

# **ALLEGORY OF PRUDENT DATA SCIENCE**

DATA SCIENCE – PAST PRESENT FUTURE

# Data, Users, Companies – A Triad

**Seller of Raw Material**

Advertisers  
(customer of  
raw material)

**Company**

**Gains Value**

**Consumes Value**

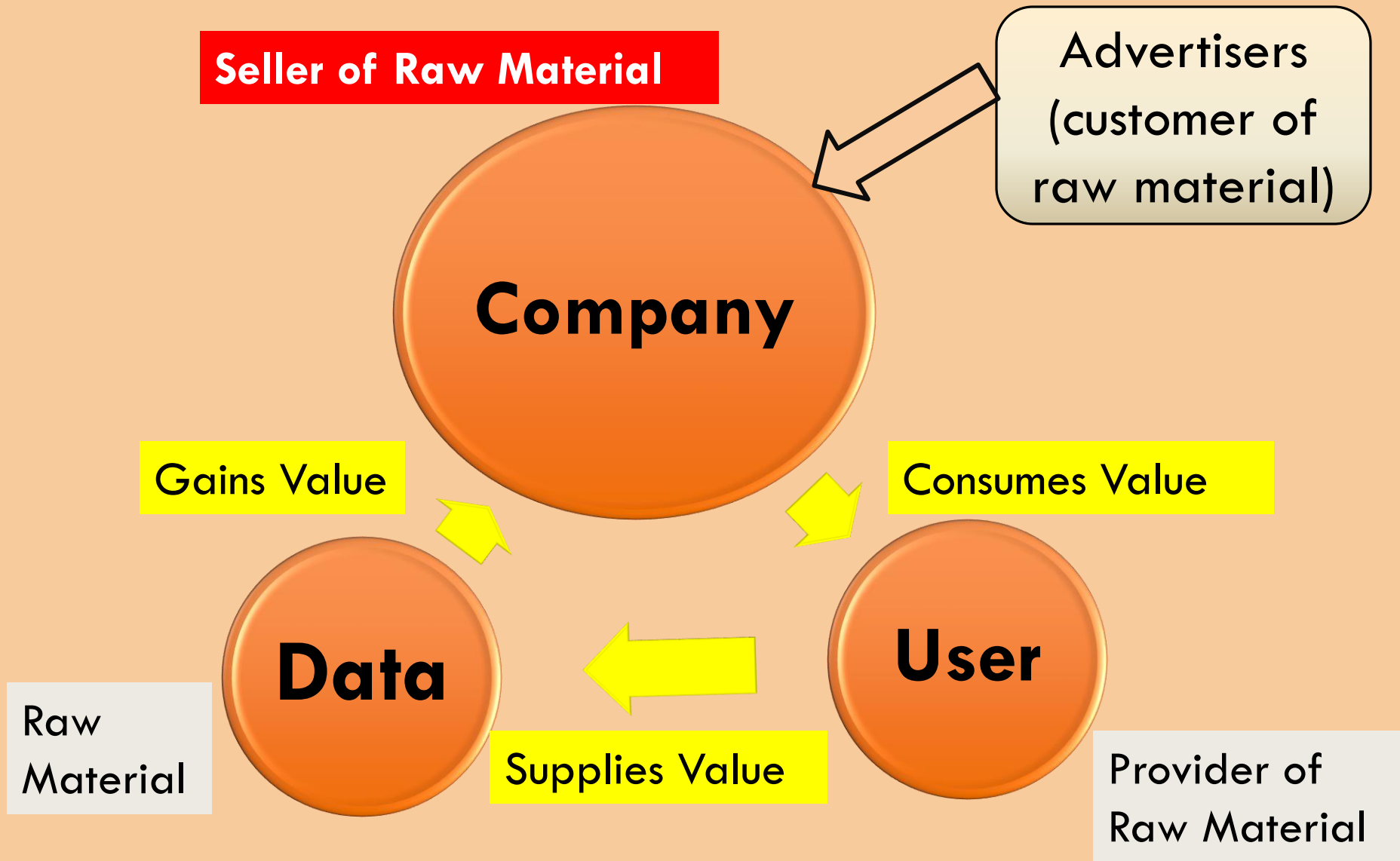
**Data**

**User**

Raw  
Material

**Supplies Value**

Provider of  
Raw Material



# WHY COLLECT SO MUCH DATA?

- Race to collect data
- Models for each micro objective
- Each model having its own metric for optimization
- Business models are being changed due to these models
  - Are the privileged or the lowest segment of society who get affected by these changes?

# WHERE WILL THIS LEAD TO?

- Big companies have bigger appetites
- They pay to get user data (most bribes are direct to users)
- **Two obvious problems**
  - May get stolen (hacked)
  - May be misused (by the company and its employees)
- **Major scandal or incident may lead to regulation**
  - Can AI be regulated?
  - Will companies want regulation?
  - Will public want regulation?

**“ARE NEITHER INTENDED NOR  
COINCIDENTAL BUT INEVITABLE”**



# CAN WE MODEL EVERYTHING?

- Models depend on data. Data and model developers have their own inherent bias
- Predictive model need objectives. We need to tell the system to do A or B
  - What if we don't know if A or B is the right decision? How can we tell the system?

# WHAT WOULD YOU DO?

A



B



# WILL MORE DATA SOLVE PROBLEMS?

- What if I collect every small action that you make every day?
  - At micro level will I get a consistent picture?
  - Using this micro-level data for micro-level models will lead to what?
  - More data implies more consistency?

# **ILLUSTRATIVE CASE STUDIES**

# **SAGA OF AIR FRANCE FLIGHT 447**



# BOEING 737

- Ethiopian Airline Flight
  - Boeing installed a new software which tells the plane to move its nose down to increase its speed and avoid it from stalling.
  - However, the faulty sensor told the computer that the plane is stalling.
  - The computer then, without the pilot's knowledge pushed the nose of the plane down further, while the pilot was trying to raise the plane.
- The computer literally flew the plane into the ocean.

# GOOGLE FLU TRENDS

- A near real-time service. Attempts to infer the intensity and location of influenza outbreaks
- Despite some initial success, the models completely missed the 2009 influenza A–H1N1 pandemic and consistently over-predicted flu cases from 2011 to 2014.
- Google's recommendation systems, which began suggesting flu-related queries to people who did not have flu

# GOOGLE FLU TRENDS

- Google Flu Trends began inferring flu cases where there were none.

➤ <https://www.google.org/flutrends/about/>

- Google Flu Trends and Google Dengue Trends are no longer publishing current estimates of Flu and Dengue fever based on search patterns.

# A PUBLIC RELATIONS DISASTER -THE TELEGRAPH

- A bot was created by Microsoft; named "Tay" - thinking about you
- Within **16** hours of its release, Tay had tweeted more than **96,000** times, yet MS suspended Tay's Twitter account
- Microsoft accidentally re-released the bot on Twitter on March 30, 2016 The bot was quickly taken offline again
- MS said that Tay "has had a great influence on how [Microsoft is] approaching AI,"

# WHY DID THE POLICE COME KNOCKING?

- In late 2013, Robert McDaniel – a 22-year-old black man who lives on the South Side of Chicago – received an unannounced visit by a Chicago Police Department commander to warn him **not to commit any further crimes.**
- The visit took McDaniel by surprise. He had not committed a crime, did not have a violent criminal record, and had had no recent contact with law enforcement.

# WHY DID THE POLICE COME KNOCKING?

- It turns out that McDaniel was one of approximately 400 people to have been placed on Chicago Police Department's “**heat list**”.
- These individuals had all been forecast to be potentially involved in violent crime, based on **an analysis of geographic location and arrest data**.
- The heat list is one of a growing suite of predictive “Big Data” systems used in police departments across the USA

# IBM WATSON AND MD ANDERSON

- Achieved an accuracy of prediction near 90 percent
- But significant updating is needed before Watson can be tested further
  - Medical oncology staff
- Watson was not in any way integrated with the hospital's electronic medical record (EMR) system!!!

# IBM WATSON AND MD ANDERSON

- “The opportunity for A.I. in health care is still nascent, but we are proud to be pioneers in this arena.”
  - Christine Douglass, IBM spokeswoman
- “IBM Watson has great AI,”
- “It’s like having great shoes but not knowing how to walk—they have to figure out how to use it.”
  - one engineer at IBM Watson Oncology Project with MD Anderson

# IBM WATSON

- Medical staff have been working with IBM since 2012, using the AI technology in a variety of ways, including
  - ✓ Watson for Genomics—a cancer knowledgebase
  - ✓ Watson for Oncology—a tool that takes as inputs a patient's information and combs through medical literature
  - ✓ Watson for Clinical Trial Matching—a system to better identify patients who might benefit from clinical trials

**LOOKING INTO THE PAST**

# SELECTED LEGENDARY

- Information and the Accuracy

CR Rao 1992

- Computers and The Theory of Statistics: Thinking the Unthinkable

Efron 1979

- The Future of Data Analysis

John Tukey 1962

- Ramsey Truth and Probability 1926

- Newton's Method published 1711

- Taylor's Series approximations 1715 (final publication)

**AI VS HUMAN SKILL  
OR  
AI & HUMAN SKILL**



# POINTS TO PONDER

- Data Science has become a *power to change our lives as we see it.*
- This has lead to governments and private companies investing heavily in data science in order to obtain a competitive edge over their rivals.
- Data Science is indispensable to the scientific community, and it is our responsibility to provide a concrete and rigorous foundation.

# POINTS TO PONDER

- Data-driven approach to model discovery derives a model from the data
- This mode of inference is becoming increasingly realistic in the data age
- This process integrates holistically theory and methodologies and computation

# *Thank you*

Dr Subbiah M

Acroama Gnan Vikas Private Limited

+91 9952726863

[msubbiah@acroma.in](mailto:msubbiah@acroma.in)

[subbiahbayes@gmail.com](mailto:subbiahbayes@gmail.com)

<http://acroama.in/>