business is saying, I can reduce my turnover; I can have more effective workforce; I can be promoting and selecting the right people. That will translate into marketplace success and enable me to create the kind of workforce I want.

They're also saying we can create a more diverse workforce, that we can create a workforce that includes people that weren't otherwise included. So, any new tool can be abused, and they will recognize that in the sales call, but they'll say we've designed it, so that that doesn't happen, and what we're doing is creating a better solution for you, your employees and your customers.

COMMISSIONER LIPNIC: Before you answer --

MR MRKONICH: Yes, agreed.

COMMISSIONER LIPNIC: So, but are they -- is the vendor saying, I have this package. I have this pre-packaged product, and you can buy this package and you're going to get the most successful workforce you've ever had, or is it to Dr. Housman's point earlier, that we come in, we're going to talk to the potential employer and figure out your workforce and what's been successful in the past, and somehow we're going to figure out this way that, you know, we're going to find these correlations for things that will make other people successful, who you haven't thought of before.

MR MRKONICH: It tends to be that each vendor has its own methodology and its own system and its own process, and that's what they're marketing. Some are more tailored. Some are more customized, just like every other kind of consulting and advice, and I think it's fair to say that, you know, it -- it's just like any test used historically. Some, you know, are tailored to your unique work environment. Those tend to be given to people going into the C-suite and those higher echelon jobs. What this does is make it economical to apply some of the same selection techniques that are used for the most important jobs in the company, in an economical way to the lower lever jobs, at the entry level jobs, and make sure those workforces are also tied to workplace success. But the -- the varying -- and some are very cookie-cutter. Some are very customized, and it really depends on the vendor.

DR. HOUSMAN: You know, that -- I think Marko is exactly right. The outcomes are almost the same, right? Better workforce, stay longer, perform better; but it could be customized to your specific pool; it could be an off the shelf assessment or a -- an approach that's been validated against hundreds -- you know, thousands and thousands of employees, and likewise, the data sources could be -- are very different. It could be scraping the web and you know, publicly available data, social media. It could be psychometric assessments. There are games that are now being used. So, I get the -- there is a wide array, in terms of the 'how', right, but the -- the 'what' does it achieve is generally pretty consistent.

DR. DUNLEAVY: Yes, I agree. I think the answer is who knows, and I've seen situations where a vendor will come with sort of a list of "off the shelf" content that they have essentially done research on, foundationally or for other clients, and then I've also seen vendors sort of take the approach where what you're getting is the expertise for them to come in, evaluate your specific data, what you have available in organizational systems, and really apply the expertise and the machine-learning to develop the algorithm to maximize prediction of whatever it is, but from an

outcome perspective, you have the story is typically, we will maximize whatever it is that you want us to predict.

COMMISSIONER LIPNIC: Dr. Lundquist.

DR. LUNDQUIST: I'm sure that's that story. I certainly have heard that story before. The problem is that many employers are being confronted with sales pitches and many people in the organization are listening to these sales pitches and what they promise is, I'm going to fix everything. Here is my black box, and it's going to work, and I think -- and I've seen that not just from external vendors; I've seen that from large companies where the internal software engineering folks want to develop their own algorithms.

There is a lot of interest in this prediction. But I think Dr. Housman and Dr. Kosinski are examples of doing it right, having a theoretical basis and blending that together.

Unfortunately, that's not always the case. That's not always what a vendor is offering. That's not always what a company is doing, and I guess I would just encourage you to be aware of that, and to the extent that there is any kind of education that gets provided, I would hope that that education gets provided both to the people who are developing these kinds of algorithms, but also to the legal community, the internal counsel, who have to advise their companies about the use of this because it's widespread.

COMMISSIONER LIPNIC: Yes.

DR. ANJUWA: Yes, so, yes, so, that's a really important question in terms of thinking about what exactly companies are getting from the big data analytics, and one thing I know for sure, they're not getting is -- and some sort of audit process. So, they might be getting a black box that claims to be able to come in and fix everything, or they might be getting something that's claiming to be more tailored, but what usually is not present is some sort of auditing process at the back end, to see what at the -- the disparate impact of what's -- what the processes are doing.

So, auditing, in terms of seeing, okay, we're -- we claim we're predicting success, but how many people are we neglecting to find or how is our pool affected by these new variables that we're using?

So, I think that is something that -- I think that's something that employers can request, and going forward, that there be some sort of auditing process that would

look at the disparate impacts of the measures being taken.

COMMISSIONER LIPNIC: Okay, I'm out of time.

CHAIR YANG: Thank you. Commissioner Burrows.

COMMISSIONER BURROWS: Yes, there's a lot to talk about. I think I would pick up right where you left off, and follow up on this question of, if I'm the employer and I'm really interested in this. It's going to save me so much time, I can really identify, personnel decisions. That takes a long time. Make it faster. Take it off my plate, right?

But how do I kick the tires on the two things that are most important to me? A) it's really going to work, it's going to be valid, and B) I don't -- you know, particularly if I'm in an industry that needs creativity, I don't want to just reify my own current workforce, which may not be particularly, you know, diverse, and you know.

Our last hearing was on the lack of diversity in tech. So, we're going to the same companies that we know themselves are not particularly diverse, in some cases, and no present company, I have no knowledge of that at all. But you know, and asking them to design something that's going to help fix a diversity problem.

So, if I've got my company, I really want to do this. How do I kick the tires on this question of, I mean, am I getting a good product or am I buying a lawsuit or at least a bad publicity problem, because I don't know if you're going to screen out people based on race, national origin, et cetera. I may not have any access to that. So, that's a question to each of you, and I'm particularly -- I don't want to put you on the spot, but maybe Mr. Housman, you talked -- Dr. Housman, you talked earlier about the fact that with respect to age or just lack of -- you know, some of the -- you might have less of a digital footprint, that there are things you can do manually to supplement for that, and so, maybe that could be reflected, as well, but those are the two things that I think if I were an employer, I'd be really worried about.

DR. HOUSMAN: Sure. Certainly, to your first point, how do I know it's working? I can tell you that the companies we work with are very scrutinizing and we meet with them quarterly and we show them results and we show them where it's working great, where there's room for improvement, because they care about the bottom line and they want to know that this thing they're paying a lot of money for, is producing results.

COMMISSIONER BURROWS: And the front end, right, because I don't want to pay a lot of money and wait for the audit. I want to -- you know, have some way --

DR. HOUSMAN: Yes.

COMMISSIONER BURROWS: -- of assessing maybe I'm talking to six different people -
-

DR. HOUSMAN: Sure.

COMMISSIONER BURROWS: -- and what questions is the employer supposed to be, you know, asking?

DR. HOUSMAN: I mean, they're certainly asking about results. You're showing case studies. Sometimes you're going into the models themselves. Sometimes they're not as interested in the black box, what's going on, under the hood.

To your second question about how do I know that it's increasing or letting -- at least not decreasing diversity? Again, we would generate an adverse impact report whenever we deploy a new assessment. So, again, you know, I know Dr. Anjuwa said those -- those aren't frequent. We would do it. I can only speak for ourselves.

Around the diversity question, that's a legitimate question. That's something that we weren't often asked to speak to. Again, we were showing that the assessment was not discriminating against any protected classes. But we -- the very few asked us, is this also promoting diversity? Are we getting diversity of thought, you know, in the door because of the assessment having been deployed.

MR. MRKONICH: With -- Commissioner Burrows, as long as we're in Washington, D.C. and there's a baseball game, I'm told tonight, Money Ball, the movie is really a big data movie. If you look at it, historically, people looked at runs, batted in, and batting averages and ERAs. Advanced cyber-metrics came in and that's what that movie is about, and if you look, teams don't find nine short-stops. Teams using big data find a person for each position, and there is no reason big data can't be used to create a creatively diverse workforce, ethnically, racially, from age -- whatever basis you want to say, diverse workforce and used properly, big data would eliminate artificial barriers that prevent that from happening.

COMMISSIONER BURROWS: I'm open to that. I just want to figure out how we can tell -- separate the wheat from the chaff in this one.

MR. MRKONICH: Ultimately with anything new, that's the challenge, and that's really why I think it's terrific and I feel honored to be here today, because this is the beginning of a dialogue, that I think has to continue because I don't think everyone has all the answers.

There are certainly disagreements and areas of agreement within this group, this panel, and I think you can only imagine if the group were just expanded for discussion, how broad the disagreements might get, and I think that shedding light, developing better understanding is the beginning, not the end of that process, and I really do think it's more talk, more discussion, more analysis, more understanding what is going on, what isn't going on, what's working and what's not working, because everyone has a shared interest in making sure that what is being done works.

There will be winners and there will be losers in the marketplace based on change. You know, big data eliminates traditional selection devices, right? I mean, so the people who have a vested interest in the traditional selection devices won't like the big data.

But what we have to do is make sure the new things that are working in every other aspect of life are welcomed into the employment community in a very fair and even-handed way. Employers want to be fair.

COMMISSIONER BURROWS: Absolutely. I think most employers do, and I think the question is, how do you do that when the folks that are the experts at whether they need the job or not, the folks who are the expert in this product, you're suddenly creating a product that the expertise that's in-house is -- doesn't really answer. So, I don't know, if, Dr. Lundquist, if you wanted to speak to that, and particularly again, I'm interested in -- you know, I for instance suggest -- believe that my son has a bigger digital footprint than I do, even though I've been around a lot longer, simply from the fact of how old the internet is, right, and how quickly people are getting onto the internet these days. He's only 10, but I bet he's ahead of me. So, speak to how you deal with that, if it's going to be so internet based.

DR. LUNDQUIST: I think it requires the kind of education and awareness among the people developing the models that we've been talking about here this afternoon.

If you're not aware that the digital footprint part of what's in your algorithm is going to vary as a function of age, you should be, if you think about it very much, but very

often, the algorithms are not built separating those out and looking at them from that standpoint.

Many times the algorithms -- I've seen some, which are developed where the internal folks say, "We won't -- we don't want you to know anything about the demographics of the group, on which you're training your algorithm." So, you're going to maximize for this high-performing group, but we're not going to tell you anything about them, because we don't want to bias the algorithm.

Well, the algorithm is already biased by the outcome of who those folks are. So, maybe algorithms need to be trained with awareness of certain kinds of demographic characteristics, if for instance, what you're looking for is an outreach to an under-served community.

So, I think that there needs to be more awareness, and I think there needs to be more focus on what are the variables that we're looking at in those equations themselves. I really think that that's the bottom line of the problem, and I -- as I said before, I think there are folks here who are talking about it done very well, but it is not uniformly done well out there in the real world.

CHAIR YANG: Well, thank you again to our panelists. We will now take about a 10 minute break and we will resume at just about three o'clock for our second round of questions. Thank you so much.

{Off the record.}

CHAIR YANG: If everyone could take a seat, we will get started in just a minute.

Thank you everyone. We will now reconvene for our last round of questions. We had really, a fascinating discussion and we know it is really just the start.

I wanted to ask of you -- each a chance to answer this question quickly, but what do you think are the most important questions that need to be answered in this space? Where are the greatest areas of disagreement or need for clarification that we should be thinking about, and I'll start with you, Mr. Dunleavy.

DR. DUNLEAVY: Great question. I think going back to the what is being measured issue is something that we absolutely can't ignore, and I would certainly advise that employers consider, you know, asking for some pretty basic foundational information, if they're approached by a vendor related to what it is that, you know,

the process is tapping into, what content inputs are being evaluated? Have you done research for others? What does that research look like? Have you evaluated sub-group differences? What do those look like, so that, you know, you can really be an informed consumer and think hard about what it is that you're thinking about potentially implementing in your workforce.

CHAIR YANG: Thank you. Dr. Anjuwa.

DR. ANJUWA: My apologies. I think the two main important questions are toward looking at the true impact of big data. So, beyond the sort of efficiency gains that's touted. So, looking at the impact of big data, both in choosing a diverse workplace, but also on the workers, once they are in the workplace, because there's also big data being collected within the workplace, and then the other -- the second question is, looking at what safeguards we can put in place, to make sure that big data, which could be a useful tool, doesn't become sort of a shield for covert discrimination.

CHAIR YANG: Thank you. Dr. Lundquist.

DR. LUNDQUIST: I go back to the question of what's being measured, for sure, and for whom is it being measured? So, how applicable are the results that are used for the algorithms for various populations and to what extent has the vendor or the developer of the algorithm explored those issues, and I would say one of the opportunities is, if we can think of ways to effectively use the -- the information that we have about people's race, gender, ethnicity, other kinds of groups, is there a way to build algorithms that would be appropriate or outreach or helpful in those situations, and to use them in a more affirmative manner?

CHAIR YANG: Thank you. Dr. Kosinski.

DR. KOSINSKI: Thank you for this important question. So, I think that one of the most important issues here is that the developers of those new big data models, they're usually software engineers. They're computation -- computer scientists, and the problem is that because of the training, they always -- not always, often are unaware of the decades of research into how to deliver assessment in the proper, fair, valid and reliable manner.

Issues that were raised before by Dr. Lundquist and others are of high importance for traditional I/O psychologists, but computer scientists are very often not even aware that considerations like that are there, and their answers to them, as well,

and so, well, in a way they're deemed to learn from their own mistakes, and I think that the important question is how we can help computer scientists and software engineers not to learn on their own mistakes, which will have -- may have dire consequences to some people in the workplace. How do we establish a system in which we can translate this knowledge to software engineers?

CHAIR YANG: Thank you. Mr. Mrkonich.

MR. MRKONICH: Thank you. Thank you Chair Yang. First, I should note one of the things that I was left with -- or the impression was, we don't use social media that often when we're evaluating current employees. There seemed to be an over-emphasis on how social media and digital footprint plays into how you deal with current workers, little bit different in the retention thing.

So, I think that's part of the answer is, we need to learn more and have a better focus and a better understanding before we act, because I think there is a risk of acting before we understand. I think there is a need to understand that job validity is a defense to -- or business necessity is a defense to a disparate impact, has a different meaning in a world based on correlation than causation. We need to have a dialogue about that and how to deal with those issues, and finally, we need to address the issue of what's the responsibility of the vendor and what's the responsibility of the employer, and we need to work at all three of those things.

CHAIR YANG: Thank you. Mr. Housman.

DR. HOUSMAN: Yes, I probably echo Dr. Kosinski's comments, in that a lot of data science and machine learning is taught that, you know, you include something in the model because it predicts, right, and if it predicts that's -- it -- it works, right, let's throw it in there, and so, that's not the case here. This is different than trying to get clicks on a website.

So, understanding the 'why', the 'why does this predict', and trying to come up with at least some theoretical model as to why some responses or something that's been gleaned about the applicant predicts on-the-job success I think is really key, and I think that training isn't, you know, there right now, and people need to be aware of these things.

CHAIR YANG: Thank you, and now, I will turn it to Commissioner Barker.

COMMISSIONER BARKER: Mr. Mrkonich, I appreciate you coming out and saying what I've been thinking, which was, you know, I think we tend to think that so much of this is -- contains a digital footprint element, when it sounded to me like, hearing from several of you that that's really not the case.

But I guess my question to you and to Dr. Housman or Dr. Kosinski, whoever would like to answer this is, okay, so, that is not the case now. But is that not the way we're going?

MR. MRKONICH: Well, again, maybe I need a big data model to tell me where I'm going, but leaving that aside, I do think that there is going to be more and more reliance on electronic information flow.

The Paperwork Reduction Act, if nothing else, is going to make sure in our governmental dealings, we're going to be using internet technology more and more often.

I think it's hard to know exactly, because you're measuring the things Dr. Kosinski talked about, in terms of talent, skill, personality traits, things that are more than just what shows up in a digital footprint.

So, I think that will be part of it, and maybe an increasing part of it, but all of us have a greater digital footprint frankly, if we're going to go back to that world, than we think we do right now.

If I bought a house, if I've borrowed money, if I have a credit card, I've got a digital footprint. I may not be on Facebook and I may not be on LinkedIn, but I do have a digital footprint.

So, yes, I think that's going to be an increasing part of it, but I don't think that's necessarily a bad thing. I think we have to make sure it's inclusive and fair, but I'd point out that this technology and this set of technologies has the ability to reach out to people who are not even able to figure out how to apply for a job. It's potentially a very inclusive process.

COMMISSIONER BARKER: All right. Thank you very much. Dr. Housman, did you have anything you wanted to add to that?

DR. HOUSMAN: Yes, I see it playing a bigger role more so in recruiting than in selection, right, we need to recognize those two are very different, that I see those

tools being more useful in terms of reaching out to applicants, encouraging them to apply for jobs. I think that's a form of kind of inclusivity and quite frankly, if you're not on LinkedIn, you're not getting an inquiry via recruiter or an algorithm, right. So, it sort of works the same way in both cases.

When it comes to selection and making decisions once someone has applied, I still think that the information that's voluntarily submitted, as far as I can see, the vast majority, if not all companies are using information that's voluntarily submitted through an application process, resume, some form of psychometric assessment, and so, I don't know of any companies that, as soon as you apply, are scouring the web for your information and then using that to make a decision.

So, I wanted to make sure -- I agree with what Marko had said, and I wanted to make the distinction between, you know, selecting someone and actively soliciting an application, right, encouraging them to apply.

COMMISSIONER BARKER: Great. Thank you. Thank you very much, and I'll waive the rest of my time.

CHAIR YANG: Thank you. Commissioner Feldblum.

COMMISSIONER FELDBLUM: So, I want to say given that I had a very skeptical question in the first round, I want to make it clear that I absolutely see the positive piece of this, that is in the human hiring manager, they can see that someone's coming in wearing a hijab or a Yakamas or what the age, gender, race of the person is. Maybe they might be more put off by someone on the autism spectrum than if you had some algorithm that correctly took into account ahead of time, some issues, didn't screen for mental illness, et cetera.

However, thinking ahead about what the EEOC can do, it does seem to me that we really do need to differentiate between the social media, Facebook, LinkedIn, et cetera, which in a research conference we had about six months ago, was clear that some HR folks are in fact, using social media in that way.

So, what is it that EEOC can do, in terms of educating those folks, and then there is a completely different dimension, which is this machine-learning, and I watched a stunning slide show of this sort of iterative use of data to keep coming to more and more predictive results, but again, often using a current workforce.

So, within those two very different approaches to data, what is it that you think the EEOC can do, in terms of education, and maybe Dr. Housman and Mr. Dunleavy and Dr. Trindel, I don't know, I think it's all of you, but --

DR. HOUSMAN: Yes, I would say potentially some guidelines, or at least some suggestions around how you can test these assessments and make sure that they're not unfairly discriminating against anyone.

Again, I think at Evolv, we were very protective. We were just -- lived in fear of lawsuits and that is going to kill any early age tech company, but so, I think we were very conservative. But I think we kind of followed rules that we thought made sense and were very conservative. I don't think there was enough guidance around what would be best practice, and so, that sort of information would certainly be helpful.

Also, because I think we were -- you know, there is a wide array of technologies and models in this space, and so, you know, you want to make sure that there are some best practices, that it's not just data scientists left to their own doing, that are doing this.

DR. DUNLEAVY: I agree, and I'm wondering if some type of blue ribbon panel would be useful for EEOC to think about putting together, pulling people from a variety of different backgrounds, HR, I/O, legal, data scientists, et cetera, just to get a handle around, you know, helping -- helping employers understand what big data is and the reality is that it's an infinite universe. It's not just one thing, and then, you know, really putting together some best practices and things that you want to look for, in the way of potentially implementing big data in what you do.

COMMISSIONER FELDBLUM: Right.

MR. MRKONICH: Commissioner Feldblum, if I might just add one thing here, because one of the things I think that's under-appreciated is how this dynamic is often happening within large companies.

It's not an initiative that's growing up within the HR function, and all of a sudden, coming forward. It often initiates in the financial function or the operations function, where they've seen it work in their supply chain management, they've seen it work in the way they market, they've seen it work in other senses, and they're saying, "If we look at our business, this is one of our most important resources," and if you surveyed them, they'd say it's their most important resource to other people.

Why aren't we using the things that work for us here? So, to me, the thing that the EEOC can best do is start to educate and learn, because all of us learn when we have a session like this, and I think the employer community, the people interested in privacy concerns and employee rights also need to get together and talk about this because we need not to be talking about what we've heard or what we've read, but what we've done and what we've literally experienced, before we can start issuing specific guidance to anybody.

So, ultimately, I think there probably does need to be some form of guidance, but before we take that step, I really firmly believe that we need to help educate our HR friends, our diversity friends, those people filling out affirmative action plans, all those folks, as to what's really going on in this space and what the opportunities are and how it needs to be done properly.

DR. TRINDEL: I just wanted to point out, I don't mean to be the devil's advocate here, but one important thing to keep in mind is, 'it works', you know, it increases ROI on the money we put into our workers, or what have you. It works.

It's just important to remember that we're talking about people here. So, it's a little bit different than, you know, in other aspects of the business.

So, it might, you know, behoove an employer to just not hire anyone with disabilities, or not hire any women that might, you know, become pregnant or whatever, and I can understand backing away and not thinking about them in terms of people, how that could increase ROI, but it's just very important, and I think a lot of us -- most of us would agree, or all of us would agree that it's just important to back up and remember that and to be mindful of that, if we're offering any kind of guidance or training or any outreach.

COMMISSIONER FELDBLUM: I've got a minute left, so.

DR. LUNDQUIST: I think I agree with what the panelists have been saying about being mindful of who needs to be educated. I would add to the group of people who need to be educated, the vendors who are out there pitching without really very much awareness, present company excluded, who are really forward-thinking, but that's not always the case, and I say that for very large companies that are out selling their wares, and in-house legal counsel, who really are struggling with trying to figure out what does validation mean in this context.

COMMISSIONER FELDBLUM: Dr. Anjuwa.

DR. ANJUWA: Yes, I think also there should be education for the workers themselves, sort of an awareness of worker rights, in terms of what data they should feel compelled to give up to their employer. So, I think that's also an important aspect to it.

And I think there is also more education for the job applicants, right, such that when they encounter situations where, they feel, you know, their information has been used in a way that's in violation of anti-discrimination laws, they can report it without thinking it's just, you know, going through the commbomulator, but you know, more recognizing the -- you know, the legally discriminatory aspects of it, and not accepting it as a par for the course.

COMMISSIONER FELDBLUM: And I'm sorry, Dr. Kosinski, that I ran out of time, but I'm hoping you can jump in on the next question. So.

CHAIR YANG: Thank you. Commissioner Lipnic.

COMMISSIONER LIPNIC: Thank you Madam Chair. So, picking up on where Dr. Trindel -- her -- your comment, because I was actually thinking the same thing about what, Marko, you had said, well, keep in mind where this is originating within companies, and so, you had started off saying, well, this is -- a lot of this is being transferred from marketing and experiences learned in marketing. But let's not forget that this is about people.

So, let me ask, starting with you, what is the difference that that makes, in terms of the application and use of big data?

DR. TRINDEL: Yes, well, in marketing, for example, and as Marko said, there's other areas of the business, of course, like logistics and we can talk about different areas. But with regard to marketing, I just see a very clear -- like, I've seen -- for example, I'm into the data. So, I've seen consumer relationship management systems morph into candidate relationship management systems. So, I've just seen a real corollary there.

So, with regard to how is it different, I think I've heard a lot of data-driven marketers argue that, you know, this is about optimizing, you know, presenting people with products that they really want to buy. So, if we know who you are and we know your background and we know what you like; we can create a situation where we can offer you products that you'll be interested in, so, you won't have to, you know, hear about advertisements for things that are just completely unrelated to you.

So, you can see an argument for that, and you can see how that area would be a little bit less regulated. But with regard to a person who wants a job, so they can live, that's a completely different situation. So, I would just argue that there's a giant difference and there's a reason why employment law is there, there is a reason why EEOC is here, to help regulate this space. So, yes, that's what I would say about it.

COMMISSIONER LIPNIC: Anyone else want to comment on that?

MR. MRKONICH: Well, thank you Commissioner Lipnic. The -- to me, the real key is that these are the most important resource that most companies feel they have, are their workforces; and if you are able to look at it in purely financial terms, I think you could justify the investment made in big data analytics, if it reduces turnover and improves output. But that's really not the only reason that companies have employees, and that's not the way that organizations deal with their own people because they want to have not only productive people, but happy people, because that works in the long run.

So, I do believe that as we go through this process, making sure employees are included and understand the way this is working a little bit better, would be good for everybody. I think right now, it's all so new and all so challenging, to just understand what's going on, that expecting people to then appreciate it and understand nuances is going to be a challenge.

But in the end, I am a firm believer that human resource departments exist to do the right thing, and including them earlier on in the process is terrific, and making sure that folks understand the intersection, because we're not talking about product here; we're not talking about a service you're selling; we're talking about the people that make that product and make that service. It's important to make sure they're included and treated fairly, so, absolutely.

DR. HOUSMAN: I'd just add, I understand -- you know, I agree with what Dr. Trindel said, that we need to be very cognizant of situations where the ROI for the employer isn't necessarily in the best interest of some segment of employees.

That said, I also want to kind of focus on the big picture which is, that these systems work and that they produce more engaged employees who stay longer, and the research shows that when you're happier on the job, the job is a better fit, you're a better fit for the culture, you stay longer, perform better and the outcomes are limitless, and so; I think by and large, the employer is in the same boat as the

employee, that that relationship, if it's stronger and it's a better match, that benefits both sides.

So, I understand the concern. You want to make sure that no one is left out. But personally, I think that these are by and large, a force for good, on not just the employer side, but the employee side, as well.

COMMISSIONER LIPNIC: Did you want to comment? Okay, because I had one other question that I wanted to ask.

So, Dr. Lundquist, you talked about this earlier, about employers reaching out across the internet and they may have a -- you know, access 1,000 data points. How are they accessing those data points? I mean, what is it that they're -- and are they really looking at 1,000 data points? What is it exactly that they are looking at?

DR. LUNDQUIST: It can be the internet. It can be even resumes that are looked at, that form the basis for the algorithms. They may be looking at the number of different sites, to which a person has posted their resume, for example. They may be looking for particular words on a resume. Do you do this kind of computer programming or not? They may be looking at your GPA.

But they may be looking at the fitness kind of data, that Dr. Anjuwa talked about. They can be looking at all kinds of a variety of things, some of which are out on the internet, but even as simple as just looking at what's on somebody's resume, you will have different combinations of variables that can be looked at, if it's words as simple as the word on your address, your zip code. It could be -- any of those variables could be in the regression equation. So, it becomes very important to look at the content of it, and I think this discussion about educating computer scientists is really important, because the computer scientists tend to be very much enamored of the correlation that's coming out, but not necessarily the impact that's coming out of those variables.

DR. TRINDEL: I would just add to that. I haven't heard much talk -- I think the issue came up about a difference between, you know, recruiting and reaching out to people based on information on the internet, but then about current employees.

There is a lot of different ways to collect information about current employees. I mean, they're there all day at your company. So, everything from their emails to -- I mean, you can really -- I've seen people construct really interesting and complicated networks, that you -- it's about the meta-data, like you email this person, you call

that person, this is your network, and then how does that translate to success, and then perhaps, we could make selections or promotions or offer incentives to people that have similar networks in the future. So, it's sort of -- there's a lot.

COMMISSIONER LIPNIC: Anyone else want to comment on that?

DR. DUNLEAVY: Yes, I would agree, there are companies with HR analytics teams, housed within HR, that do an enormous amount of research along those lines, and can identify an assortment of different organizational systems that may have metrics, that really do represent performance, and you know, in those situations, of course, fairness considerations and diversity and inclusion may also be areas of interest and being housed in HR, those conversations can't happen.

COMMISSIONER LIPNIC: Thank you Madam Chair.

CHAIR YANG: Thank you. Commissioner Burrows.

COMMISSIONER BURROWS: Yes, I had a quick question. Hopefully -- my question is quick. Your answer can be as long as it needs to be, for Professor Anjuwa.

With respect -- you talked a bit about employee privacy in your written testimony, and I wanted to -- you mentioned that there might be some things that employers could do to protect employee privacy, and I wanted you to sort of expound on that and tell us what kinds of things those would be.

DR. ANJUWA: Yes, thank you very much for that question. So, I think in respect to employee privacy, the most sensitive point is when it comes to health data, that's collected as part of wellness programs, workplace wellness programs. And these workplace wellness programs are generally administered by outside vendors, some of which are part of health insurance companies, and some of which are not.

The distinction there is that the ones that are not, are not considered HIPPA protected entities, such that the HIPPA rules would not actually apply to the health data being -- that they are collecting.

In an article, 'Health and Big Data', my co-authors and I talk about several steps that the employer can take when running a wellness program, to ensure that worker privacy is still protected. And some of those steps include a commitment to keeping employee data private from the employee itself -- employer itself, such that the health data is then not used for employment decisions going forward, and also a

commitment to ensuring that the worker has informed -- gives informed consent for collection of the data. So, the worker has all the information as to what the data being collected, is being used for and also, what it will represent in terms of the worker's profile, and also in terms of the data flow of that data, in terms of, can the wellness vendor now sell that data, which actually is a very commonplace practice.

So, these steps, you know, if an employer were to commit to them, would at least start to reduce some of the privacy risks posed to the worker, and protect the worker also from employment discrimination, while still benefitting both the employer and the employee, if the employer -- employee wants to participate in the wellness program.

COMMISSIONER BURROWS: Thank you. Mr. Mrkonich. Sorry.

MR. MRKONICH: It's actually Mrkonich.

COMMISSIONER BURROWS: Mrkonich. Exactly the way it's spelled, okay, Mrkonich.

MR. MRKONICH: Mrkonich is actually the Croatian word that means dark.

COMMISSIONER BURROWS: Okay, I'll remember that. So, Mr. Mrkonich. I wanted to follow up on something you said about the way that these decisions are making -- these decision making processes are happening in companies, in the real world, and it reminded me that I wanted to understand better, what it is that employers actually want these products to do, because we've talked about a host of things, from screening to retention and everything in between.

So, if you could speak a bit to that, and I'd love to hear others, just so that as I'm thinking about this, I have the right sort of set of scenarios in mind.

MR. MRKONICH: Thank you Commissioner Burrows. In fact, I think Dr. Dunleavy correctly noted, the number of analytics departments within HR departments is growing dramatically. So, we're talking about a historical trend and where are companies.

But basically, there are a couple things. One is, HR departments and businesses generally have been impressed by the way they're able to select their senior leaders using very advanced selection techniques, and they said, you know, if we could -- if we have 10,000 employees, for example, if you're a big employer, in an entry level job, if we could get -- select people who would stay on average, six months longer,

we would have happier people, a better trained workforce, we'd deliver better service to our customers.

So, retention tends to be very big topic among people, because what you're trying to do is leverage big data -- use big data to leverage your knowledge, to a level of the company you couldn't afford to do it with before, and it's a very inclusive, everyone wins sort of opportunity.

There are other opportunities with performance management, where I think we're all familiar with the recency bias. If you have an annual performance cycle that ends in December, make sure you screw up in January and February and do a really great job in December. Computer learning, big data learning process can monitor annual performance in a truly annual way, and do that kind of thing. So, you find people who aren't satisfied with the way -- their performance management system, saying, "We've got to be able to do better than this," by tracking what's really going on.

I remember an employer a long time ago said, "Well, when we're in a layoff mode, I want to layoff the person who is not as good." I said, "Yes, I agree," but who is that, and these sorts of tools can help you evaluate that free from the perception and interpersonal bias that you might have based on, you know, who is most like me? Do I need someone who can type 100 words a minute, because I've got to turn things around quickly, or do I want this steady person who is creating 10 letters a day instead of the five that my 100 words a minute person is doing?

All those questions, we have data that help us answer. So, what people are saying is, we can use this to create a more efficient, better, more practical workforce. So, as an employer's HR department, I'm sitting back and saying, "One, how can I find the people to hire?" Right now, it's hard to find people with the skill set you want. I can outreach using some of the hiring techniques. How can I get people who will stay longer, if I'm lucky enough to have some choice? How can I manage them once they get here? If I'm in a reduction force setting, if I'm in a promotion setting, how can I make sure I'm doing the right person?

COMMISSIONER BURROWS: In the little bit of time I have, I wanted to also ask, and I guess this maybe might be something for Ms. Trindel, but for Professor Trindel, but for whoever wants to jump in on this.

You know, one question I have is with the screening piece and that broadening the -- the attempt to broaden the applicant pool on the one hand, but on the other hand,

there have been a lot of reports about attempt at sort of narrowing of advertising, and that the same employer, or in this case, advertiser will send people very different advertising based on their demographic background and that's the kind of thing that is really interesting maybe if you're an advertiser and absolutely not permitted in most cases, if you are talking about employment.

So, I wanted to speak to how we prevent that kind of creep, and what you all are aware of going on in that area.

DR. TRINDEL: I just wanted to mention that there has been discussion about the idea that like passive recruiting, so recruiting of passive applicants, that's supposed to be where it's at because good people have jobs and they're well paid and they're not necessarily on the job market. So, you should go find those people, you hire them, you get good ROI. I understand the logic, and perhaps you do open up your, you know, hiring by doing that. But as you say, Commissioner Burrows, and this is really an important take-home point, is that you might be opening yourself up to different kinds of people who wouldn't apply, usually because those different kinds of people already have jobs and they're happy and they're not looking, I would say.

And so, as you say, you may be narrowing. In fact, and to speak to your second point about marketing and putting jobs in front of people, in a way that might be an issue, there's been a couple studies that have been published, specifically with regard to -- I have one in front of me about Google ads.

So, there has been a publication, or a couple publications about, you know, when you identify yourself as female, the job ads that are shown to you are for lower paying jobs. You identify yourself as African American, or there are some data points fed in, that indicate your race, you're shown a different set of jobs.

So, that's where I really see the crossover very obviously between marketing and employment, because these are marketing of job ads, rather than marketing of products, and obviously, that's problematic.

COMMISSIONER BURROWS: Thank you. I'm out of time.

DR. KOSINSKI: So, if I could just comment on it, because I think it's a very strong and very important point. Yes, I believe that basically targeting of this kind, so if you can see employers biasing their message and basically targeting the messages, let's say targeting women with lower paid jobs, this is simply because they have no -- no better way of judging someone's talent and potential.

So, if you could go beyond just race, gender or income, and we could start targeting people based on their psychological profiles or skills, which big data enables us to do, even in a passive way, we could basically overcome this problem. Not to mention that we could also target people with messages that are attractive to them.

So, let's say in terms of computer science, I spend quite some time at Stanford's computer science department, and there were big efforts to recruit more female students. One of the big issues there is that the entire computer science field is speaking basically to males about things that are interesting to males. The buildings are even -- has been -- it has been shown in a study, are basically built in such a way that males like it more.

Now, if we can target adverts and invite people to job interviews, and distinguish between males and females, we can use it in a malicious way, so basically, target the females with less attractive jobs, but we can also use it in a very productive way, let's say, target women with messages that stresses the attractiveness of this job, specifically for women and thus, increase the rates of their applications.

CHAIR YANG: Thank you. We have come to the close of an incredibly valuable, fascinating meeting. I think each of us up here would have tons of questions for each of you separately to continue this discussion, and we really do see this as a springboard for future discussion.

We're going to have an opportunity now for each of my fellow Commissioners to make a brief statement, to close out the meeting. I would just like to share with you all, how valuable we have found this. There is so much that I think we all need to learn about this space, from the Commission's perspective. We want to learn more about what employers are using, how they're doing it, why, what ways we can be most helpful as part of that process.

We know there is a lot for employers to learn about what is actually happening in some of the algorithms that vendors may be marketing and what -- to shine the light on sort of the black box that's out there.

So, we do hope that we can continue that discussion; that we can engage in this dialogue, and think how we can be most effective in this space. I do want to mention that across our agency, one step that we're taking is forming our own internal working group. Kelly Trindel will be leading that effort, and we will be reaching out to all of you, and I encourage others who may be watching this, to

really reach out to us, as well, because we want -- we know that there are a lot of people who aren't in this room, whose perspectives that we need to hear, from the computer scientists and the HR folks, and others.

So, per our typical practice, the Commission will hold our meeting record open for 15 days. We invite members of the public to submit written comments on any issues or matters discussed at the meeting, and we hope those of you who are interested in staying engaged on these issues discussed today, will submit those comments for the record. You can mail them to Commission Meeting EEOC Executive Officer at 131 M Street, Northeast, Washington, D.C. 20507, or you can email that to commissionmeetingcomments@eoc.gov, and that Commission meeting comments is all one word in the email address.

All comments will be made available to the Commission and to Commission staff working on the matters discussed at the meeting. In addition, comments may be disclosed to the public and by providing comments in response to the solicitation, you are consenting to their use and consideration by the Commission and to their public dissemination.

Accordingly, please do not include any information in submitted comments that you would not want made public -- you would not want made public, such as your home address, telephone number, et cetera. Also note that when comments are submitted by email, the sender's email address automatically appears on the message, and with that, I will turn it over to my fellow Commissioners. Commissioner Barker.

COMMISSIONER BARKER: Thank you Chair Yang. I just want to thank all of you. I'm leaving this meeting today with a whole lot, I think more understanding of the issues and how this whole thing works. One thing that we didn't have a chance to touch on is to what extent big data is used by the federal government and recruitment and selection, and that would be a topic for a whole new meeting.

But thank you again, for lending us the benefit of your time and expertise today.

CHAIR YANG: Thank you. Commissioner Feldblum.

COMMISSIONER FELDBLUM: Thank you again. Thank you for convening this meeting, starting us off on what I think is going to be an important effort. I'm very glad that it was made clear that our digital data footprint is a lot more than social media. I think people tend to forget that. For example, you send an email with your comments, your email address, that's part of your digital data footprint. Sure,

people will be able to figure out your name, as well as your zip code and home address.

I have to say that I'm leaving this hearing a little bit wondering about what best we can do. So, often, we have hearings and it's very clear what we should do as an enforcement agency. We had a hearing on harassment. People asked, let's have legal guidance. Let's have work and prevention. Commissioner Lipnic and I did the select task force, and you know, basically in prevention of harassment.

So, sometimes that's pretty clear. I think this is really a different animal, and I will say that the two things that I have gotten, as to looking forward, and I look forward to people adding to this; that there is potentially something of a convening role that EEOC can play, the sort of blue ribbon panel and best practices. I do think it seems like there are some very specific educational pieces, even without saying what the best practice should be yet, just educating employers about questions they might want to ask, educating computer engineers, how they might want to educate themselves, and but I think third, because we are primarily an enforcement agency, I think we really need to give some thought to what role does law play in this, particularly given some of the more potentially complex issues on disparate impact, using this means of data versus others.

But you have certainly given us a lot to think about, and I have much confidence in our Chair and in our working group, to help guide us on what we should do next.

CHAIR YANG: Thank you. Commissioner Lipnic.

COMMISSIONER LIPNIC: Thank you Madam Chair. Thank you again, for convening this meeting, and thanks so much to all of our witnesses for your excellent testimony today.

I too, am left wondering a bit. One of the things I am wondering in the big data world, is it a good thing to have an Ivy League degree or a bad thing to have an Ivy League degree? I'm not really sure about that right now.

But overall, I look forward to our future engagement on this issue. I agree with the comments of my colleagues and hope that we can come to both some better understanding of how this is all being used in the employment context, and figure out what is the appropriate and necessary thing for the EEOC to be involved in this, but and look forward to working with all of you in the future. Thank you.

CHAIR YANG: Thank you Commissioner Lipnic. Commissioner Burrows.

COMMISSIONER BURROWS: Yes, I'd just like to thank the witnesses again, for your excellent testimony, and it was highly informative. I am very much looking forward to us sort of digging in and just looking at this in more detail, and I do think that as each of you has made clear, this is a tool that can be used either well or very poorly. It can either help advance equal employment opportunity or it can be another thing that causes a lot of litigation, that you know, as a lawyer, I understand why that's necessary, but it's not necessarily the best thing for the employment situation.

So, I am very hopeful that we will be able to actually steer and contribute to this process, and meet those who are experts in the data analytics, with some of the expertise in the employment and the fairness aspect, because I think that's ultimately where we all are trying to go. So, I thank you. I want to thank again, the Chair, and the excellent team of our staff that put this together, as well as Commissioner Lipnic's staff, and I want to associate myself with Commissioner Feldblum's comments about my dear friend G.C. Lopez, we will be very sad to see you go. Thank you all.

CHAIR YANG: And as we close out the meeting, I just want to thank you for your invaluable comments and the time that you spent in helping us really understand this issue better, and I would like to thank you and extend that invitation for you to reach out to us at any time.

With that, is there a motion to adjourn?

COMMISSIONER FELDBLUM: I move to adjourn.

CHAIR YANG: Is there a second?

COMMISSIONER BARKER: Second.

CHAIR YANG: All in favor?

{Chorus of ayes.}

CHAIR YANG: Opposed? Thank you. This meeting is adjourned.

{Off the record at 3:43 p.m.}

