



# Affecting Change with Data

## A Primer on Hype and Reality in Data Mining

Patrick Hall

H<sub>2</sub>O.ai

July 26th, 2018

- 
- 
- 

# Agenda

Disambiguation

Hype and Reality

Short History

Affecting Change

Past

Present

Future

Precautions



## Preface

*“We are drowning in information and starving for knowledge.”*

— Rutherford Roger, *Librarian (1915 - 2015)*

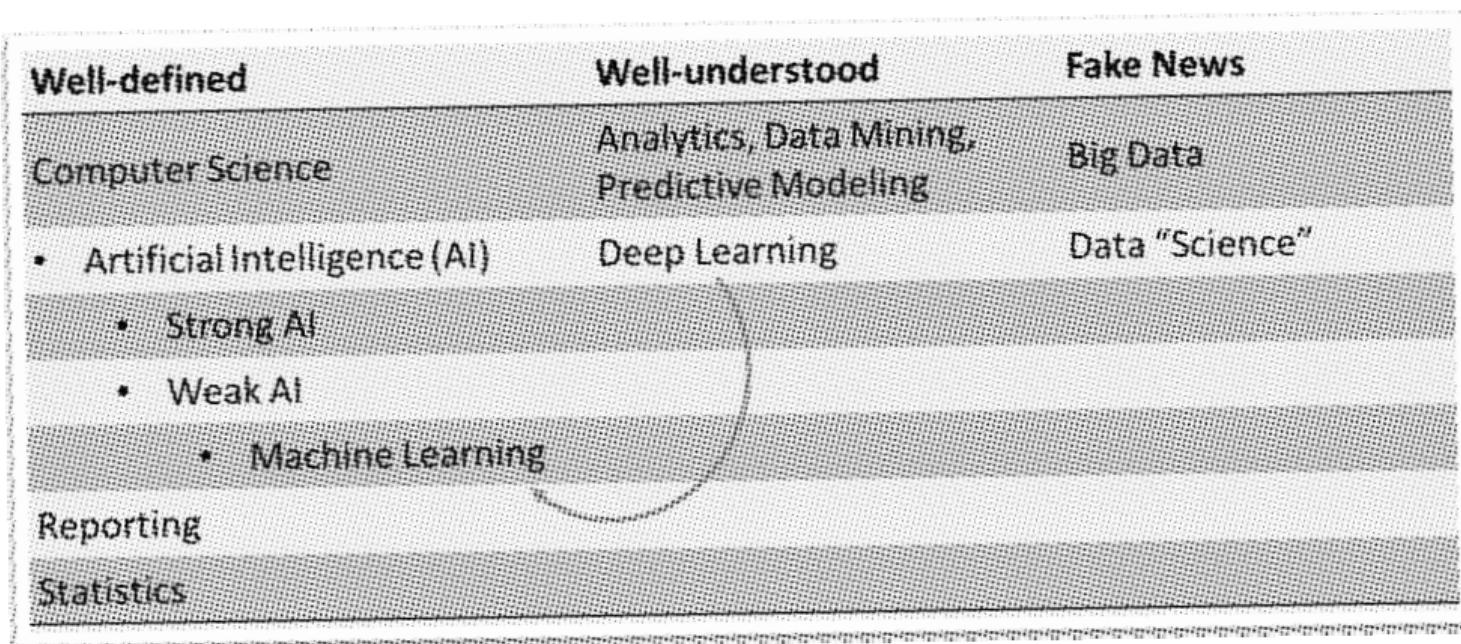
*“Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth. Truth is not beauty.”*

— Frank Zappa, *Musician (1940 - 1993)*

*“All models are wrong, but some are useful.”*

— George Box, *Statistician (1919 – 2013)*

# Buzzword Disambiguation



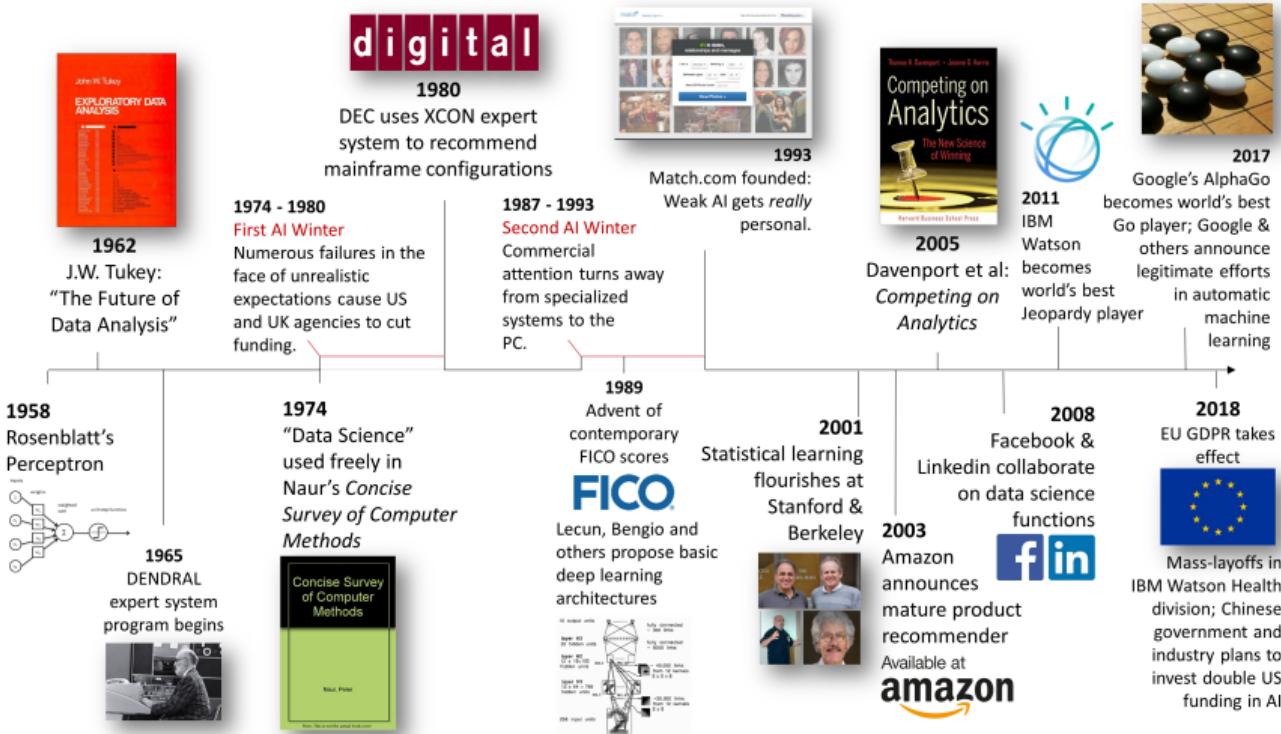
# Hype and Reality



- Data mining has been used for decades in traditional industries to improve financial margins.
- Big tech. made a windfall by overselling new but minor conveniences often at the cost of our personal privacy.
- As usual, we *may* be on the brink of a nonlinear leap in technology.

Image credit: Shutterstock — KimSongsak

# A Very Short History of Machine Learning





# Affecting Change with Data



Data mining is a blunt tool for revisiting the past, understanding the present, and seeing into the future.

Image credits: Flickr — chengyee, Andy Castro, Roo Reynolds



## Learn from the Past

- Often easier to conduct as an individual as tangible results include papers, data visualizations, stories, lessons, and morals.
- Use common sense to learn about human or financial gain or loss.
- Look for trends, groups, and outliers in data.

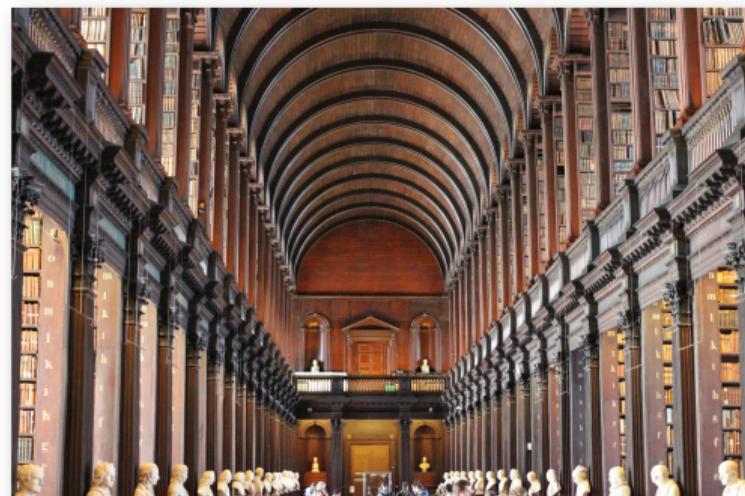


Image credit: Flickr — David P. Whelan



## Understand the Present



Typically requires organizational information technology (IT) support as tangible results include real-time reports and alerts created from data and IT systems.

Image credit: Flickr — Sérgio Bernardino

## Prepare for the Future

- Usually requires advanced IT support.
- Tangible results are typically decisions about customers, patients, students, equipment, or other valuable entities made by sophisticated computer systems.
- Common commercially viable applications include:
  - Quantifying future risk
  - Personalized promotions
  - Predicting churn
  - Recommending products or content



## Take Necessary Precautions



- Data can be biased, incomplete, or just wrong.
- Results can be ignored or used politically.
- Avoid:
  - Confirmation bias
  - Testing multiple hypotheses
  - Unintended discrimination
  - Privacy violations
  - Unmanageable complexity

Image credit: Flickr — Thomas Hawk

○  
○  
○

## Stay in Touch



[linkedin.com/in/jpatrickhall](https://www.linkedin.com/in/jpatrickhall)



@jpatrickhall