

# **Emerging Technologies in HR Analytics: Opportunities and Challenges**

**Burris Logistics**

Professor Prasanna (Sonny) Tambe

## Some data and teams I have worked with:

**careerbuilder®**



**LinkedIn**

**AngelList**

**Glassdoor**

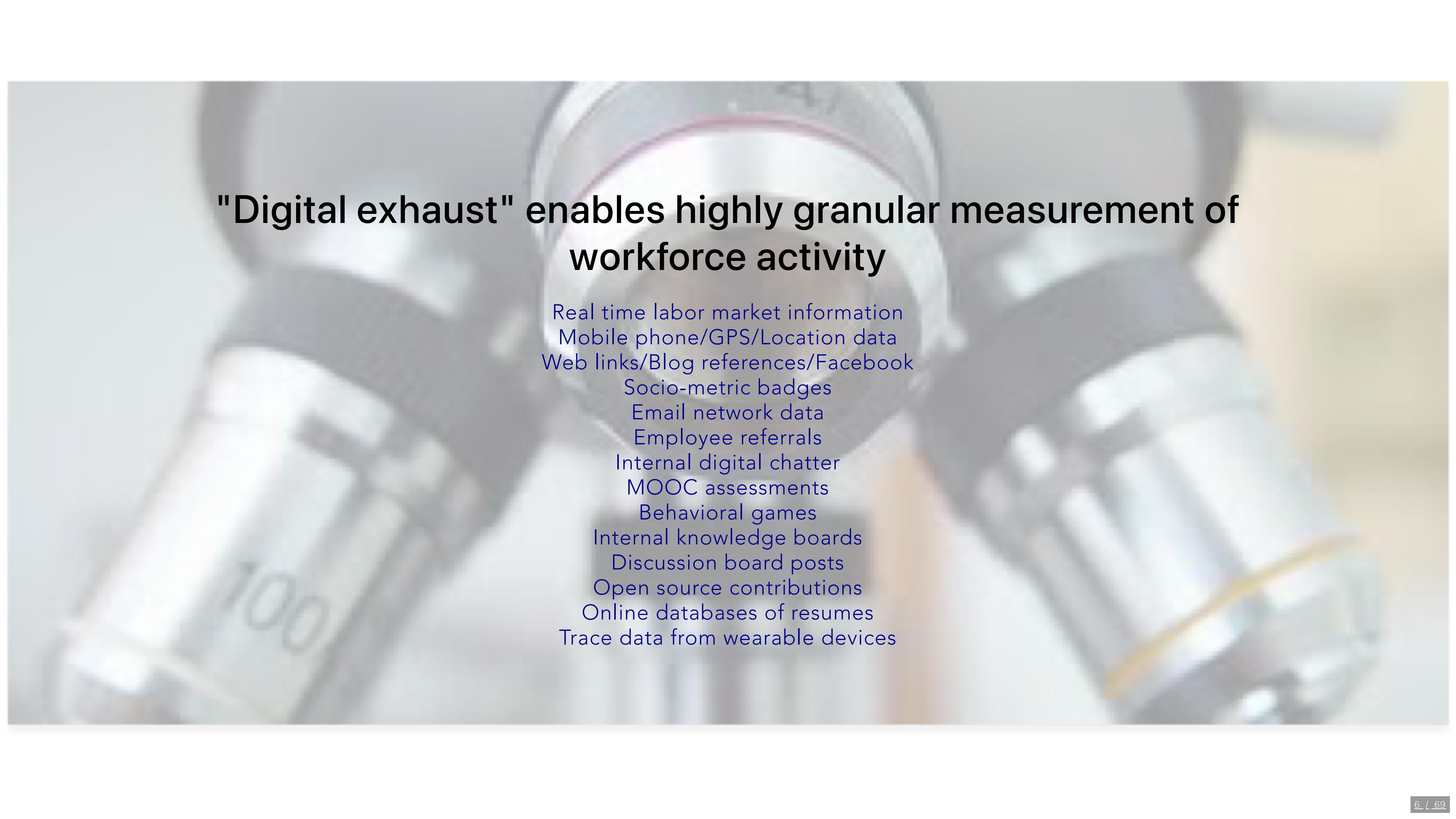
**Up**

## Three topics to be covered today

- Part 1. Why HR is an **exciting** context for AI
- Part 2. Why HR is **difficult** for AI
- Part 3. A potential role for **COVID-19** restructuring

# Part 1. Why HR is an exciting context for AI

# Key question: How does a change in the granularity of observation transform HR decision-making?



# "Digital exhaust" enables highly granular measurement of workforce activity

- Real time labor market information
- Mobile phone/GPS/Location data
- Web links/Blog references/Facebook
- Socio-metric badges
- Email network data
- Employee referrals
- Internal digital chatter
- MOOC assessments
- Behavioral games
- Internal knowledge boards
- Discussion board posts
- Open source contributions
- Online databases of resumes
- Trace data from wearable devices

## Data vs. intuition

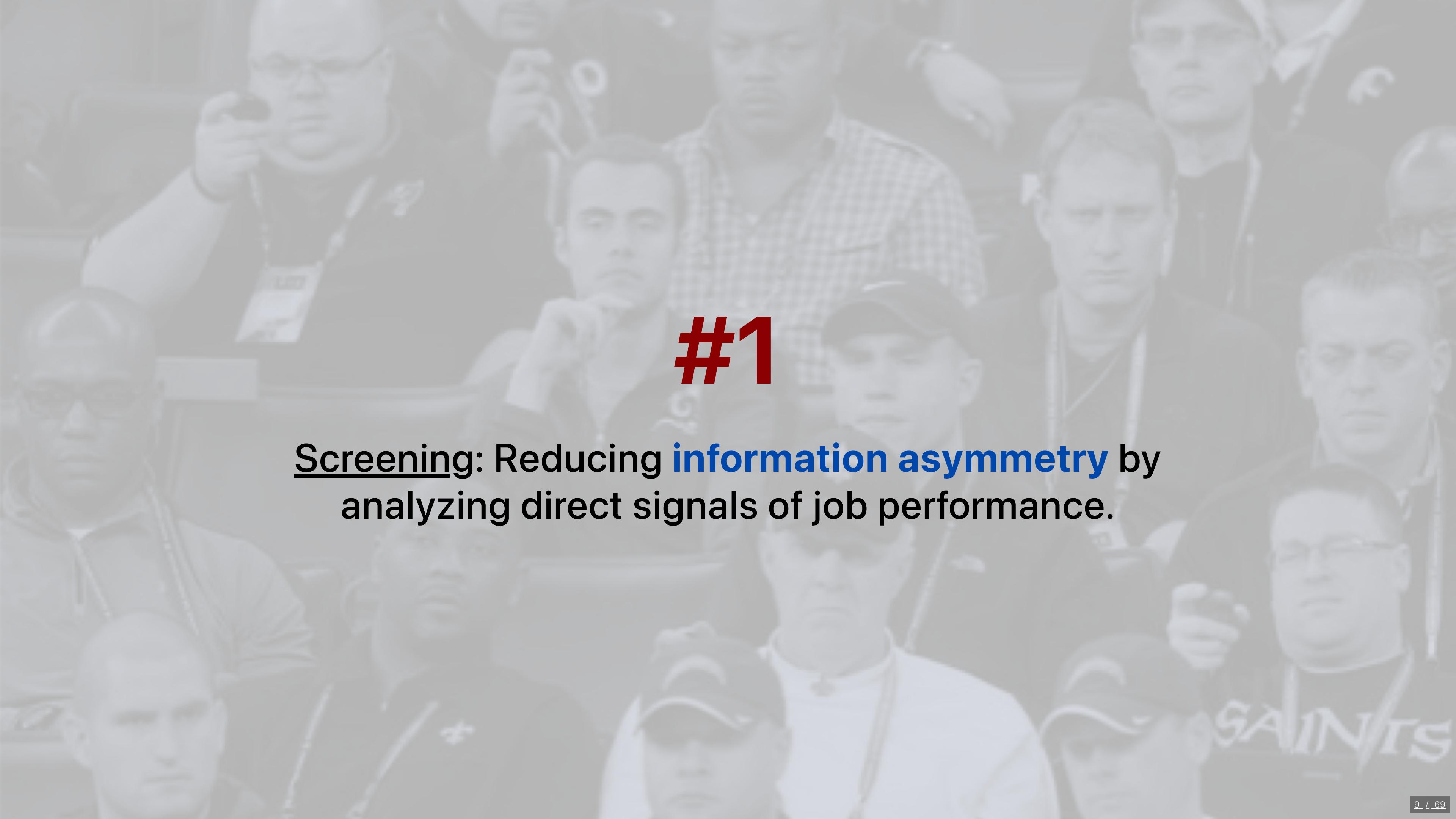
HURRICANE FRANCES was on its way, barreling across the Caribbean, threatening a direct hit on Florida's Atlantic coast. Residents made for higher ground, but far away, in Bentonville, Ark., executives at Wal-Mart Stores decided that the situation offered a great opportunity for one of their newest data-driven weapons, something that the company calls **predictive technology**.

A week ahead of the storm's landfall, Linda M. Dillman, Wal-Mart's chief information officer, pressed her staff to come up with forecasts based on what had happened when Hurricane Charley struck several weeks earlier. Backed by the trillions of bytes' worth of shopper history that is stored in Wal-Mart's data warehouse, she felt that the company could "start predicting what's going to happen, instead of waiting for it to happen," as she put it.

**Source:** [NY Times](#)

# Where is data-driven decision-making useful in HR?

- Recruiting
- Selection
- On-boarding
- Training
- Performance management
- Advancement
- Retention
- Employee benefits



#1

**Screening: Reducing **information asymmetry** by  
analyzing direct signals of job performance.**

A blurred background image showing a group of people in an office environment. Several individuals are seated at desks, working on laptops. The scene is somewhat dimly lit, with strong highlights from computer screens. Some papers and office supplies are visible on the desks.

**There are a growing number of ways to collect  
strong signals of on-the-job performance**





Dropbox



Pinterest



tumblr



fitbit



airbnb



Locu

Bloomberg



Boston  
Scientific



lyft



BLIZZARD  
ENTERTAINMENT

Goldman  
Sachs



Synaptics®



AppDirect



ORACLE

MailChimp



mopub



MINISTRY OF SUPPLY



wayfair

Amicus



a16z



AlchemyAPI™



asana:

Constant Contact®



GE



KPCB



motorola

optimizely



PARSONS



Wingman



Google

StartUp Academy  
Bain Capital Ventures



GitHub



facebook



Meteor

twilio



Etsy



eBay



Rakuten  
Loyalty

wunderground.com®



mongoDB



Firebase



FLIGHTCAR™

intel



SAILTHRU



THOMSON REUTERS



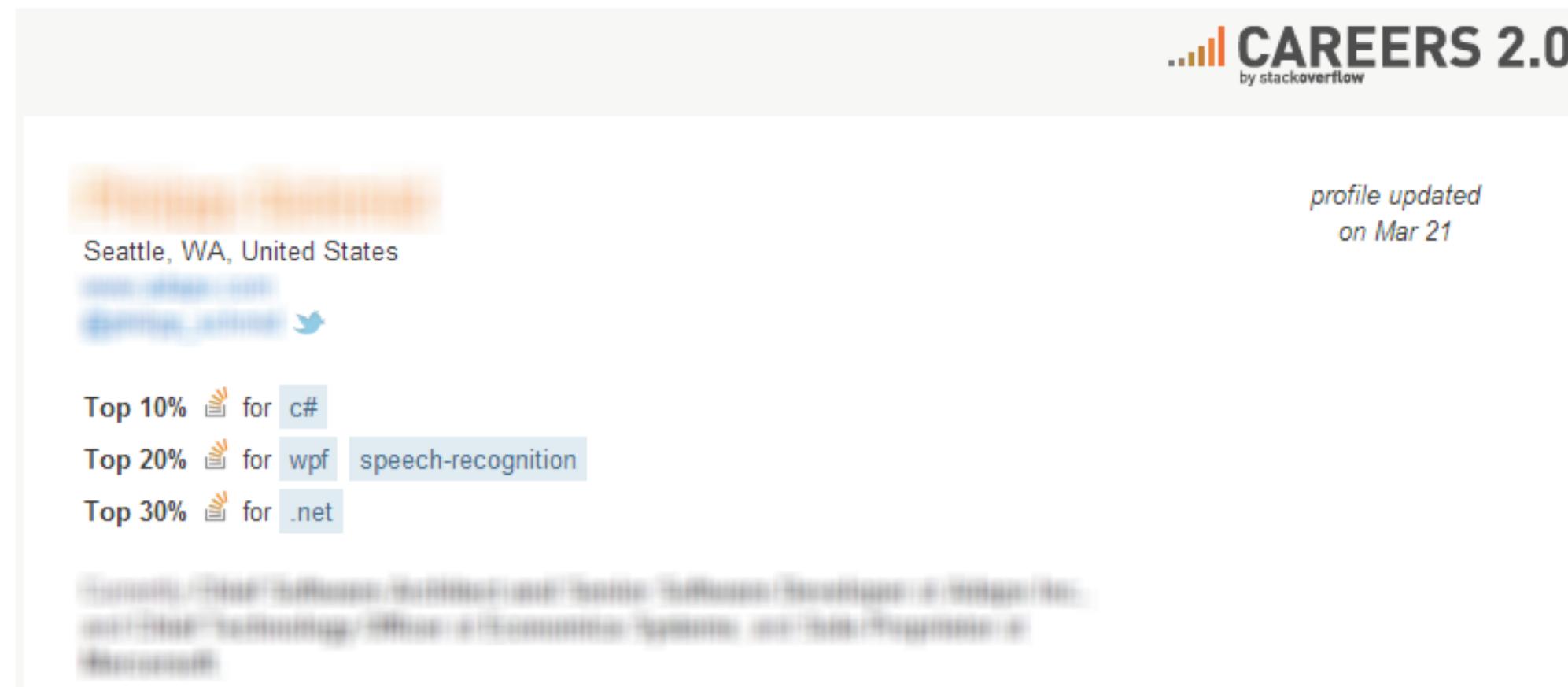
General Catalyst  
Partners

Square

## Why do employers sponsor hackathons?

Effective, but recruiting scale is limited ...

# Scale through archival data: What can online activities tell us about worker quality?



The Simple T...2016) AAUW The gender p...n - LA Times Pando: What...ployees get? Microsoft bus...ness Insider Apple, Micros...h - GeekWire In Drug Merg...ayoffs - WSJ Technology B...s Year - WSJ Elizabeth War...n Wire - WSJ

git - Push existing project into Github - Stack Overflow

list comprehension - Google Search

list comprehension

All Videos News Images Shopping More Search tools

You're speaking our language. Up for a challenge? I want to play No thanks Don't show me this again

About 1,010,000 results (0.44 seconds)

A **list comprehension** is a syntactic construct available in some programming languages for creating a **list** based on existing **lists**. It follows the form of the mathematical set-builder notation (set **comprehension**) as distinct from the use of map and filter functions.

[List comprehension - Wikipedia](https://en.wikipedia.org/wiki/List_comprehension)  
[https://en.wikipedia.org/wiki/List\\_comprehension](https://en.wikipedia.org/wiki/List_comprehension)

[About this result](#) • [Feedback](#)

**Python: List Comprehensions**  
[www.secretix.de/olli/Python/list\\_comprehensions.hawk](http://www.secretix.de/olli/Python/list_comprehensions.hawk) ▾  
Everything else is output from Python. Python supports a concept called "list comprehensions". It can be used to construct lists in a very natural, easy way, like a mathematician is used to do. The following are common ways to describe lists (or sets, or tuples, or vectors) in mathematics.

**5. Data Structures — Python 2.7.12 documentation**  
<https://docs.python.org/2/tutorial/datastructures.html> ▾  
Sep 20, 2016 - Remove the first item from the list whose value is x. .... A list comprehension consists of brackets containing an expression followed by a for ...

**Python List Comprehensions: Explained Visually - Trey Hunner**  
[treyhunner.com/2015/12/python-list-comprehensions-now-in-color/](http://treyhunner.com/2015/12/python-list-comprehensions-now-in-color/) ▾  
Dec 1, 2015 - Python List Comprehensions: Explained Visually Dec 1st, 2015 10:30 am | Comments Sometimes a programming design pattern becomes ...

**List Comprehensions - Python 3 Patterns, Recipes and Idioms - Read ...**  
[python-3-patterns-idioms-test.readthedocs.io/en/latest/Comprehensions.html](http://python-3-patterns-idioms-test.readthedocs.io/en/latest/Comprehensions.html) ▾  
A list comprehension consists of the following parts: An Input Sequence. A Variable representing members of the input sequence. An Optional Predicate expression. An Output Expression producing elements of the output list from members of the Input Sequence that satisfy the predicate.

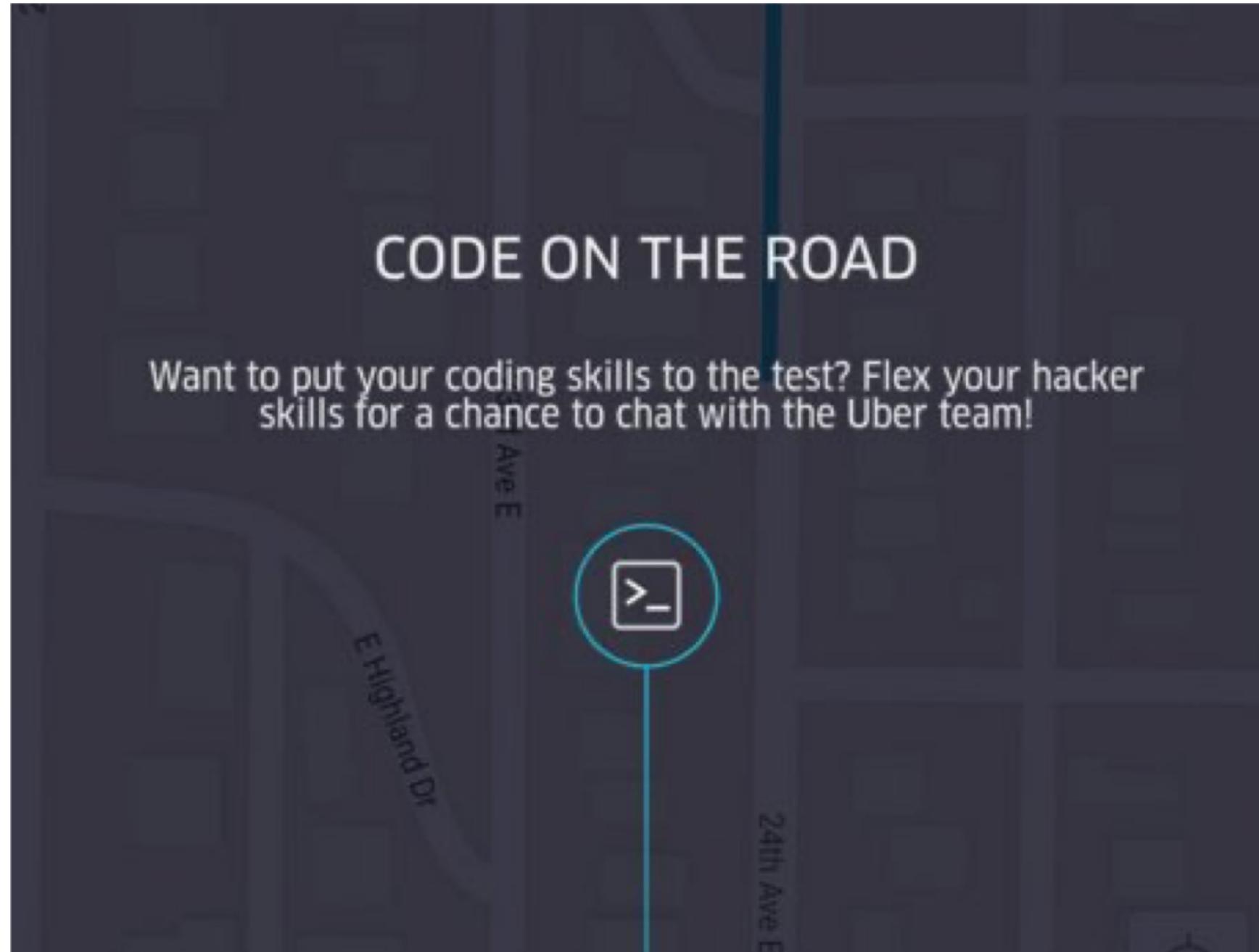
**List comprehension - Wikipedia**  
[https://en.wikipedia.org/wiki/List\\_comprehension](https://en.wikipedia.org/wiki/List_comprehension) ▾  
A list comprehension is a syntactic construct available in some programming languages for creating a list based on existing lists. It follows the form of the ...

**Python Tutorial: List Comprehension - Python Course**  
[www.python-course.eu/list\\_comprehension.php](http://www.python-course.eu/list_comprehension.php) ▾  
We learned in the previous chapter "Lambda Operator, Filter, Reduce and Map" that Guido van Rossum prefers list comprehensions to constructs using map, ...

**List comprehension - HaskellWiki**  
[https://wiki.haskell.org/List\\_comprehension](https://wiki.haskell.org/List_comprehension) ▾  
Apr 23, 2016 - List comprehensions are syntactic sugar like the expression import Data. ... Strings in Haskell are lists of characters; the generator, c <- s,

**List Comprehensions in Python - Python For Beginners**  
[www.pythontutorforbeginners.com/basics/list\\_comprehensions\\_in\\_python/](http://www.pythontutorforbeginners.com/basics/list_comprehensions_in_python/) ▾

# Filling the top of the funnel at Uber



× HACKER CHALLENGE    × HACKER CHALLENGE    × HACKER CHALLENGE

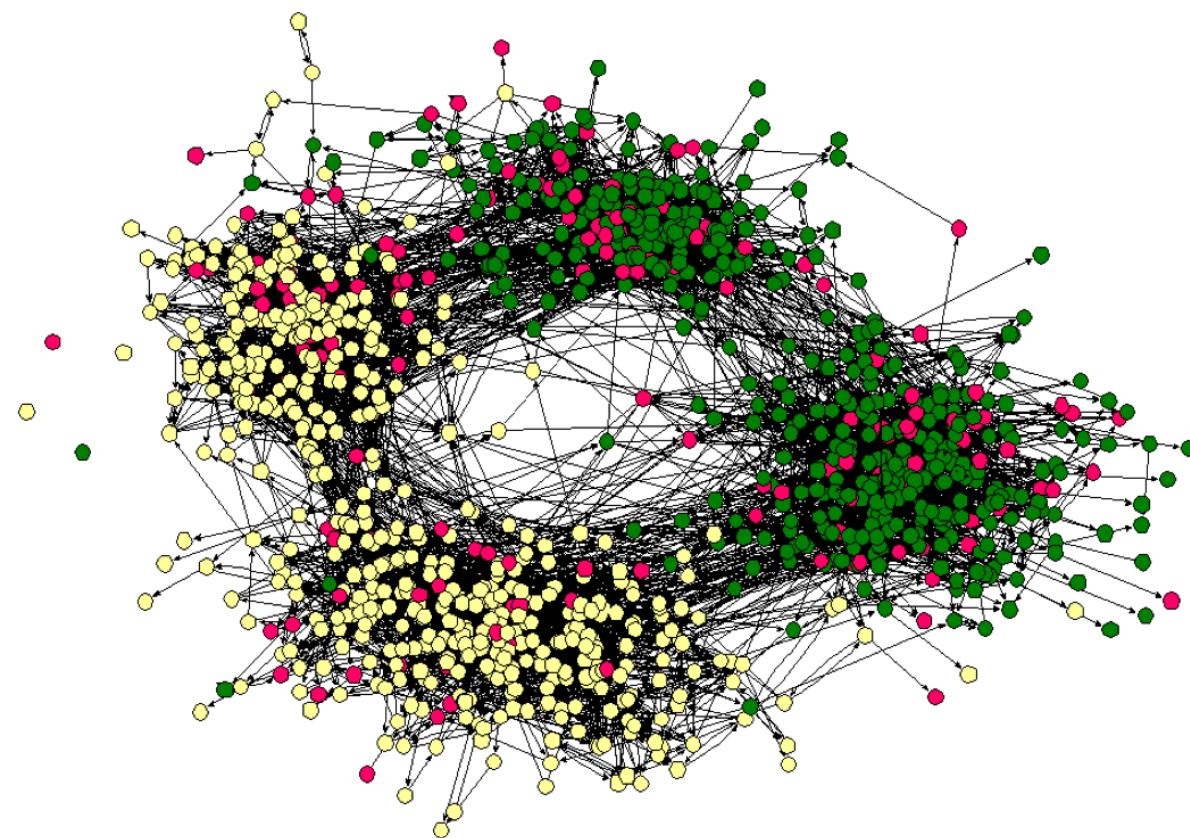
```
1 TIME LEFT: 58s          SCORE: 00110 1 TIME LEFT: 11s          SCORE: 00110 1 You are pretty good at this.
2 ======          ====== 2 ======          ======
3 4 Question 2          4 CORRECT!
5 5 You are helping design our dispatch          6 UR133tl33tl33tl33tl33tl33tRU
6 system. When a trip is requested, you          6 U'cs.           .cs'U
7 need to return the driver with the          7 U  'cs.           .cs'  U
8 shortest ETA. If a driver canceled,          8 U  'cs.           .cs'  U
9 the next driver with the shortest ETA          9 U  'cs.           .cs'  U
10 out of k ETAs is dispatched instead.          10 U  cs'           'cs.  U
11 11 Q1 SCORE          110 11 U.cs'           'cs.U
12 12          / \          12 U!CMSBYEBYEBYEBYEBYEFTDIU
13 Which data structure would you use to          13
14 store the k drivers and dispatch the          14 Uber is hiring engineers with your
15 driver with the shortest ETA?          15 talent. Would you like us to send you
16          -----          16 more information on what it's like
17          -----          17 working at Uber?
18          -----          18
19 [A] Array          19 [ NO ]      [ YES ]
20 [B] Heap          20
21 [C] Hash Table          21
22 [D] Binary Search Tree          22
```

# #2

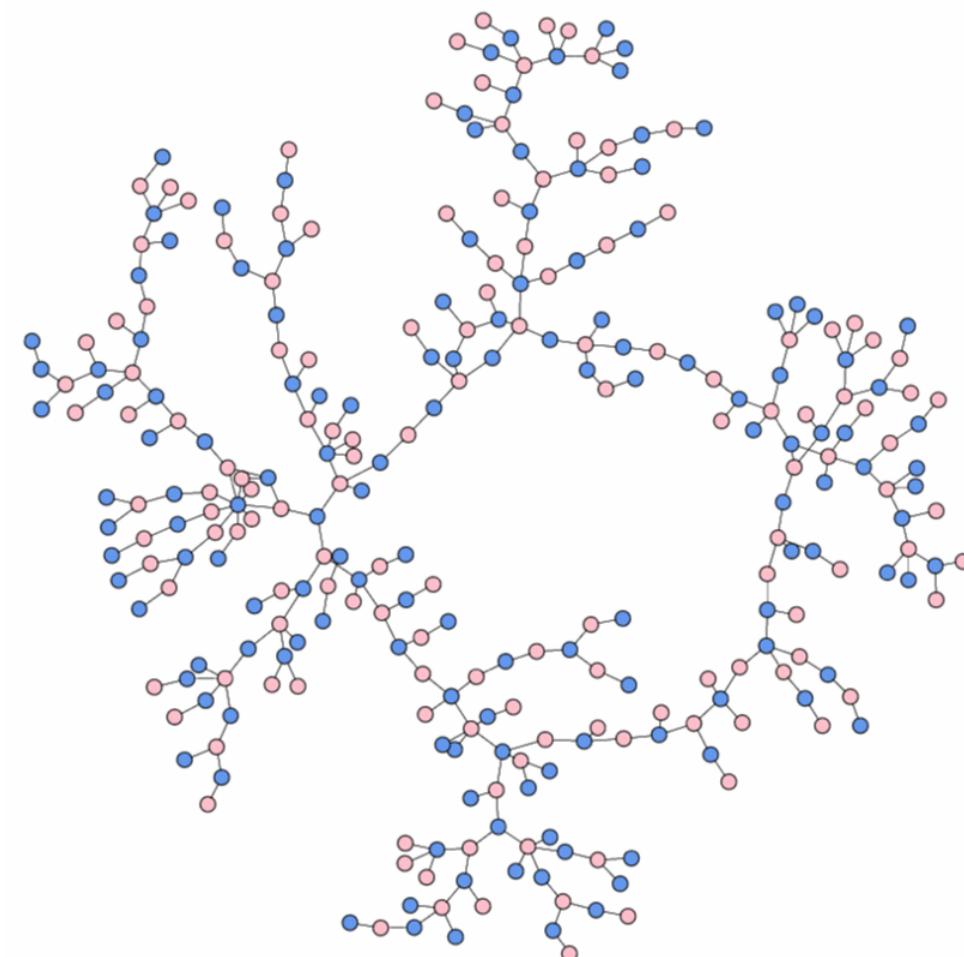
Using granular data for better management of information work.

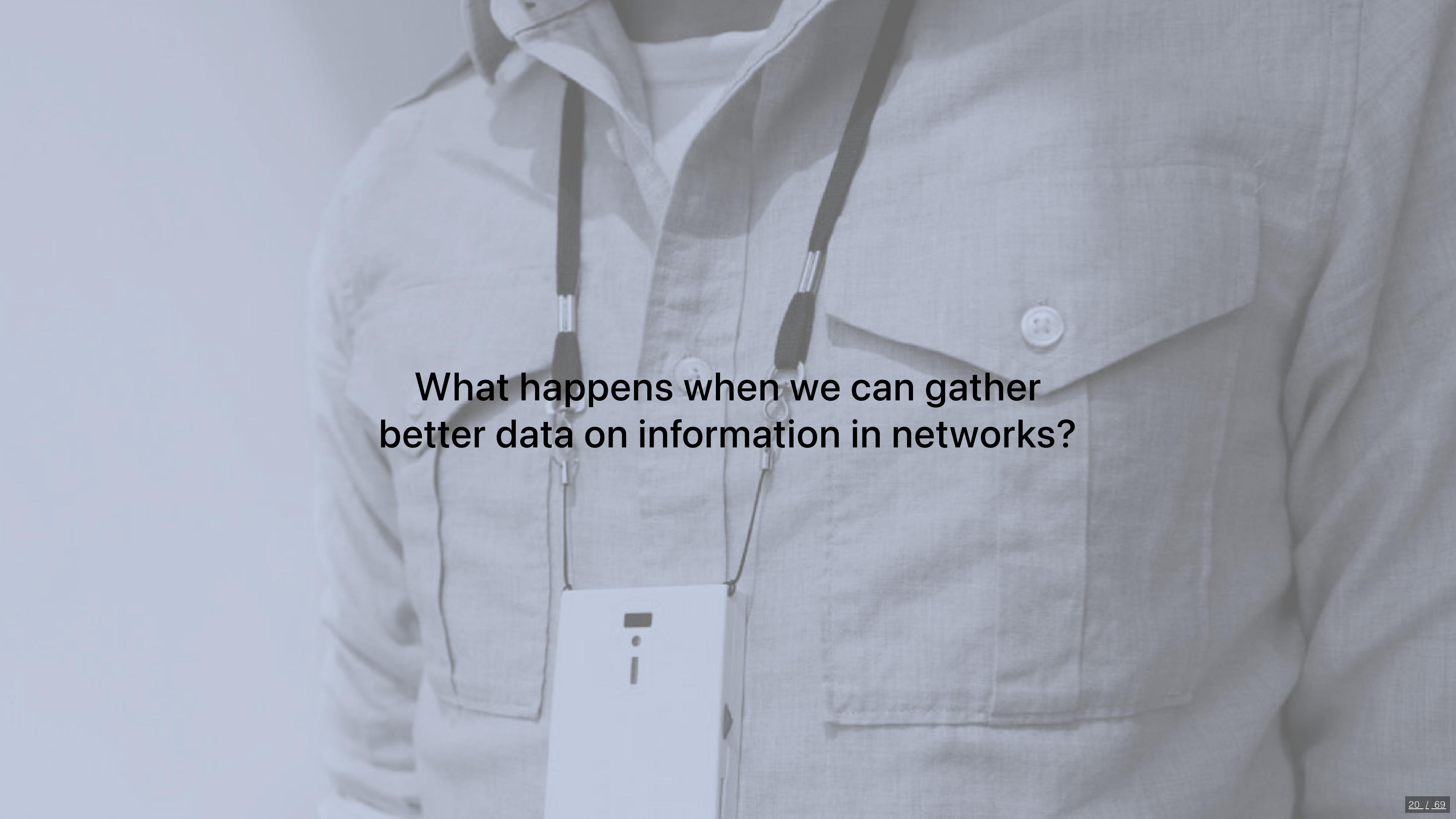
**Example:** Tracking the flow of information in services work

# How might new information flow through this network?



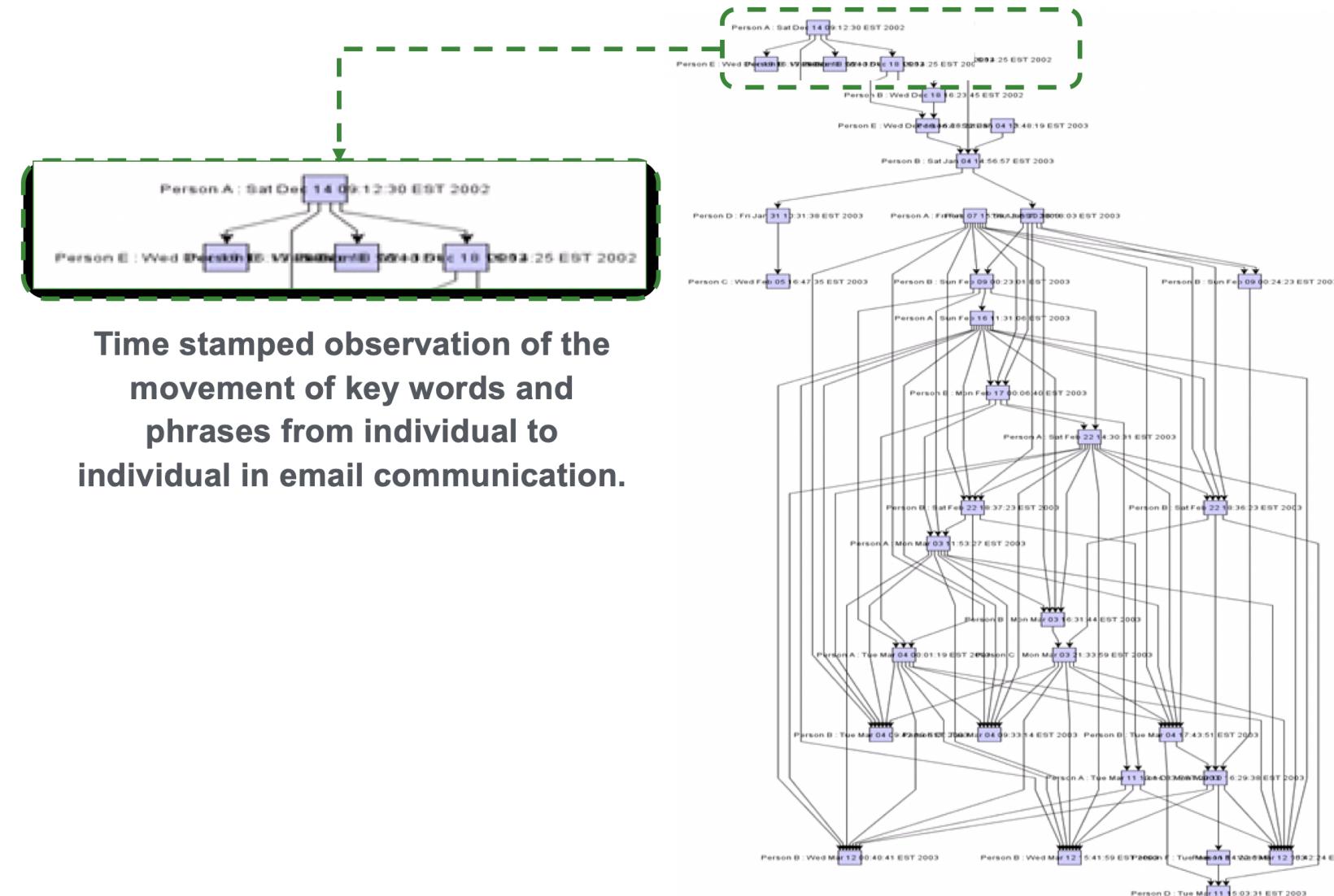
How might new information flow through this network?





What happens when we can gather  
better data on information in networks?

# We can begin to connect information work to concrete outcomes



**Source:** Aral, Brynjolfsson & Van Alstyne "Productivity Effects of Information Diffusion in Networks."

## Connecting information work to revenue

- Access to information diffusion predicts individual productivity.
  - Each **additional 'keyword seen'** is associated with about **\$70 of additional revenue** generated.
- Seeing information sooner also predicts higher productivity.
  - An additional word seen **within the first week** of its emergence in the network is **worth ~ \$321**.
  - An additional word seen **within the first month** of its emergence in the network is **worth ~ \$115**.

## Part 2. Why HR is challenging for AI

**What types of problems may arise?**

## Contrast HR with **marketing** or **finance**

- The objectives in marketing and finance are relatively clear.
- The outcomes are easily measurable (e.g. sales or clicks).
- User activity generates very large data sets (number of clicks or purchases).
- In terms of social norms, it is accepted that companies are going to try to sell more of their product or to make more money.



**Renewal by Andersen**  
LLC

LIKE PAGE

Sponsored • 3

Are you a craftsman who wants to bring your skills and expertise to one of the premiere home-improvement service providers in the country? We're hiring installers in your area:



## Now Hiring Installers

Love working in the great outdo...  
[surveymonkey.com](http://surveymonkey.com)

APPLY NOW

## HR constraints commonly clash with algorithms

Supplying ad engines with job ads led to instant EEOC violations

Source: Lambrecht and Tucker 2018

## **SIX** specific challenges HR poses for AI/ML

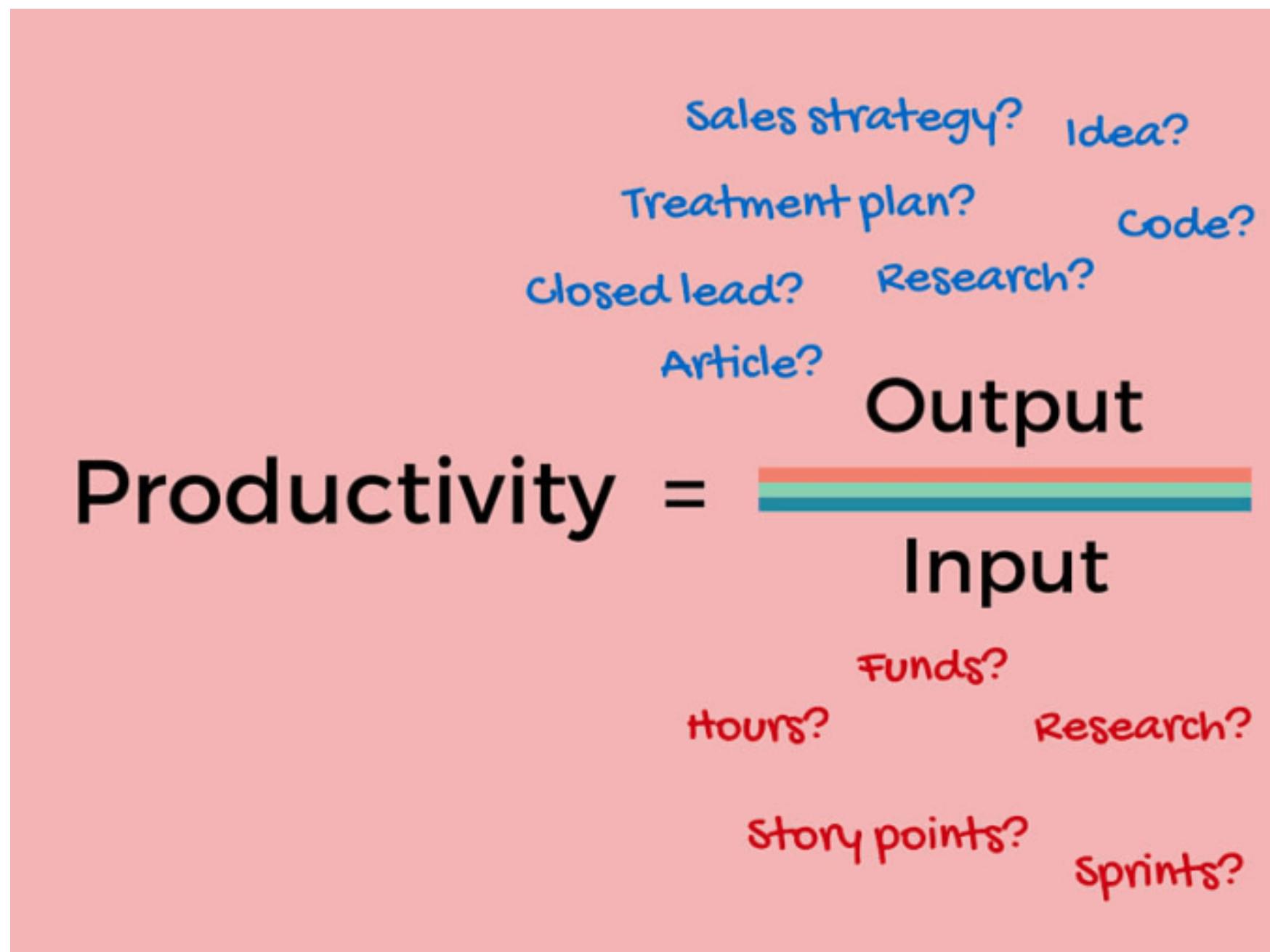
$$y = X\beta + \epsilon$$

Problems that arise when transporting a prediction model to an HR context.

## **SIX** specific challenges HR poses for AI/ML

1. Defining a "good employee"
2. Potential for bias
3. Compliance and legal liability
4. How employees react to the use of algorithms
5. Assembling the right data
6. Concerns about privacy

# 1. Defining the dependent variable is difficult



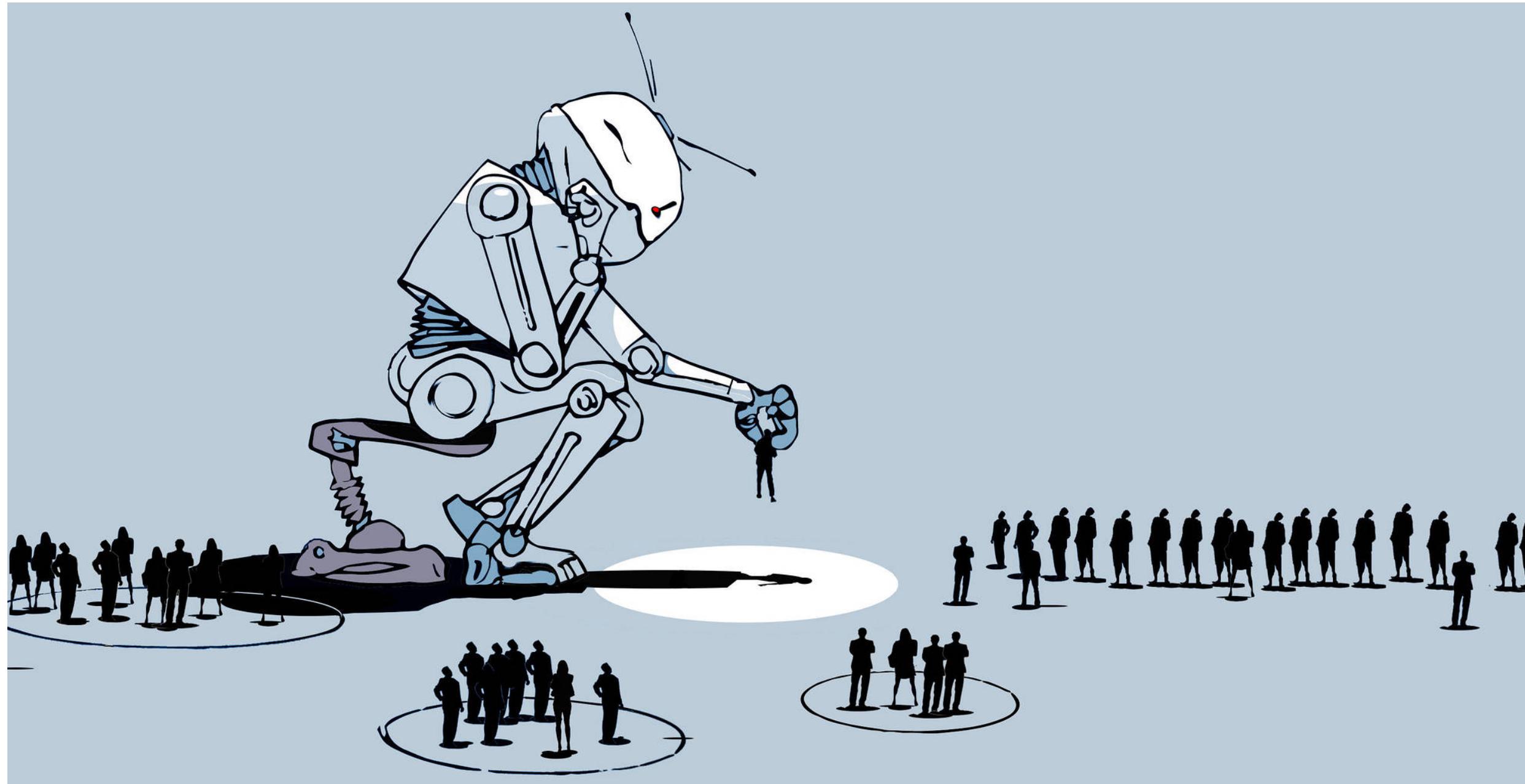
## What makes an employee good?

- Many dimensions to this construct, job performance is interdependent among employees, and not everyone agrees what good means in this context.
- Even if we know the answer, digital traces of what we would like to measure may not exist.
- Traditional measures, such as performance appraisal scores, have been **roundly criticized** in the literature (e.g. Coen and Jenkins 2002).

Almost no HR practitioners we have spoken with could define what an ideal measure should look like, let alone how a proxy measure could be generated from the data that is available.

--Especially problematic when proposing a lengthy, resource-intensive project.

## 2. Attribution and bias



**Source:** The Wall Street Journal

## The economics of discrimination

- Key hiring problem: Information asymmetries in labor markets. Workers know their own abilities but employers do not.
- Group variables are used as a **proxy** for unobservable or unknowable individual characteristics such as productivity.
- Can create a tradeoff between efficiency and equity.

## How does this translate to EEOC guidelines?

- There are **many** open legal questions here.
- Closest existing guidance is the Uniform Guidelines on Employee Selection Procedures (UGESP).
- Selection tests (including scoring models) cannot cause adverse impact to protected groups.
- Adverse impact determined by the rate of selection of each group relative to the highest rate selection group ("the **four-fifths rule**").

### 3. The importance of the paper trail

- Algorithms are biased but managers are also biased (affinity bias, confirmation effects, endowment effects, etc.)
- **Ban the box** : High-profile campaign to remove the box asking if applicants have a criminal record.
- Makes it easier for ex-offenders to get a job and makes them less likely to re-offend.

## Effects of throwing away this information?

- Decreased employment opportunities for low-skill black and hispanic workers (Doleac and Hansen, 2016; Agan and Starr 2017).
- In the absence of the box, employers infer this information from demographic variables.
- Biased algorithms are easy to describe, hard to resolve

Why are people preferable to  
algorithms in this context?

## The importance of the paper trail

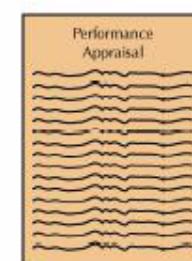
- **Key point:** For most civil action suits, the burden of proof lies with the plaintiff.
- In cases regarding termination of an employee's contract, the **burden of proof is reversed**.

## This makes **ML interpretability** critical

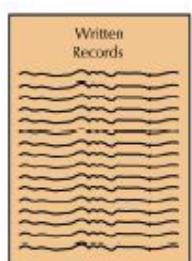
- Plaintiffs' burden of production:
  - Belong to a protected class
  - Must be qualified for the job
  - Experience an adverse employment outcome
  - Job goes to someone not of the protected class
- A principal HR function is to meet the burden of production when faced with allegations of wrongful termination based on race, gender, or age.
- **Contrast this with the explainability and fairness challenges that arise in most algorithmic contexts**

# This makes the paper trail very important

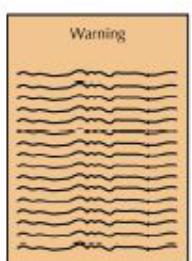
## *Keys for Defense in Wrongful Discharge: The “Paper Trail”*



**Performance Appraisal**  
Make sure performance appraisals give an accurate picture of the person's performance.



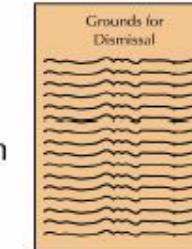
**Written Records**  
Maintain written records on behaviors leading to dismissal.



**Written Warning**  
Warn employees in writing before dismissal.



**Group Involvement**  
Involve more than one person in termination decision.



**Grounds for Dismissal**  
Put grounds for dismissal in writing.

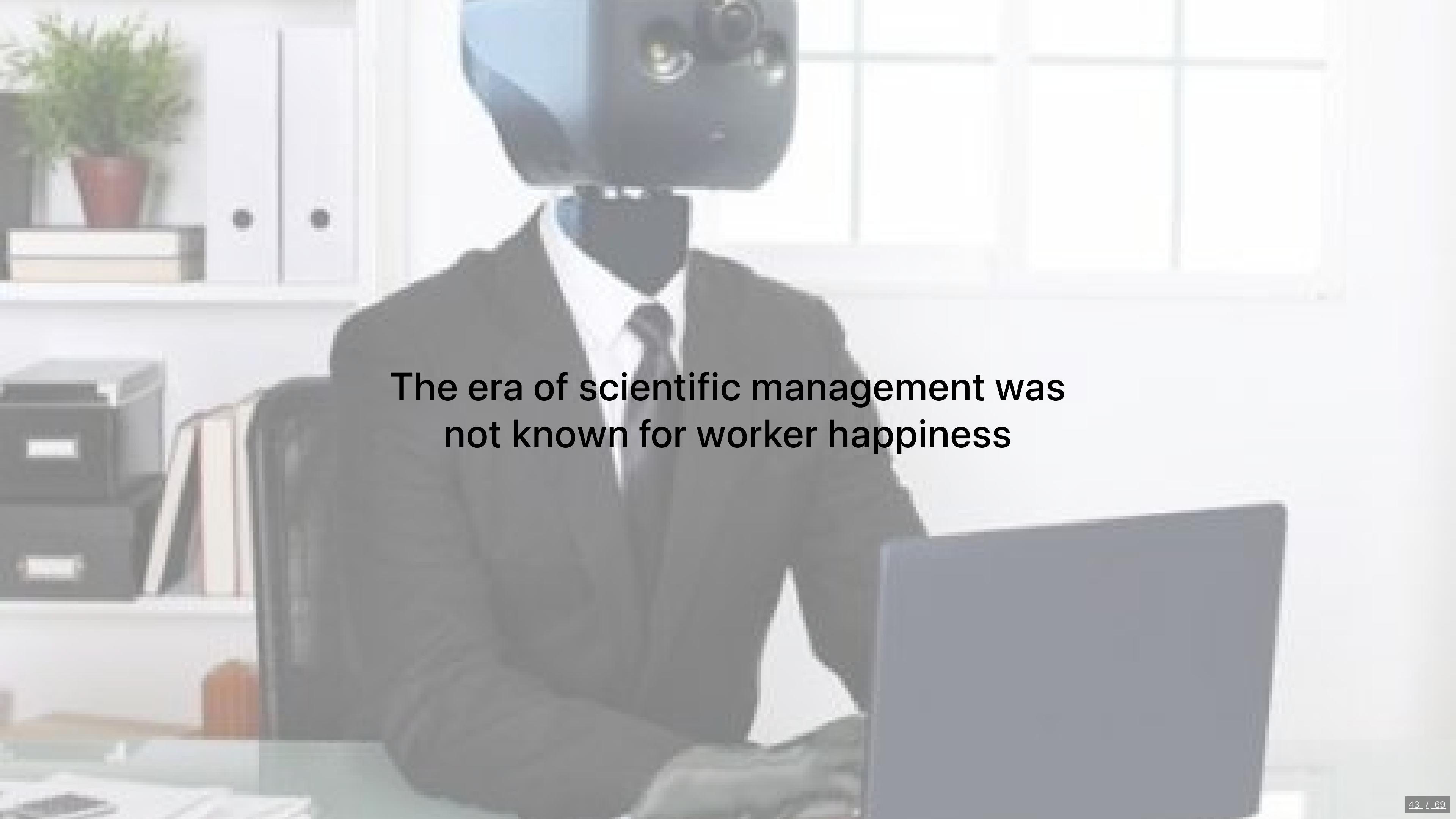
**Source:** Human Resource Management, Tenth Edition, Mathis and Jackson

# New data privacy laws raise the stakes for explainability

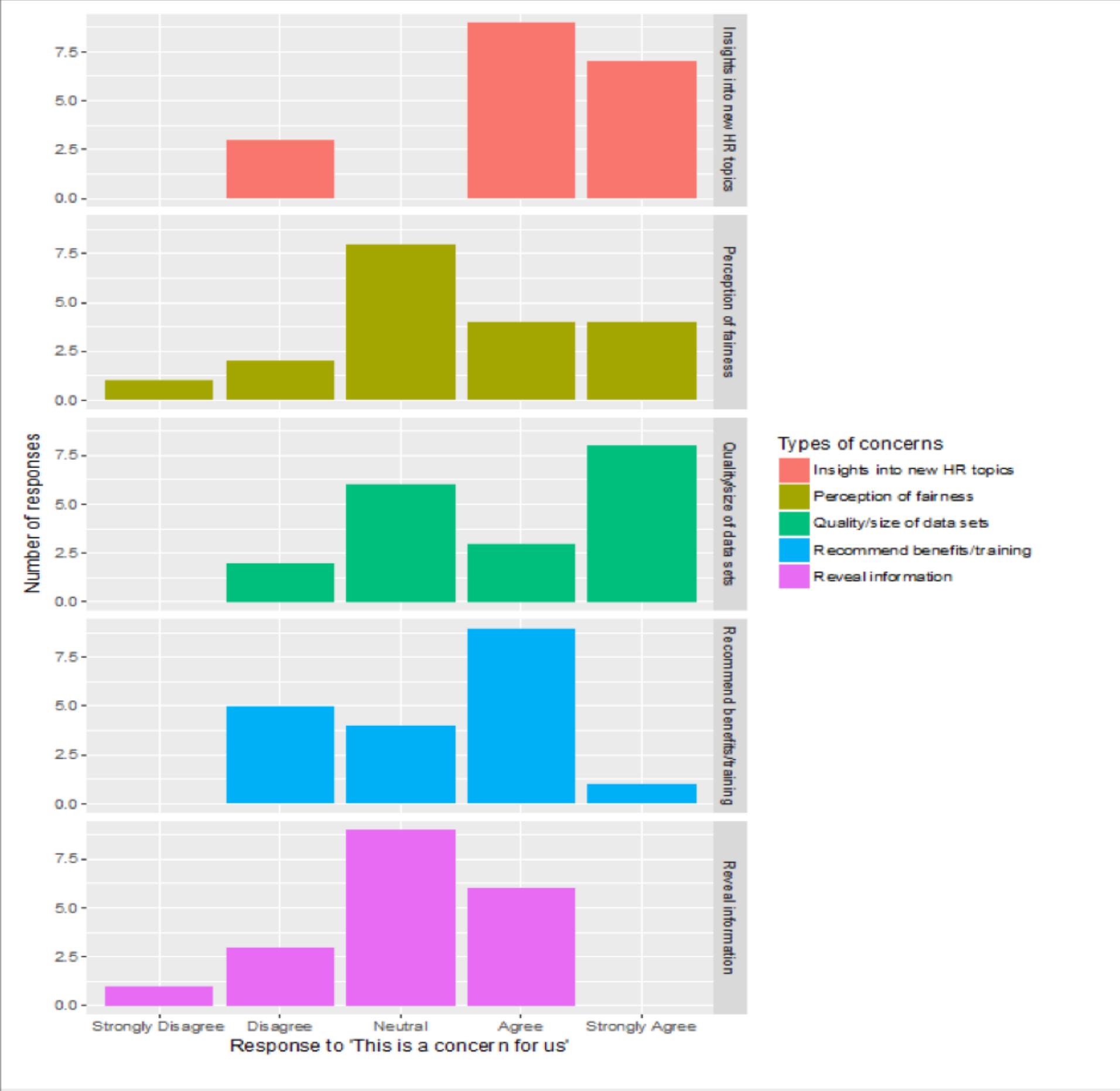
- The **GDPR** does not prohibit machine learning, but imposes a heavy burden on compliance.
  - ... **Article 22**: Must opt-in to automated decision making.
  - ... **Article 17**: Right to be forgotten.
  - ... **Article 13**: Has a right to a meaningful explanation of logic.
- The guard rails for HR data usage that exist even in the most progressive companies are formidable and becoming increasingly difficult to navigate.

## 4. Implementation: Employees vs. algorithms

- Employees may not react well to algorithmic decisions, especially if the news is bad.
- Like with any system, employees may adjust their behavior once they learn the incentives, making data inauthentic.
- A great deal of coordination within the firm is based on "relational contracts".
  - You let me leave early yesterday, so I don't mind staying a little later today ..."

A black and white photograph of a man in a dark suit and tie, sitting at a desk in an office. He is looking down at a document or piece of paper on the desk. Behind him is a window with a grid pattern, and to his left is a filing cabinet and a potted plant.

**The era of scientific management was  
not known for worker happiness**



## 5. Assembling the right data sources

## It is difficult to get HR data at scale

- There are **few observations per worker** and algorithms perform poorly when predicting rare outcomes (e.g. toxic employees).
- Firms often use different HR vendors for different tasks, and it is not easy to merge data across sources.
- Unlike accounting, there is no standard list of HR variables.

# HRIS systems are challenging to work with

HRIS	HCM	HRMS
Recruiting / ATS	HRIS	HCM
Core HR	Onboarding	Payroll
Benefit Admin / OE	Performance	Time & Labor
Absence Management	Position Control	
Compensation	Succession	
Training	Salary Planning	
Workflow	Global	
Self-Service	Analytics	
Reporting		

## HRIS systems are challenging to work with

- HR information systems are **data islands**.
- Difficult to integrate systems when project goals and outcome variables are poorly defined.
- Further compounded by compliance differences across borders.
- **Combined with the lack of a robust DV, this makes exploration and learning costly.**

## What about vendors?

- Vendors have the ability to aggregate data from many employers to generate superior performance, which can alleviate the small data issue (and the sample bias issue.)
- But challenges remain:
  - For each employer, a question that often arises is "**how distinct is our context**"?
  - This raises questions, skepticism related to prediction accuracy.
  - There are legal (privacy) issues as well about employee data sharing.

## 6. Privacy

Data capture promises new battles over **worker privacy**.













5.89 MI

We already **trade privacy for discounts** in many markets



**Great drivers get  
GREAT RATES with  
Snapshot®**

As with credit histories, **opting-out** may not be a choice



As digital footprints become more important,  
growing potential for structural labor market inequality

white many people, many IP addresses

red few people, many IP addresses

blue many people, no IP addresses

# The landscape of data privacy is shifting

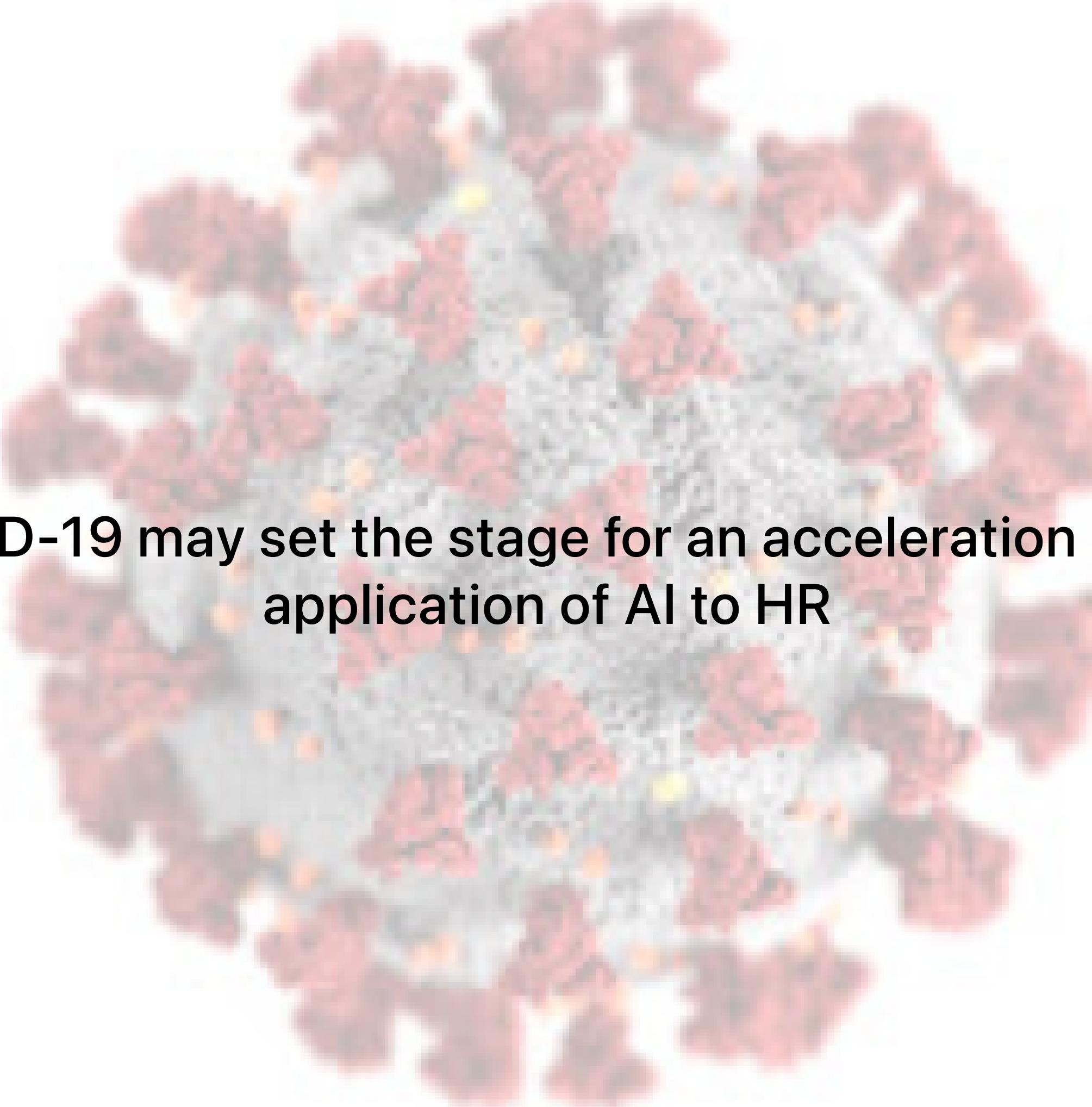
- GDPR
- CCPA
- Personal Information Protection Law (China)

**How might data privacy get balanced with AI benefits?**



**How might data privacy get balanced with AI benefits?**

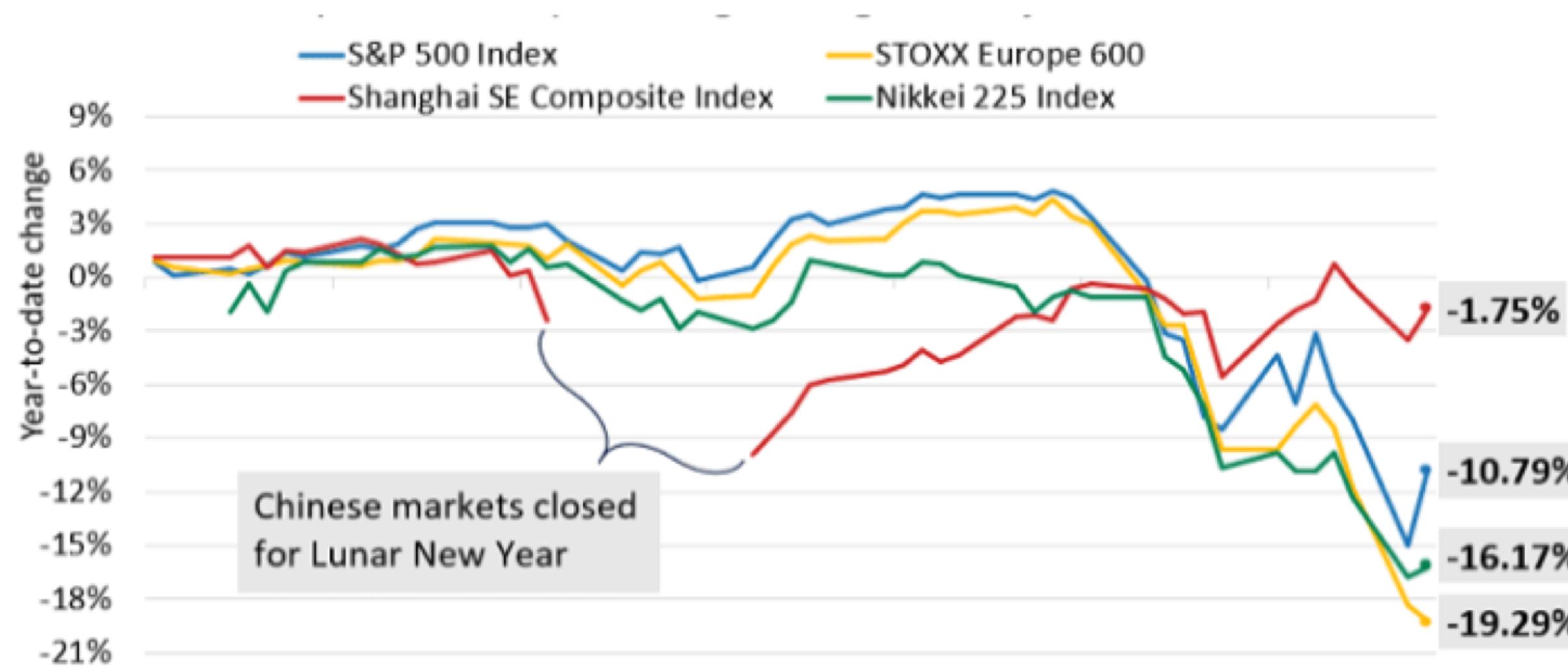
## **Part 3. A potential role for COVID-19 restructuring**



**COVID-19 may set the stage for an acceleration in the application of AI to HR**

**How?**

# Economic shocks accelerate technology adoption



# More collection of digital trace data



# Estimote launches wearables for workplace-level contact tracing for COVID-19

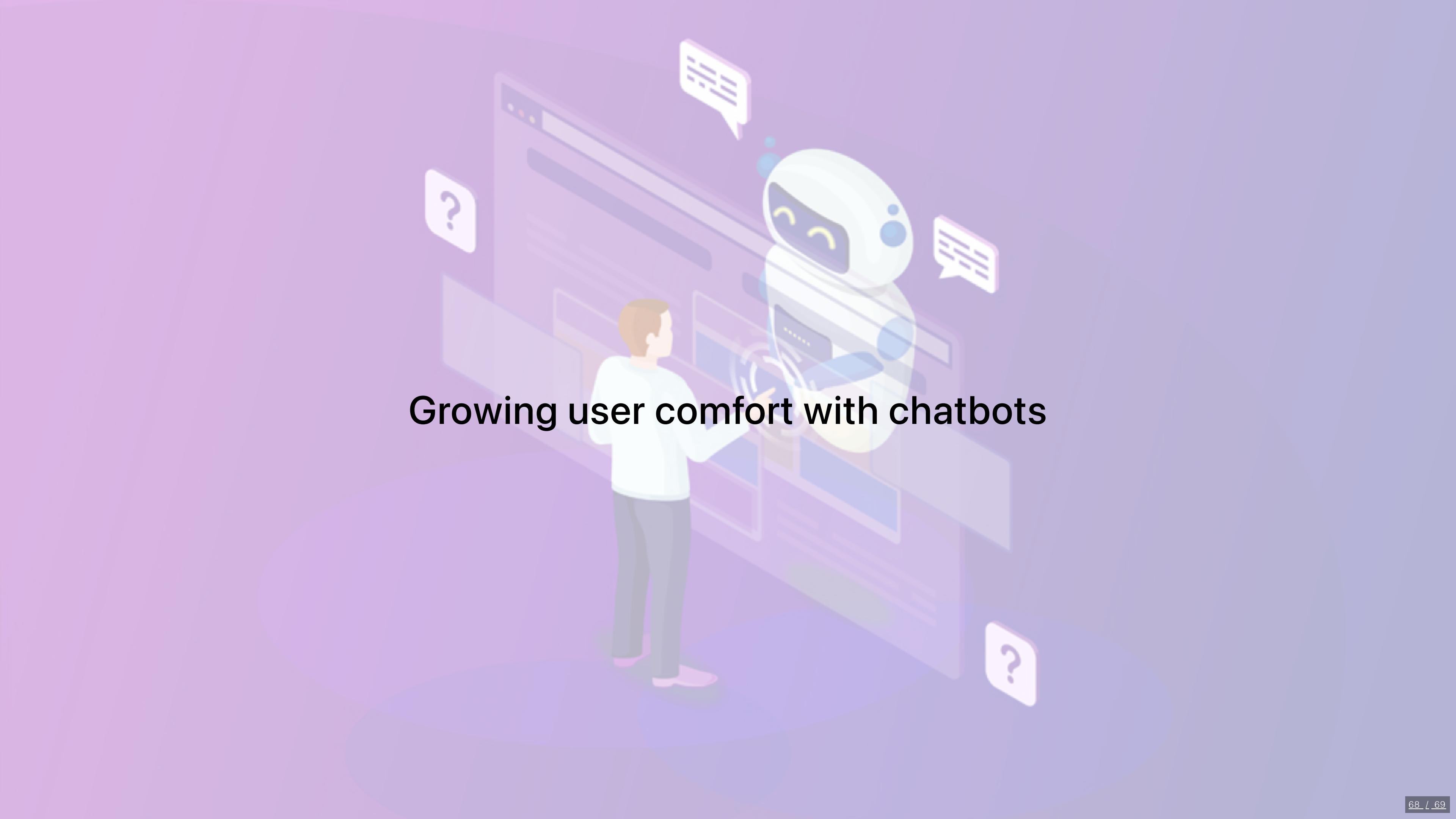
Darrell Etherington @etherington / 1:00 pm EDT • April 2, 2020



Comments



Increased use of workplace sensors



**Growing user comfort with chatbots**

In **sum**, applying AI to HR is very promising but poses extremely difficult challenges ...

... but **COVID has set the stage for rapid changes.**

Thank you! [tambe@wharton.upenn.edu](mailto:tambe@wharton.upenn.edu).