Some Open Questions about the Value of Data and Simple but Useful Concept of Data Q

Michael Schwarz, CVP & Chief Economist Microsoft Corporation

Economics of Data Workshop
University of Pennsylvania, December 2018
Is AI like no other labor saving tech?

“...the epoch that began early in this century, and an analysis of its latest manifestations: an economic order in which knowledge, not labor or raw material or capital, is the key resource.”

"The Age of Social Transformation," by Peter F. Drucker (November, 1994)

- It is difficult to overestimate importance of AI
  --but it could not stop us from trying.

- From economic point of view AI is not unlike other major technological innovations; from philosophical point of view it is like no other
Where will productivity gains go?

• Consumers/employees
• Capital owners
• Purveyors of AI technology (established and/or new)
• Gains dissipation
AI tech is a technical triumph—AI business is...

- What will emerge as the dominant business model for AI?
- Macro productivity growth due to AI is less than measurement error
- Selling APIs is a logical model for monetizing AI
  - Is it sustainable?
  - Is it a good idea?
  - An AI API that works “perfectly” will likely become a commodity
- There is no monopoly on AI technology--is data the strategic asset?
Is data the most important strategic asset?

• Is data scarce like gas or plentiful like water?
• A baby creates AI using surprisingly small amount of data
• AI tends to be more data thirsty than humans-- will it stay this way?
• For tasks that AI can do perfectly marginal value of data is likely low

Bottom line: it is a certainty that data leads to first mover advantage and the first mover advantage brings in more data but this virtuous cycle by itself is probably not sufficient to prevent commodification of AI on the long run
KPIs in the age of AI

• In the age of AI KPIs becomes critically important secret sauce
• The case of customer service bots
• The case of recommender systems
Data can cause monopolization of AV tech

• Should vendors of AV technology be required to share accident data?
• FAA requires public disclosure of near accident and accident data
• Is there a rational reason not to require the same of car makers?
The data to die for— the case of AVs

What will be the number of miles per fatality for future AVs?
A. Around one million miles per fatality
B. Around ten million miles per fatality
C. Around one hundred million miles per fatality
D. Around a billion miles per fatality
The data to die for— the case of AVs

What will be the number of miles per fatality for future AVs?
A. Around one million miles per fatality
B. Around ten million miles per fatality
C. Around one hundred million miles per fatality— correct
D. Around a billion miles per fatality

Currently there are about 85 million miles per fatality
The Impact of Big Data on Firm Performance (Caricature)

Big Data Is Key Ingredient to Increase Firm Performance

Big Data

Machine Learning

Increase Profits

+ =

Increase Revenues

and/or

Reduced Costs

“It’s not who has the best algorithm that wins. It’s who has the most data”
What is the difference between people and corporations?

• People can retrieve from memory information that they know they have
• Retrieving data that organization is known to have might require enormous time and effort.

• Here are example of questions that might be answered in minutes or months depending how the data is organized:
  • List of employees who were on parental leave three years ago
  • Percentage of users who churn after not logging into a service for 6 months
  • Increase in 2018 revenues among customers with three service calls in 2017
  • Etc. etc.
Defining Company’s Data Q

• We have neither a term to describe nor a measure to quantify “data agility”

• **Data Q of a company is the maximum amount of time (or resources) a company needs to pull data that is known to be available on its servers**

• Average Data Q is the average amount of time it takes to obtain the data for data requests that come from decisionmakers
  • Average Data Q is misleading because decisionmakers avoid asking data questions that could not be answered at a reasonable cost
What causes poor Data Q?

• Some companies achieve much lower Data Q than others

• Hard to find datasets
• Poorly design or poorly documented data schemas
• Difficult to navigate permissions for data access

• If we were to start measuring Data Q things will get better
• Low Data Q is a compliment to investments in AI
• Are companies underinvesting in reducing Data Q?
Instead of a Conclusion

• What is your company’s Data Q?
Another Conclusion

WE ARE HIRING