Machine Learning Ready

Machine Learning Overview

- Machine Learning is a general purpose technology, what does this mean?

Machine Learning Overview

- A **general purpose technology** or **GPT** is a term coined to describe a new method of producing and inventing that is important enough to have a protracted aggregate impact.

- Similar to electricity or the internet, in that it can be applied across domains and work to improve market outcomes.

Machine Learning Overview

- What is driving the continued progress of machine learning techniques?
  - Increased Access to Computing Power
  - Larger Training Datasets
  - Improved Algorithms - DNN

- Facebook moved from phrase-based translations to DNNs for roughly 4.5 billion language translations each day
- Error rates on ImageNet (10,000 labelled images) have been driven down from 30% in 2010 to less than 3% today.
  - 5% is important because that’s typically the human error rate

Machine Learning Overview

- Machine Learning adoption depends on the type of actively being completed and the value proposition between trading manual labor for automation.
  - Manufacturing automation is ripe for machine learning.

- TrendForce – Estimates smart manufacturing is a $200 billion a year industry and will increase to over $320 billion by 2020.
  - That is a projected compound annual growth rate of 12.5 percent.

- Similarly, the International Federation of Robotics estimates by 2019 the number of operational industrial robots installed in factories will grow to 2.6 million from just 1.6 million in 2015.

Machine Learning Overview

- However, before we all turn into robots consider two important facts:
  1. We remain remarkably far away from what would be consider a similar general intelligence that can be compared to humans
  2. Machines cannot do the full range of tasks that humans can do

- We can then refer to jobs or activities that might be good cases for Machine Learning as SML or Suitable for Machine Learning

- What are other examples of jobs that might be seen as SML and how do we know if our organizations are ready?

Machine Learning Overview

- Successful implementation of ML requires very detailed specifications on what is to be learned and data to support that learning activity.
  - Including the development of engineering features through a series of trial and error and
  - Then most importantly embedding these products into **normal business operations** in such a way that efficiencies can be realized.

- What then are the implications for ML teams?
- Learning apprentice?

Machine Learning Overview

- What tasks are most suitable for ML to take over:
  - Most recent successes are predicated on supervised learning
  - Competency is narrow as compared to the complexity of human decision making

1. Learning a function that maps well-defined inputs to well-defined outputs
   - If can predict Y given any value of X – still might not produce the actual causal effect
2. Large Data is present or can be created containing input-output pairs
   - The more training data available the more arcuate the model
3. Task provides clear feedback with well definable goals and metrics
   - If we know what to achieve – (optimize flight patterns not a single flight)
4. Where reasoning and diverse background knowledge is not necessary
   - Good at empirical associations but terrible at decision making that requires common sense of historical knowledge
5. No need for why the decision was made to be clear
   - NN could use millions numerical weights

Machine Learning Overview

6. A tolerance for error or sub-optimal solutions
    - ML use probabilistic outputs which means some error is always assumed
7. Function of item being learned should not change rapidly over time
    - Work best when the distribution of future test examples is the same roughly as the training set over time
    - If not the case systems need to be in place to refresh algorithms

Potential Future Scenarios, The Economist Forecast out to 2030

- Impact on GDP and Productivity
  1. Greater human productivity through upskilling
     - Assumes a high degree of complementarity between human skills and AI systems.
     - Will require greater investment by governments than is seen currently
  2. Greater investment in technology and access to open source data
     - Requires government policy and investment in access to open source data, potentially tax incentives to encourage private adoption and continued advances in computing technology to drive down hardware cost.
  3. Insufficient policy support for structural changes in the economy
     - Projects a clear substitution effect for labor due to a lack of upskilling or training
     - Could also result from a lack of national data sharing policies

- 1 and 2 both have positive long term impacts on GDP where scenario 3 could have slightly negative or at a minimum reduce nations' ability to realize all the potential gains from Machine Learning/AI.

The Economist Intelligence Unit, *Risks and Rewards: Scenarios around the economic impact of machine learning.* 2017

Machine Learning Gone Wrong

- Google's Cloud Natural Language API was launched in 2016. In fall of 2017 Andrew Thompson from Motherboard Inc. experimented with the tool and discovered bias results.
- "I'm Christian" had positive results but, "I'm a Jew" and "I'm a gay black woman" resulted in negative results.
- Trained on news and social media data

The Economist Intelligence Unit, *Risks and Rewards: Scenarios around the economic impact of machine learning.* 2017

Machine Learning Gone Wrong

From Google: “We dedicate a lot of effort to making sure the NLP API avoids bias, but we don’t always get it right. This is an example of one of those times, and we are **sorry**.”

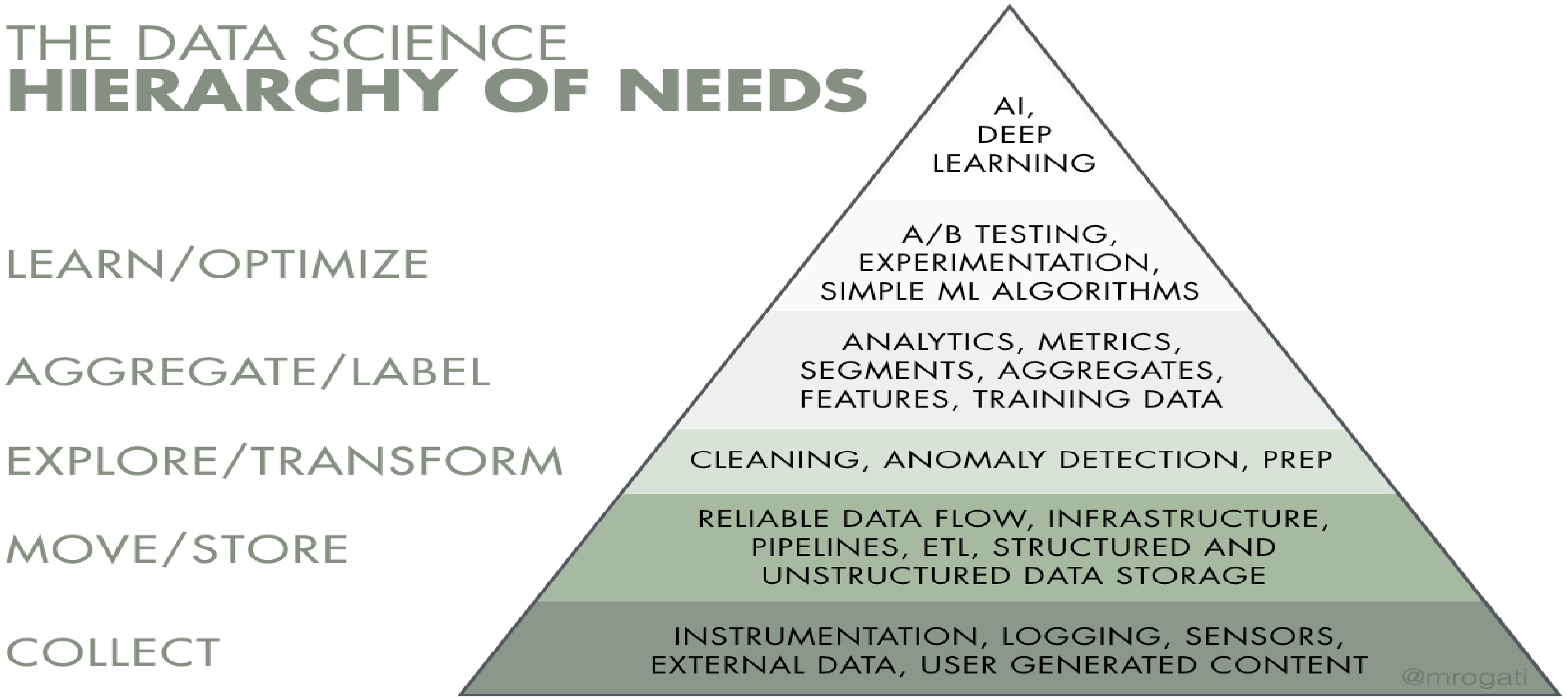
Machine Learning Gone Wrong

- Amazon Recruiting Engine
  - Used a AI tool to give candidates a score from one to five stars based on resumes.
  - “They (Amazon) literally wanted it to be an engine where I’m going to give you 100 resumes, and it will spit out the top five, and we’ll hire those.”
  - The problem was that data was used from 10 years of hiring that was mostly men, so the ML tool systematically scored women lower
  - Amazon just this year abandoned the project because it couldn’t be proven reliable.

Machine Learning Gone Wrong

- So if Amazon and Google sometimes get it wrong what can be done…?

  - Need to make sure we are asking the right questions and designing hybrid platforms with human expertise working along side intelligent machines.
  - Ensure we train our workforce to understand what the outcomes of the systems actually mean and most importantly….
  - How mistakes happen and what pitfalls are common
  - **Training and Organization Structure to include Data Governance are critical first steps.**



Intellectual Capital is Present

Educational Landscape

	Traditional Higher Education	Immersion Programs	Online	Boot Strapping
Example	Undergraduate & Graduate	Springboard, General Assembly, Data Society	Coursera, Udacity, DataCamp, etc.	MOCs, books, free courses
Goal	Industry recognized validation of skills	Gain new skills at a low cost, rapidly	Enhance current skills or gain awareness of field	Gain or enhance skills at personnel pace
Investment (Time and Money)	High/High	Low/Medium	Med/Low	High/Low

Steps for Entering the Pyramid

1. Prioritize areas in your organization that could be early adaptors
2. Develop a training plan
3. Establish a open data framework and culture – very hard
4. Establish Governance to include roles, responsibilities and guidelines for the organization
   - Includes accountability for data integrity
5. Pilot programs that work to create data pipelines that support established business operations
6. Create feedback mechanisms and adjust as needed

Advantages of Data Science/ML Approaches

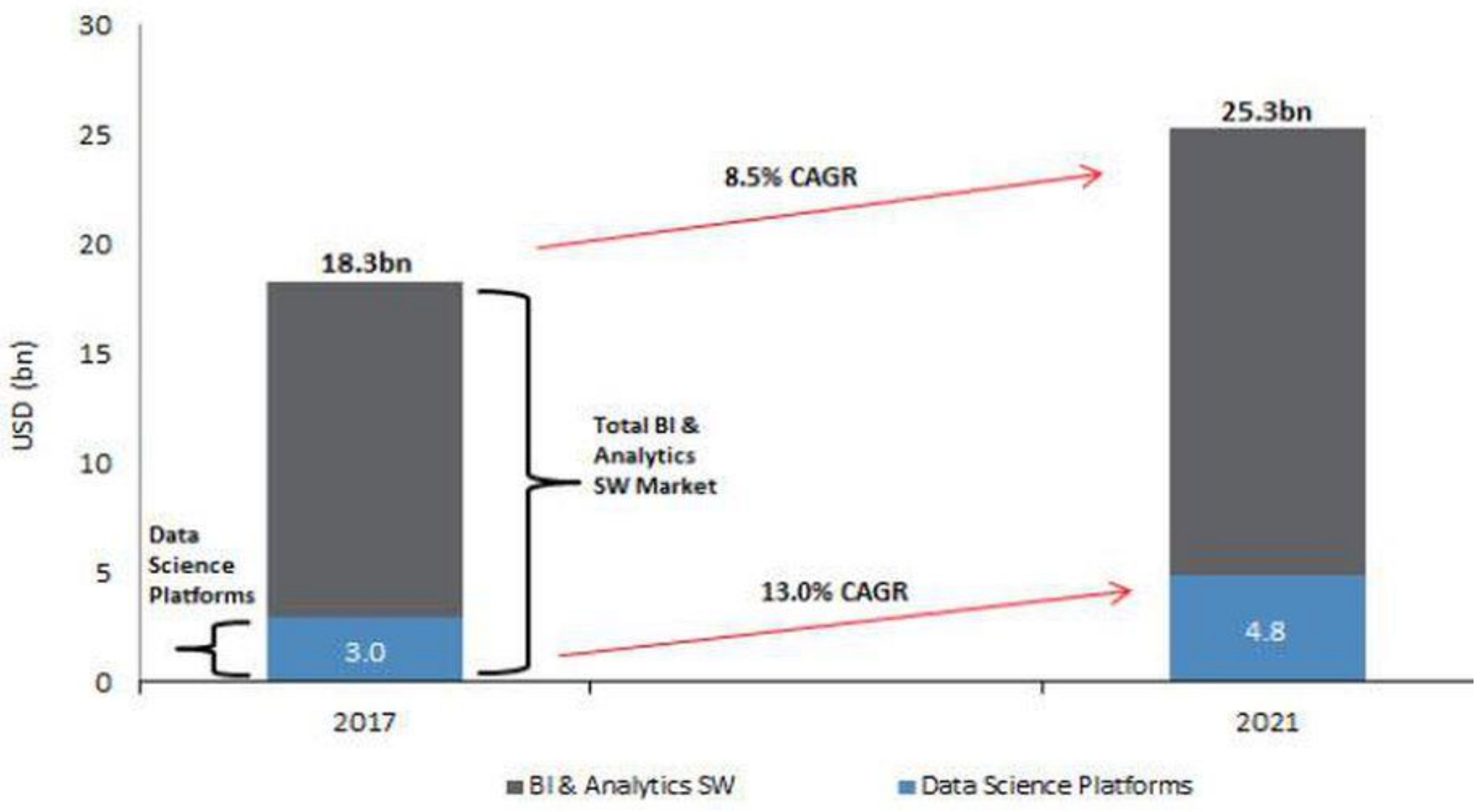
Traceability

Repeatability

Scalability

Data Science Platform Growth

Figure 29: Data Science Platforms are growing faster than overall BI & Analytics SW



Source: Gartner.