Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Where Does Bias Come From?

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

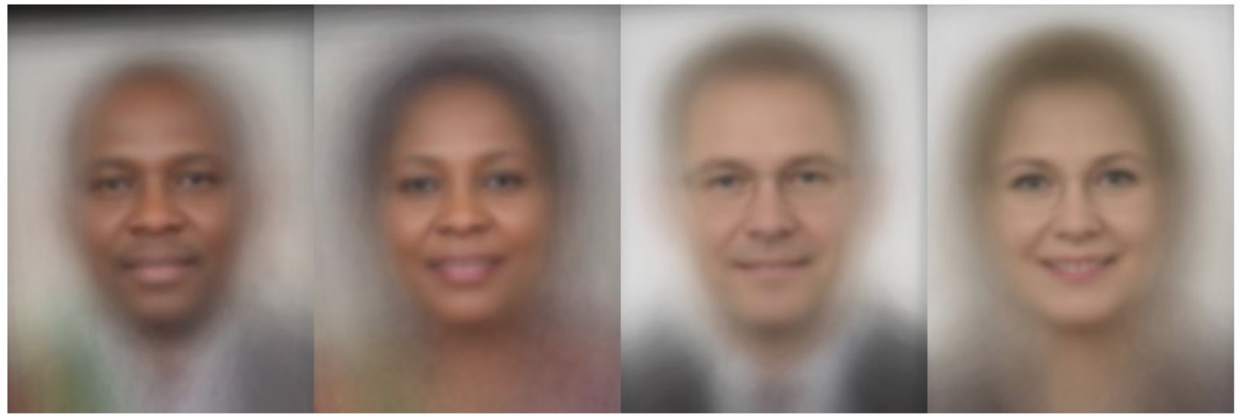
Where Does Bias Come From?

- Machine learning algorithms learn the best mapping from the training data
- If prior decisions encode historical bias, algorithms will necessarily learn to be biased as well

Example: Data Adequacy Bias

- Bias does not have to be intentional
- The lack of training data for some populations leads to poor predictive performance for certain groups

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



Machine Learning Bias

- Bias happens all the time, even inadvertently
- Example: STEM job advertisements

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Bias is Difficult to Manage

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Machine Learning Bias

- Bias is difficult to manage
  - Involves value judgements, which requires a holistic view

Machine Learning Bias

Example: Propublica/Northpoint

- Northpoint used algorithms to make bail decisions
- Propublica reported that black defendants were being unfairly detained
- Northpoint said that among those who actually went on to offend, they were being detained equally
- Is it worse to let the guilty go free or unfairly punish the innocent?
  - Blackstone's ratio

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Bias is Complicated to Fix

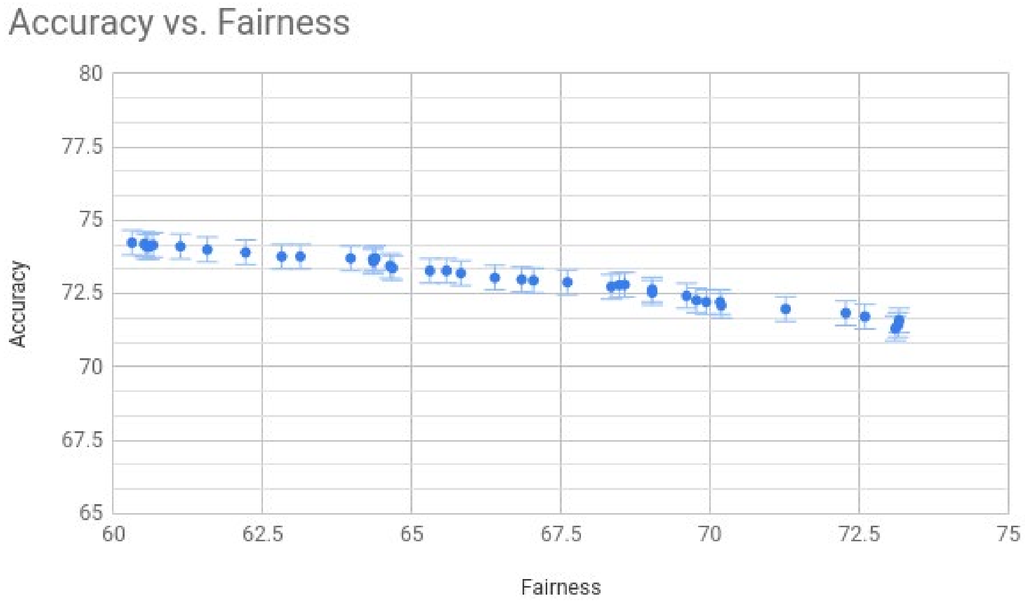
Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Bias is Complicated to Fix

- When we think about resolving bias, it can come at the cost of affecting other performance metrics we care about, like accuracy
- Satisfying conceptions of fairness for some groups requires sacrificing it for others

Model	Accuracy	Fairness Value
Removal of Sensitive Attribute	77	38
fairNN	71	70

Implementation of fairNN: Fairness value almost doubled, but accuracy did decrease.



95% confidence interval plot for accuracy vs fairness

Bias is Complicated to Fix

- Getting better training data is hard

Bias is Complicated to Fix

- Who should even be dealing with bias?
- New positions
  - Chief data officers
  - AI councils

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Some Approaches to Fixing Bias

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Some Approaches to Fixing Bias

- Improve training data
  - Change labels (e.g. more positive outcomes for protected groups)
  - Adding weights so that some observations are more important than others
- More information
  - Pipeline diversity
  - Interpretable models
  - Google’s nutrition cards for machine learning
- Training engineers/developers
- AI councils

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

What is Explainable AI?

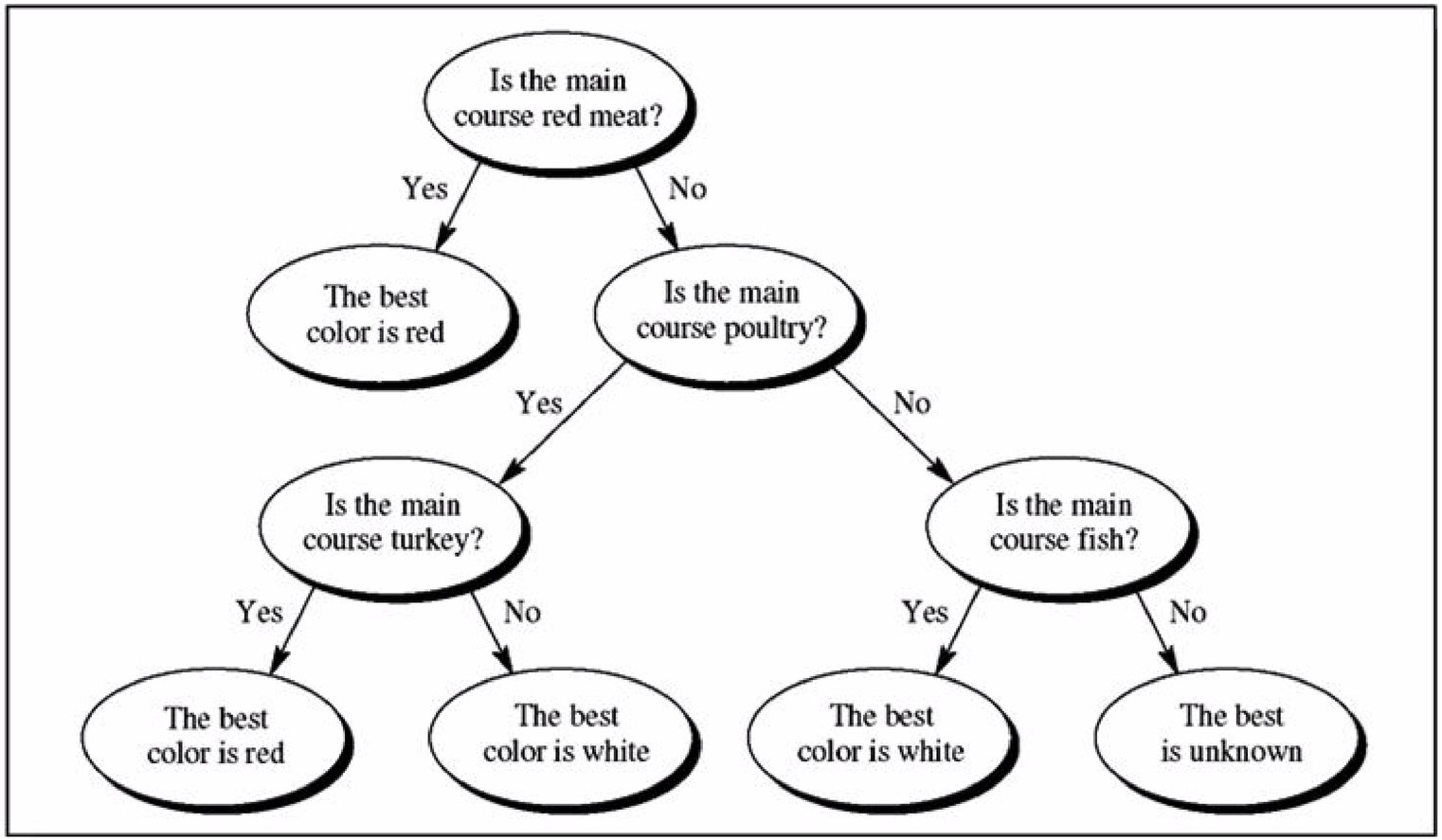
Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Explainable AI

- Explainable AI relates to using methods where how and why the algorithm arrived at the results can be understood by human experts
- This contrasts with some types of machine learning that operate more like a “black box” (e.g. deep learning)

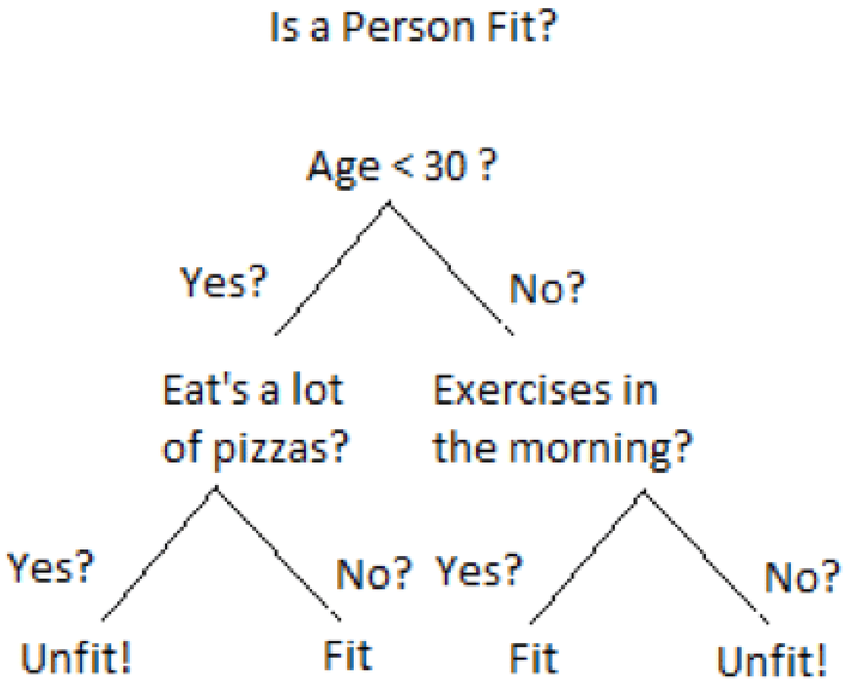
Explainable AI

- Rule-based systems are easy to explain



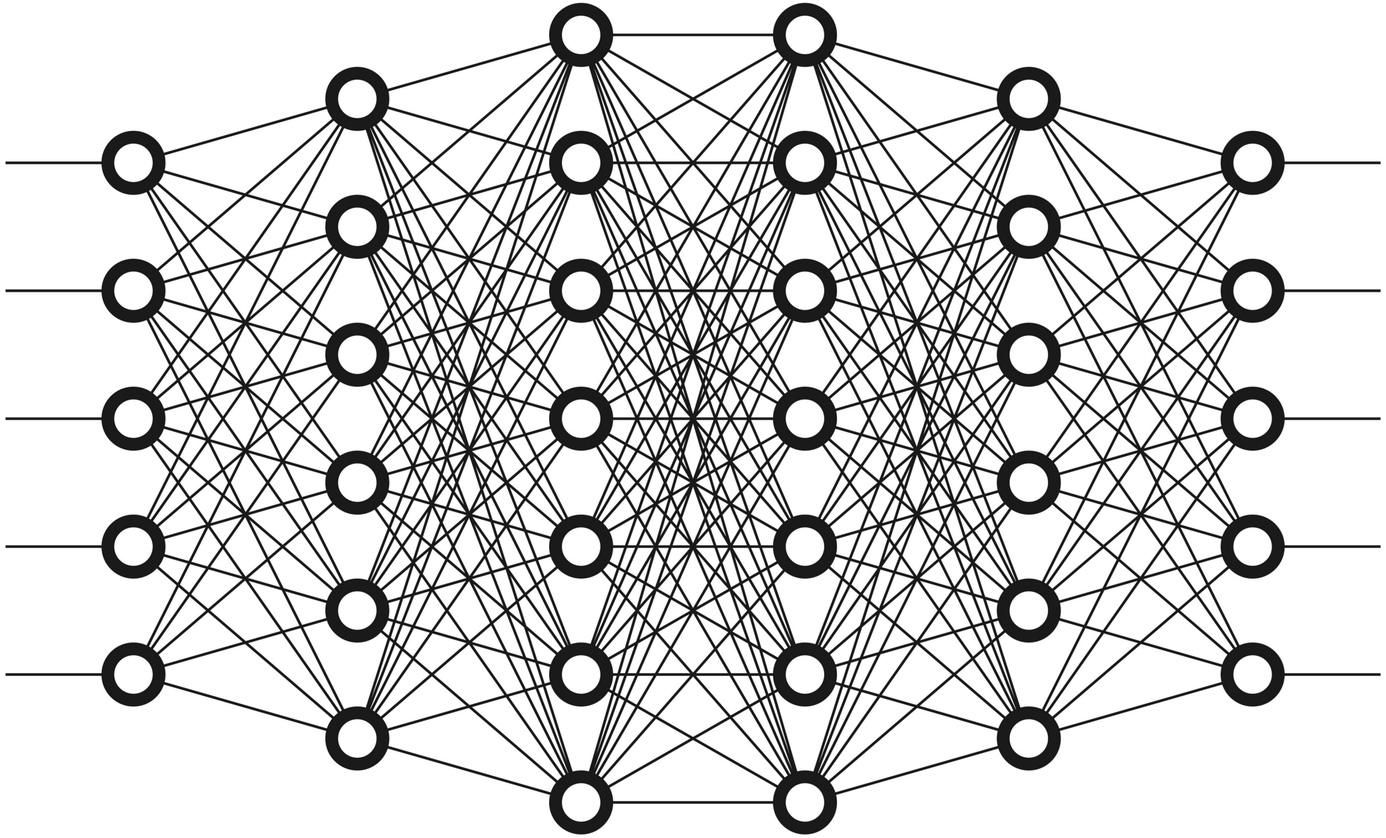
Explainable AI

- Machine learning decision tree output is also relatively easy to explain



Explainable AI

- Deep learning output can be much more difficult to explain



Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Examples of When Interpretability is Important

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Explainability, or Interpretability

- In some business contexts, being able to understand why the algorithm comes to a decision it does really matters

Importance of Interpretability in HR

- Many guidelines and legal protections require HR organizations or departments to clearly document how they come to decisions
  - Equal Employment Opportunity Commission (EEOC) rules

Importance of Interpretability in HR

- New data privacy laws
  - General Data Protection Regulation (GDPR) in the European Union
  - California Consumer Privacy Act (CCPA)
- Place restrictions on employer ability to use data provided by potential applicants
  - Need to be able to provide the logic that was used when arriving at a final decision

Investment in Explainable AI

- Companies, especially big tech, are increasingly invested in this issue
- Government, through various agencies, is also invested in providing resources to build more explainable AI systems for different industries

Wharton
University of Pennsylvania
ONLINE

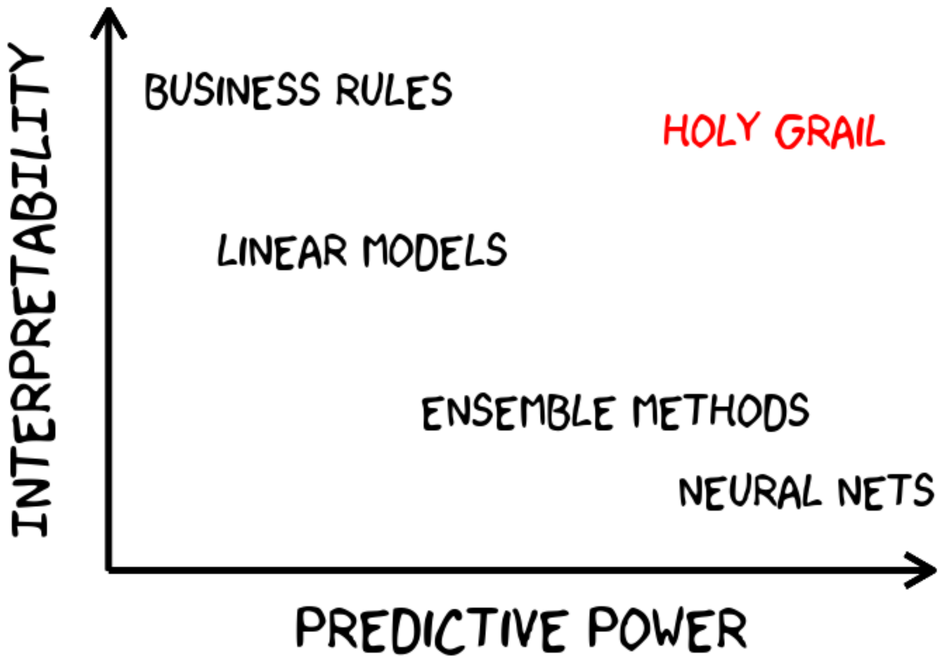
AI Applications in People Management

Tradeoffs between Interpretability and Performance

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

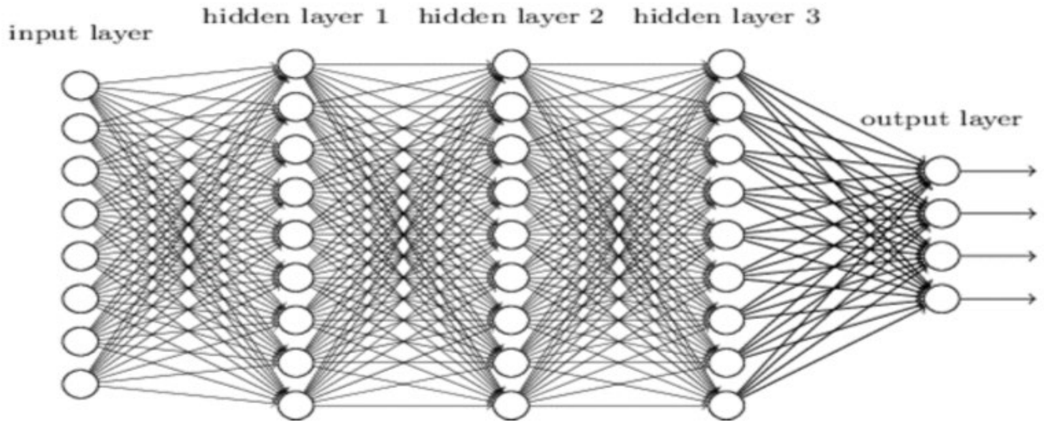
The Performance Penalty of Interpretability

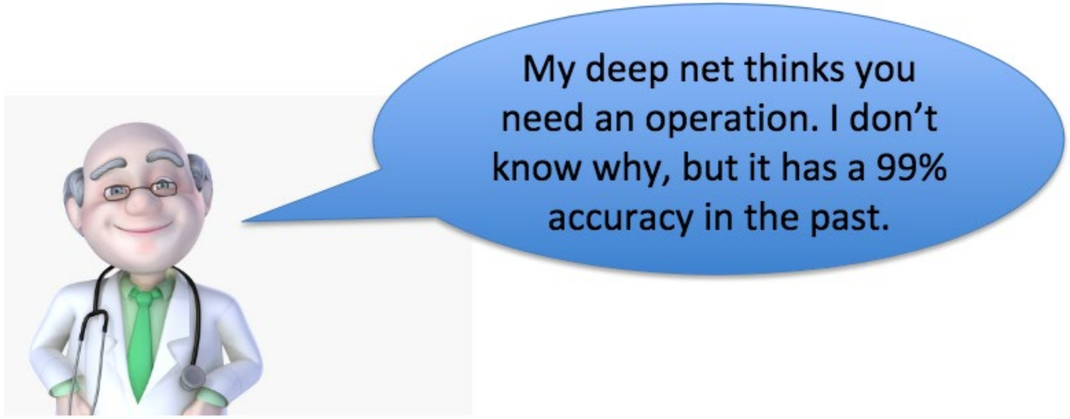
- There is generally a tradeoff between explainability and performance
- In practice, a precision prediction model often needs to be balanced by the ability to justify the model
  - Loan processing



The Performance Penalty of Interpretability

- When is explainability more important than accuracy?
  - User clicks or buying/selling assets — interpretability doesn't matter much
  - Promotions — interpretability matters a lot





Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Some Approaches to XAI

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Some Approaches to Explainable AI

- SHAP: Separates out how much each feature is contributing to the output/prediction
- LIME: Can start with a model that's very complex and difficult to interpret and generate a simpler comparison model that is locally accurate
- Surrogate trees: Generates a simpler model (e.g. a decision tree) that mimics the performance of the more complex model and is easy to interpret
- Auto-encoders: Boils data down into a small set of features to make it easier to interpret

Some Approaches to Explainable AI

- There are initiatives led by a number of companies
  - IBM - Explainable AI 360
  - Microsoft - InterpretML
  - H2O - Driverless AI

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

What Problem Can the Blockchain Solve in HR?

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

What Problems Can the Blockchain Solve in HR?

- AI solutions require massive amounts of data
- Potentially new business models using fine-grained data to predict match quality and performance
- At the same time, there are growing concerns about ownership and privacy of data
- Blockchain could help solve this problem
  - It is a way for two parties to exchange information without the need for a third-party to own or hold the data

Questions Emerging Around Where Data is Best Stored

- Why can’t we just trust third parties to keep our records safe for us?
  - Growing regulatory issues (e.g. GDPR)
  - Market failure where the right incentives for commercial investment do not exist
  - Concerns about use and abuse of personal data

Potential of Blockchain

- Blockchain is a data storage technology that is immutable
  - Transactions cannot be changed once confirmed, and therefor it is the “truth”
- It does not require a central authority/owner for verification
- Has the potential to be the data storage solution that ties private data together with AI prediction

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Why is the Blockchain so Transformative?

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Why Do We Need an Alternative?

- Limitations to how a database/spreadsheet works
  - Digital bits can be written and rewritten
  - The validity of any information stored in a database requires trust between counterparts
  - Requires a centralized and trusted broker
- This means we need to rely on institutions to hold highly personal data and this may not be acceptable

What the Use of Blockchain Offers

- Blockchain is a software protocol that achieves trust algorithmically, so a third party is not needed
- Individuals can own personal data and companies can use it for HR analytics, and the data can be trusted

Why is a Change to Database Technology so Important?

- Databases are the “factory floor” of information-based businesses
  - Requires a centralized and trusted broker
- Massive amount of economic value in trust-brokering
  - Car rentals, hotels
- Disintermediation of the need for trust institutions
  - Government issued IDs and failed nation states
- In HR, many institutions exist, at least in part, so you can trust that they are truthfully maintaining databases
  - Universities
  - Learning platforms

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

How Does the Blockchain Work?

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

How Does the Blockchain Work?

- The blockchain can be viewed as a ledger/spreadsheet
- For it to be the “truth” we need to know:
  - That the transaction (information) was verified by each of the counterparts (source/determination)
  - That the transaction entry was subsequently never falsified, updated, or deleted
- If we know these two things, we can be sure that the information is correct

How Does the Blockchain Work?

- Identical copies of this ledgers are stored across all participating “nodes” of the network
  - Information (transactions) are recorded (differences) in a sequential chain of blocks, hence the name
  - Each block holds changes in the world since the last block
  - The current state of the world is computed by summing up all the blocks

Blockchain Design Points

- Any transactions between two parties are verified using “PGP encryption”
  - These transactions are then /confirmed/ by the network and stored in this chain of blocks
- Blockchain is designed in a way that makes it exceedingly difficult to try to fake data or cheat the system
  - Doing so requires solving a problem that requires massive (prohibitive) computational resources
  - Because of this, it is a data storage technology that prevents faking or rewriting entries
  - If it is written in the blockchain, it must be true (within reason)

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

What are Some Emerging HR Applications?

Prasanna (Sonny) Tambe, Associate Professor of Operations, Information and Decisions

Emerging HR Applications

- Can broker between the world of HR data with the use of machine learning for prediction
  - On-tap credentialing
  - Instant payroll across national borders

Emerging HR Applications

- A number of firms emerging around the use of blockchain for HR data
  - Credentialing
  - Payroll
  - Performance evaluations

Emerging HR Applications

- Potential for new business models
  - Wearable data
  - Code check-ins
  - Skill-based credentials

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Effects of AI

Peter Cappelli, George W. Taylor Professor of Management and Director of Wharton's Center for Human Resources

Predicting the Future

- Actuarial science
- Economic forecasting — assumes the future will look like the past
  - If the future does not look like the past, the forecasting models won't work very well
- Asking people — polling, market research
- Expert judgement questions (Phil Tetlock 2017; Tetlock and Gardner 2018)
  - People who are confident, deep subject matter experts, and are guided by theory do worse
  - People who do better question assumptions, consider counter arguments, look for similar situations elsewhere

Where Do the Predictions Come From?

- Competition for “eyeballs” in private sector
- Money wins
- Some examples of predictions that are more hype than substance:
  - Early 2000’s – the coming labor shortage
  - The “millennials”
  - Driverless trucks

What Have the Effects Been of IT?

- Biggest changes come when technology pushes management to make different decisions about work organizations
- IT is not designed to eliminate jobs, but to introduce new functions that didn’t exist before
  - Online shopping recommendations

What Have the Effects Been of IT?

- Effects of applying IT to existing jobs, don’t necessarily eliminate those jobs
  - Even in the context of robotics
  - Low skill jobs are still around because labor is cheap
  - Driverless cars
- Simply because things are possible to do with robotics and artificial intelligence doesn’t mean that there is a market for it, or that it will be cheap enough to do

Is AI Different?

- Data science can focus on one task
  - Radiologists
  - Driverless trucks
- Management decisions matter enormously
  - Numerically controlled machines (CNC machines) and the use of computers in engineering decisions
- The big impact on jobs and productivity comes when managers use the IT to restructure how work is organized

Wharton
University of Pennsylvania
ONLINE

AI Applications in People Management

Course Conclusion

Peter Cappelli, George W. Taylor Professor of Management and Director of Wharton's Center for Human Resources

Conclusions

- Data science is a different approach to solving problems and making decisions
  - Puts priority on making more accurate predictions
  - Not necessarily fair, explainable, or consistent with other goals
- The challenge in applying these tools to managing people is about balance
- Artificial intelligence and data science will push other changes in the way we manage employees and work, which could matter just as much as the data science itself
  - Collecting more data, measuring, and analyzing
  - Fresh eyes looking at old problems

ONLINE