Gartner Data & Analytics Summit
Summit 2018
22 – 23 May 2018 / São Paulo, Brazil

Gartner Summits

The Essentials of Data Science and Machine Learning: How Machine Learning Extracts Knowledge From Data

Peter Krensky

AI Renaissance or Apocalypse?

Gartner®

Key Issues

1. How do you navigate the hype and semantics and deal with the shadow of AI?
2. How does it all work?
3. What can you do right now with data science and machine learning?

Key Issues

1. How do you navigate the semantics and deal with the shadow of AI?
2. How does it all work?
3. What can you do with data science and machine learning right now?

se·man·tics (səˈman(t)iks) *n.*
The meanings of words and phrases in a particular context

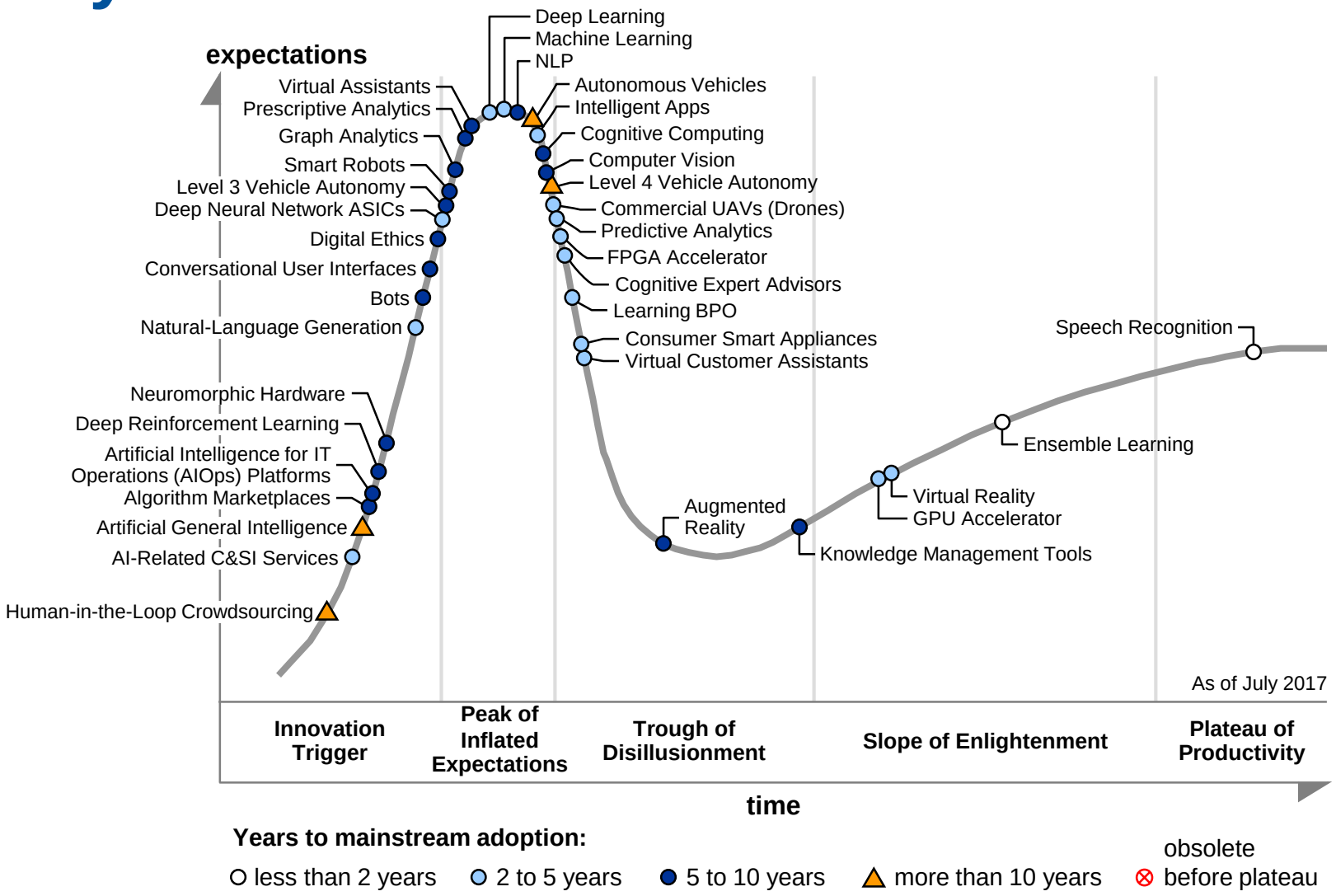
The Gartner AI Hype Cycle

Very early maturity levels

- **86%** of tech profiles (dots) headed to the bottom of Trough of Disillusionment
- **54%** not expected to plateau and deliver reliable productivity for mainstream buyers until 2022 or later

Huge potential

- **41%** offer transformational benefits
- **44%** offer high benefits



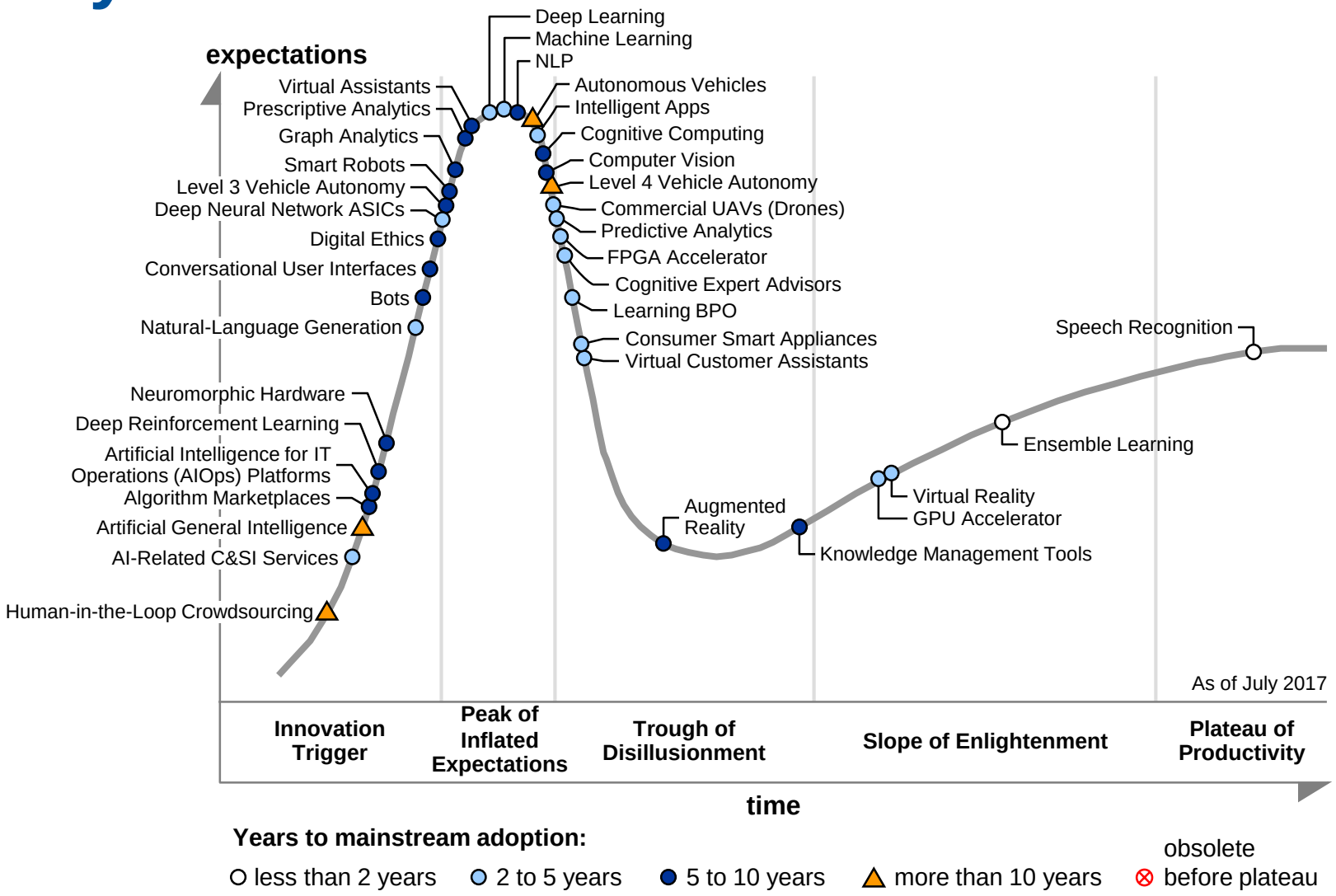
Source: "Hype Cycle for Artificial Intelligence, 2017," 24 July 2017 (G00314732)

Gartner®

The Gartner AI Hype Cycle

Though the lens of D&A

- **>50%** of hype has nothing do with data and analytics
- Marketers stick the AI moniker on anything remotely algorithmic
- Key data science concepts are crowded around the peak



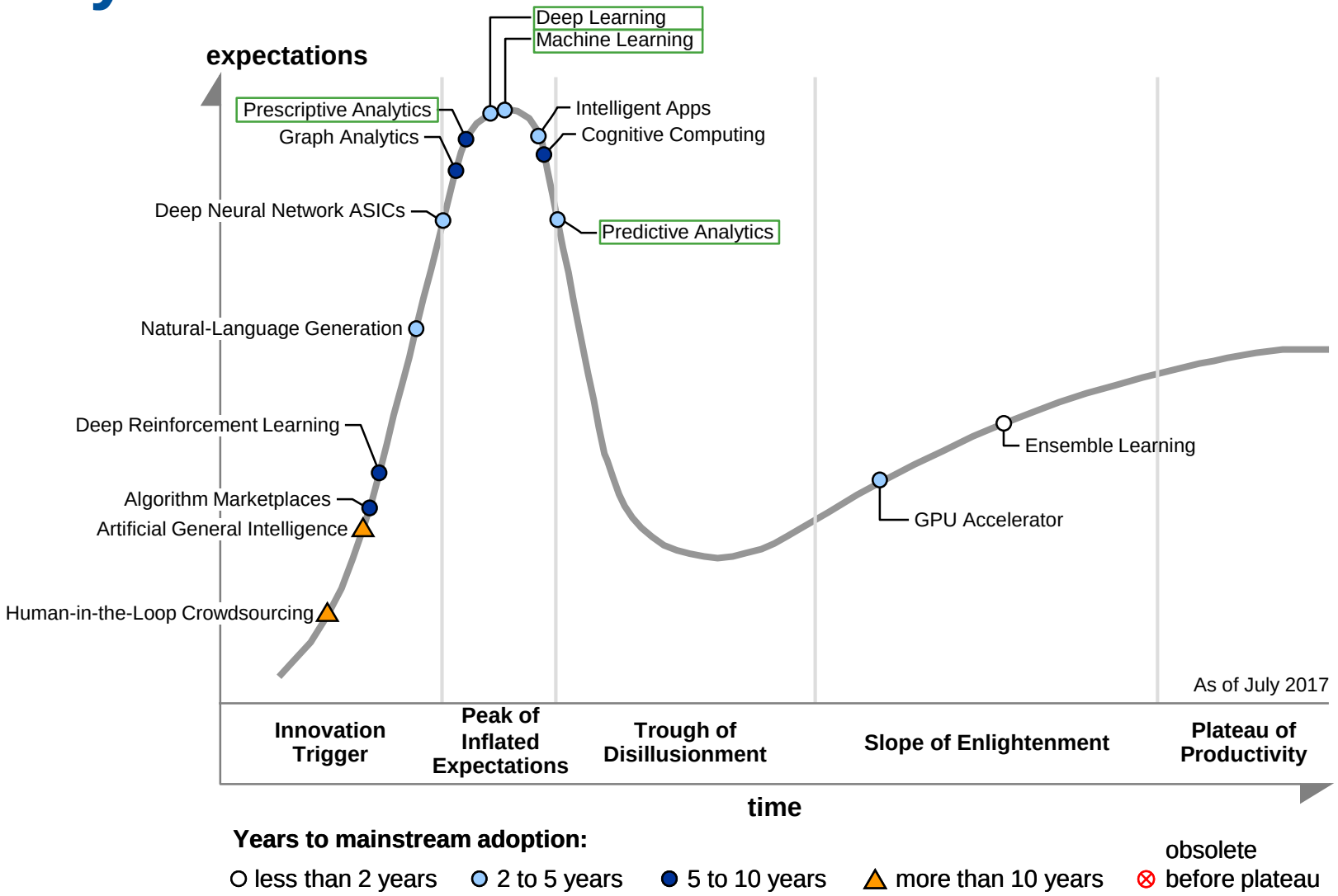
Source: "Hype Cycle for Artificial Intelligence, 2017," 24 July 2017 (G00314732)

Gartner

The Gartner AI Hype Cycle

Though the lens of D&A

- **>50%** of hype has nothing do with data and analytics
- Marketers stick the AI moniker on anything remotely algorithmic
- Key data science concepts are crowded around the peak



Source: "Hype Cycle for Artificial Intelligence, 2017," 24 July 2017 (G00314732)

Gartner®

Hierarchy of Terms From a D&A Perspective

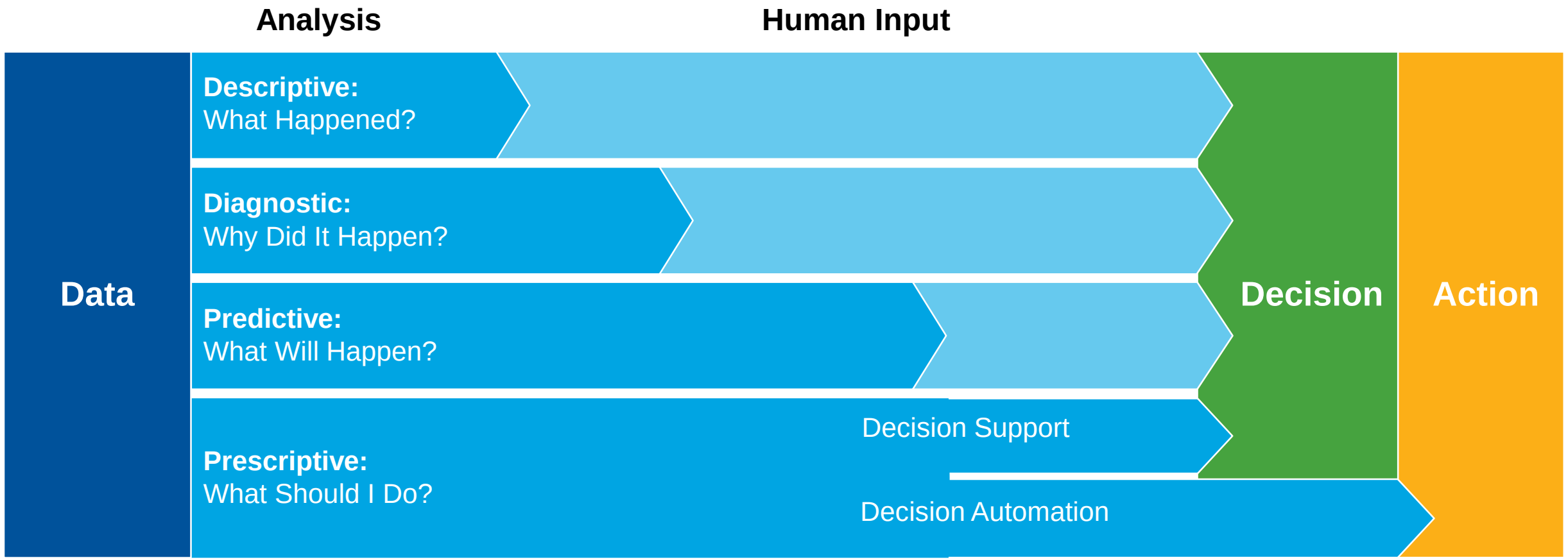
Artificial intelligence can be defined as a general approach to the simulation of cognitive processes by means of computer programs.

Data science and machine learning at its most basic is the practice of using algorithms to parse data, capture knowledge, learn from it, make a deterministic or predictive model and deploy that output into a business decision.

Deep learning is a subset of ML algorithms that creates knowledge from multiple layers of information processing.

Gartner®

The Analytics Spectrum



Gartner®

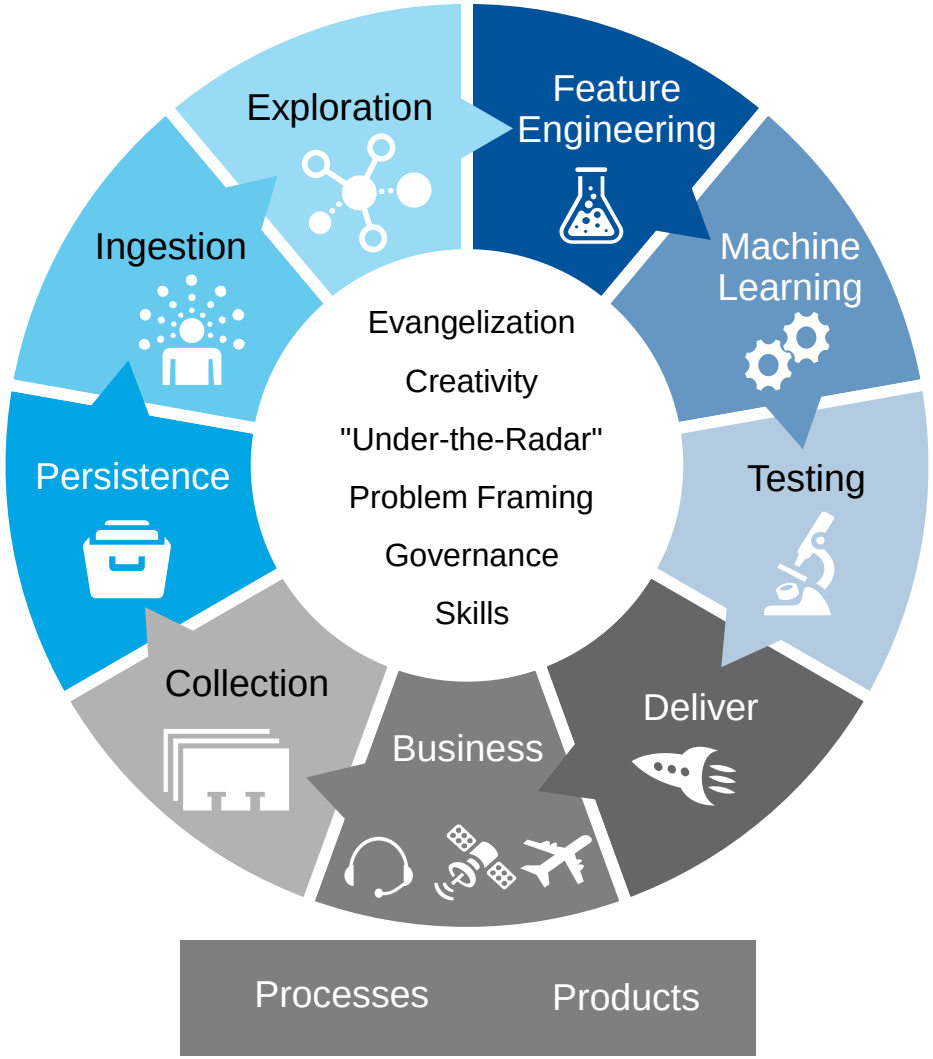
Data Science Is a Model Creation Cycle

Business hacking Flat files
Discover data shape

Bias

Many layers Data lake In-transit
RAM Bias

"Filter" Bias



Compression Relevance
Domain knowledge

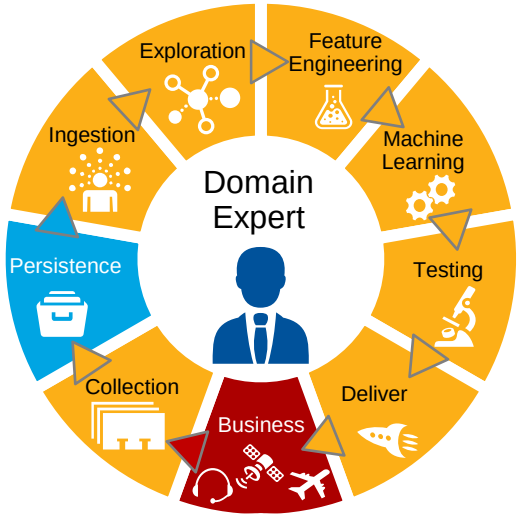
Grid search Machine learning

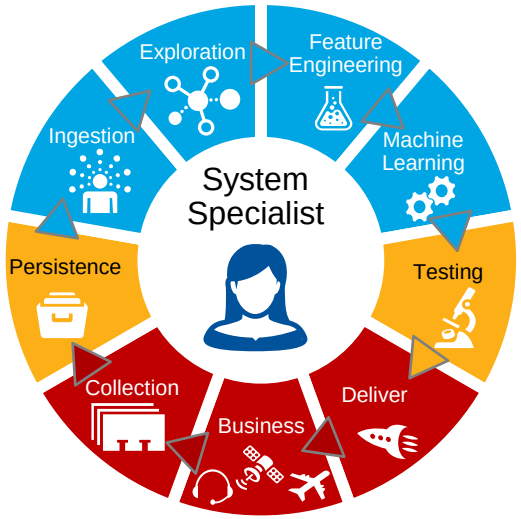
K-fold cross validation ROC, lift curves, cost matrices

Champion-challenger Deployment
Storytelling Value communication

Gartner®

Skill Map





Knowledge Spectrum

Poor	Little	Ok	Good	Strong

Citizen Data Scientist

Gartner

… at the Core, ML Is About Creating Mappings Informed by Input/Output Pairs

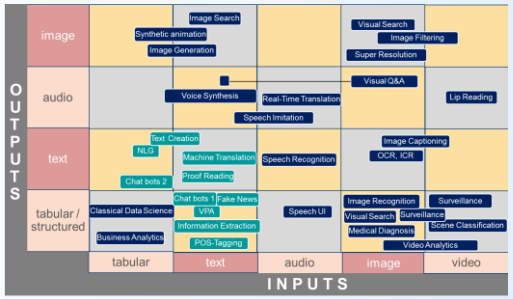
Type of Problem	Inputs	Outputs
Loan Application	Application data	Will the applicant repay the loan? (0 or 1)
Demand Prediction	Market situation	How many products will be bought? (n)
Self-Driving Cars	Car sensory data	Break, accelerate, tilt the wheel? (x, y)
Propensity to Buy	Profile and transactions	Will the customer buy or not? (0 or 1)
Failure Prediction	Sensor readings	Will a failure happen with 4 weeks (0 or 1)
Customer Churn	Profile and activities	Will customer cancel the contract? (0 or 1)
Medical Diagnosis	Pixel data from a retinal scan	Will the disease break out? (0 or 1)
Advertisement	Ad + context + user profile	Will the user click on ad? (0 or 1)

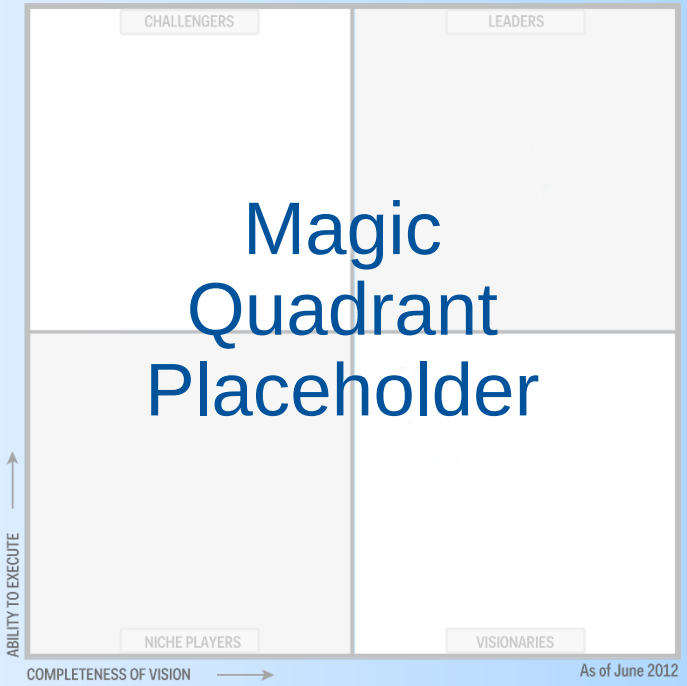
Landscape of ML Solutions

DYI

Salesforce Einstein

SAP Clea





Magic Quadrant Placeholder

From "Magic Quadrant for Data Science Platforms," xx February xxxx (G00xxxxxx)

Skymind's DL4J

Caffe

Google's TensorFlow Theano

Microsoft Cognitive Toolkit

H2O.ai's Deep Water

Baidu's Pebble Intel BigDL

Amazon Web Services' (AWS)

Apache MXNet

Business Users

Application Engineers

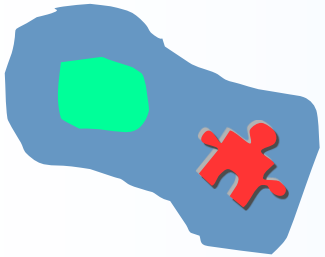
ML Engineers

Embedded Machine Learning

Machine-Learning APIs

R, Python, Scala, Matlab

Deep-Learning Frameworks



Data Analysts

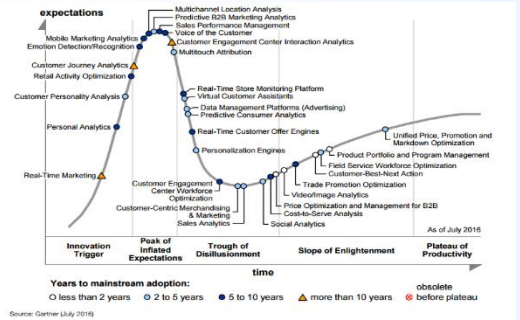
Data Scientists

Augmented Analytics

Data Science and Machine-Learning Platforms

Data Analysis Software

Buy



Intel Nervana

Microsoft Azure

Rescale AWS

Google Cloud Platform

Deep-Learning Cloud Platforms

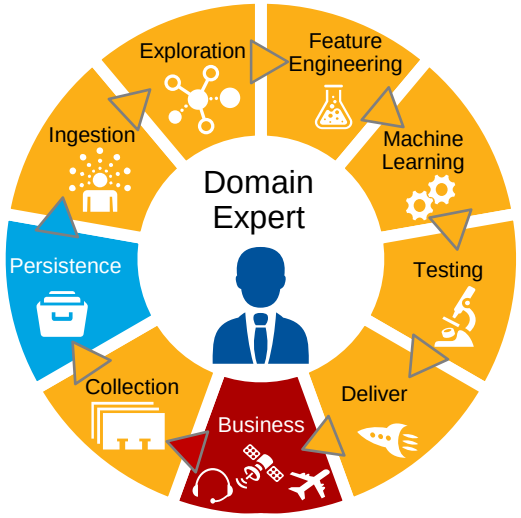
Deep-Learning Hardware

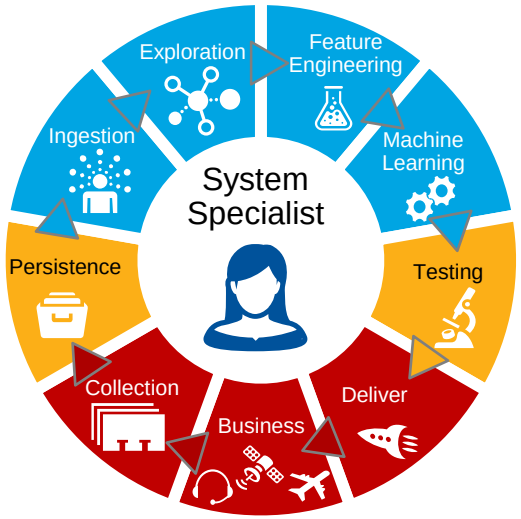
Nvidia, AMD, IBM, Intel

Key Issues

1. How do you navigate the semantics and deal with the shadow of AI?
2. How does it all work?
3. What can you do right now with data science and machine learning?

Skill Map





Knowledge Spectrum

Poor	Little	Ok	Good	Strong

Citizen Data Scientist

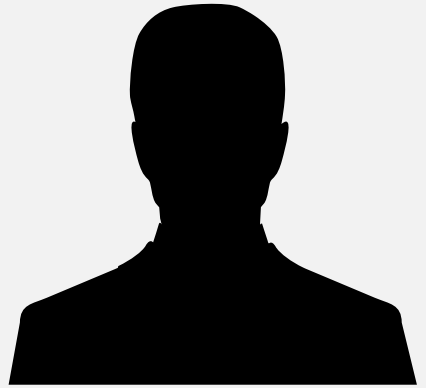
Gartner

Finding, Keeping, Nurturing Skills

By 2020, **>40%** of data science tasks will be automated

Core Data Scientists

Citizen Data Scientists



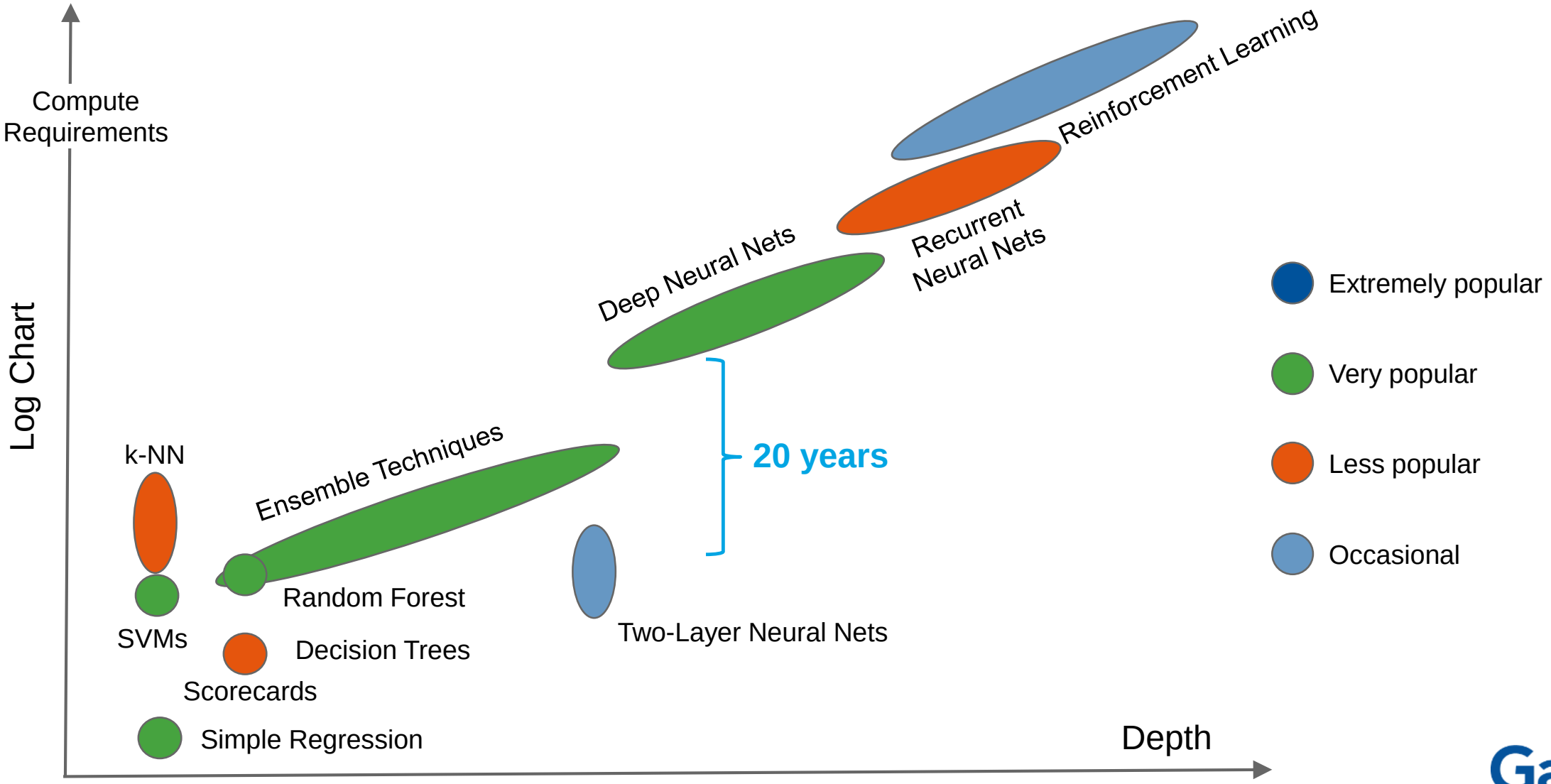
Academia

Consultants/ Freelancers

79% of DS teams currently use R
75% currently use Python

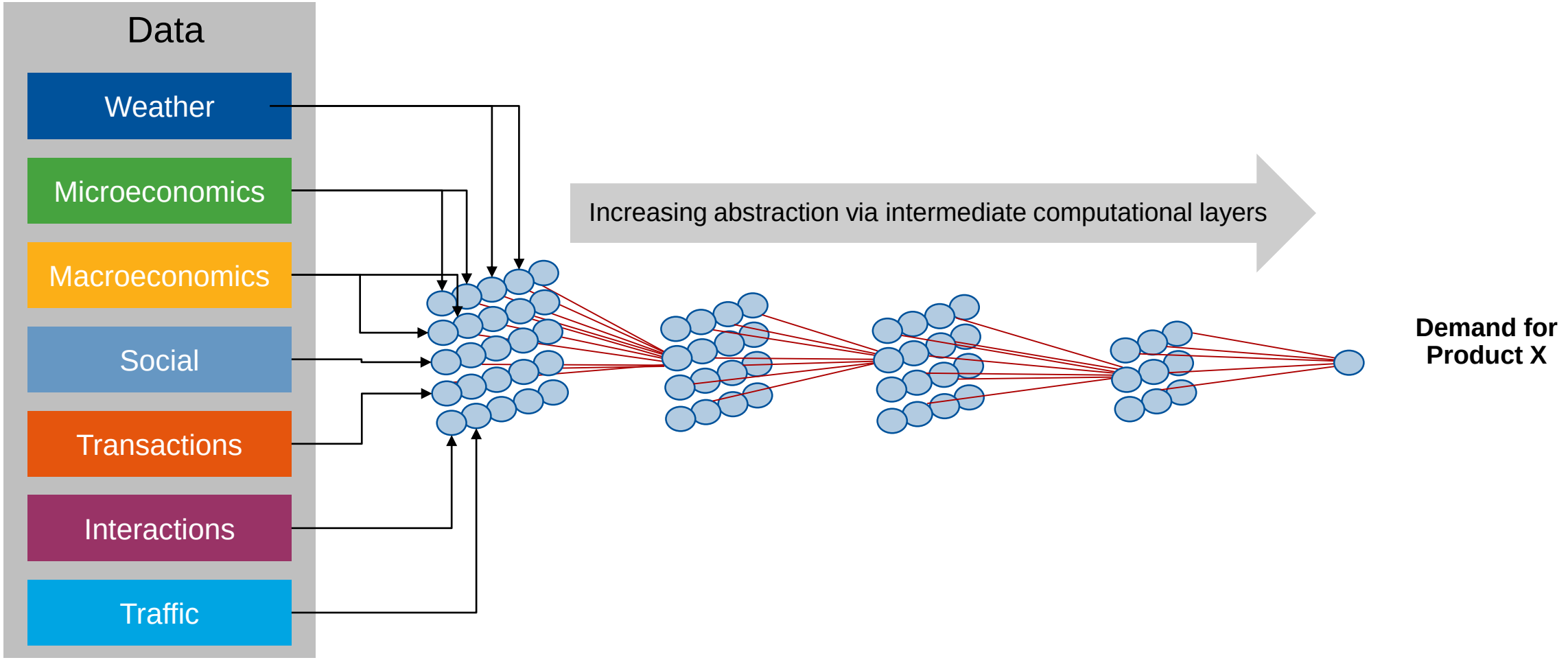
Gartner®

There Is a Zoo of Machine-Learning Approaches Out There ...



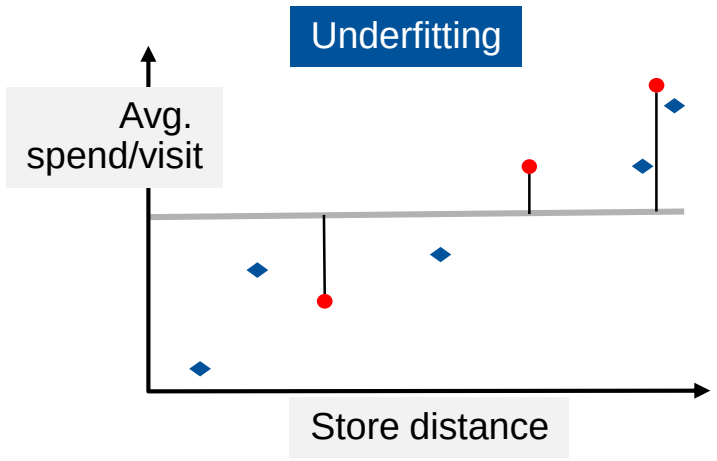
Gartner

Deep Learning Addresses One of the Biggest "Big Data" Challenges: Data Fusion

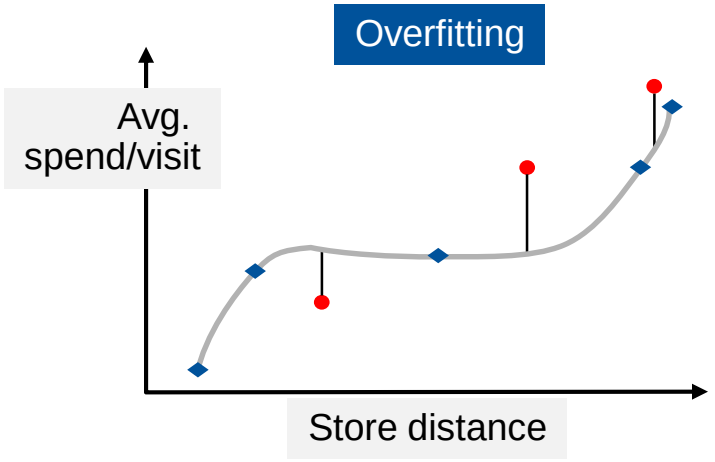


Gartner®

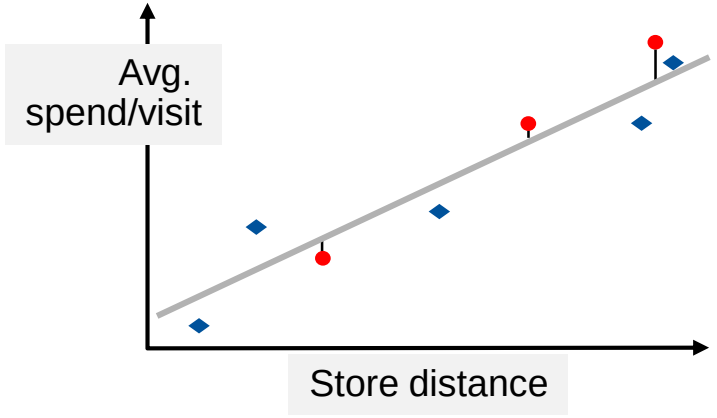
The Challenge of Machine Learning: Under and Overfitting

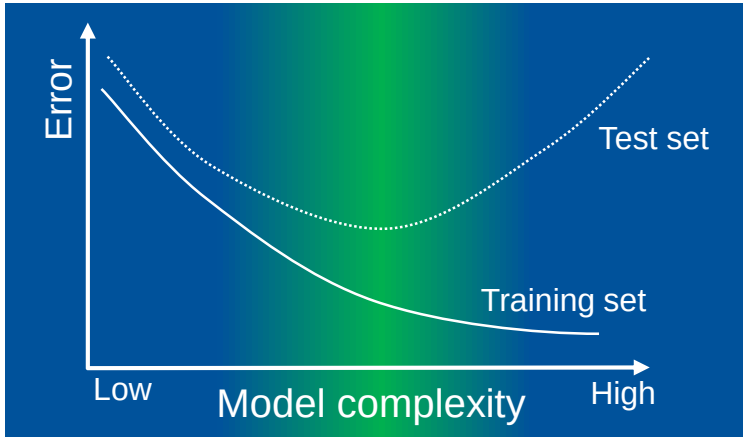


- Predictor is too "simplistic"
- Cannot capture the pattern



- Easy to be good on the training data
- Predictor is too "powerful"
- Rote learning



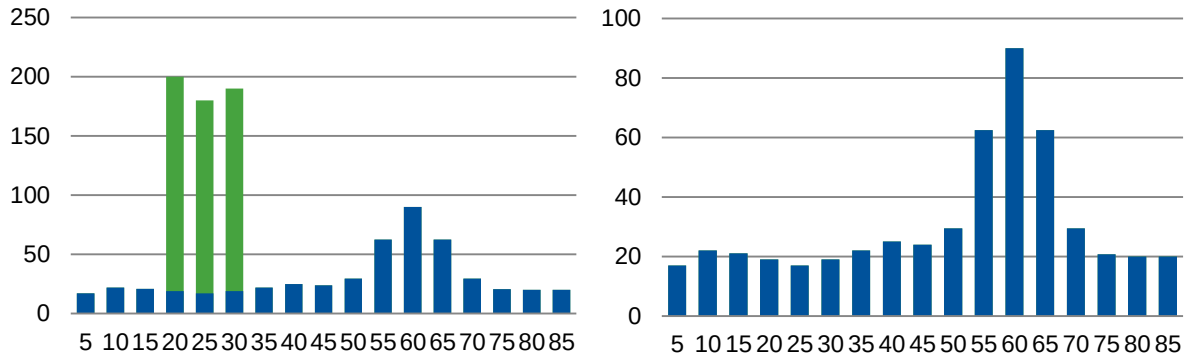


Gartner®

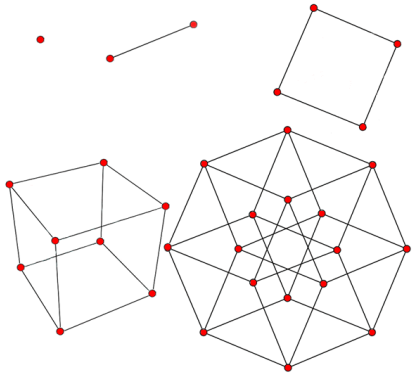
Machine Learning — Further Issues

Boundary problem

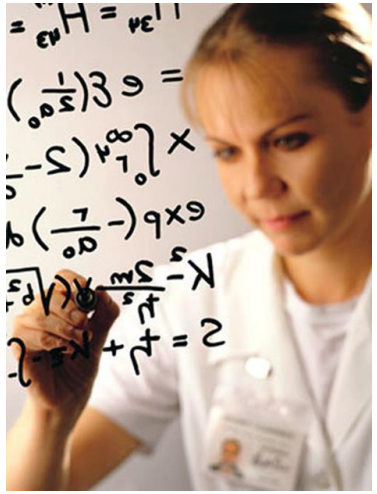
Bias



Poor data



No. of required data points to just occupy the corners	Dimensions
2	1
4	2
8	3
16	4
256	8
65.536	16
~ 4 billion	32
~ 4 billion x 4 billion	64



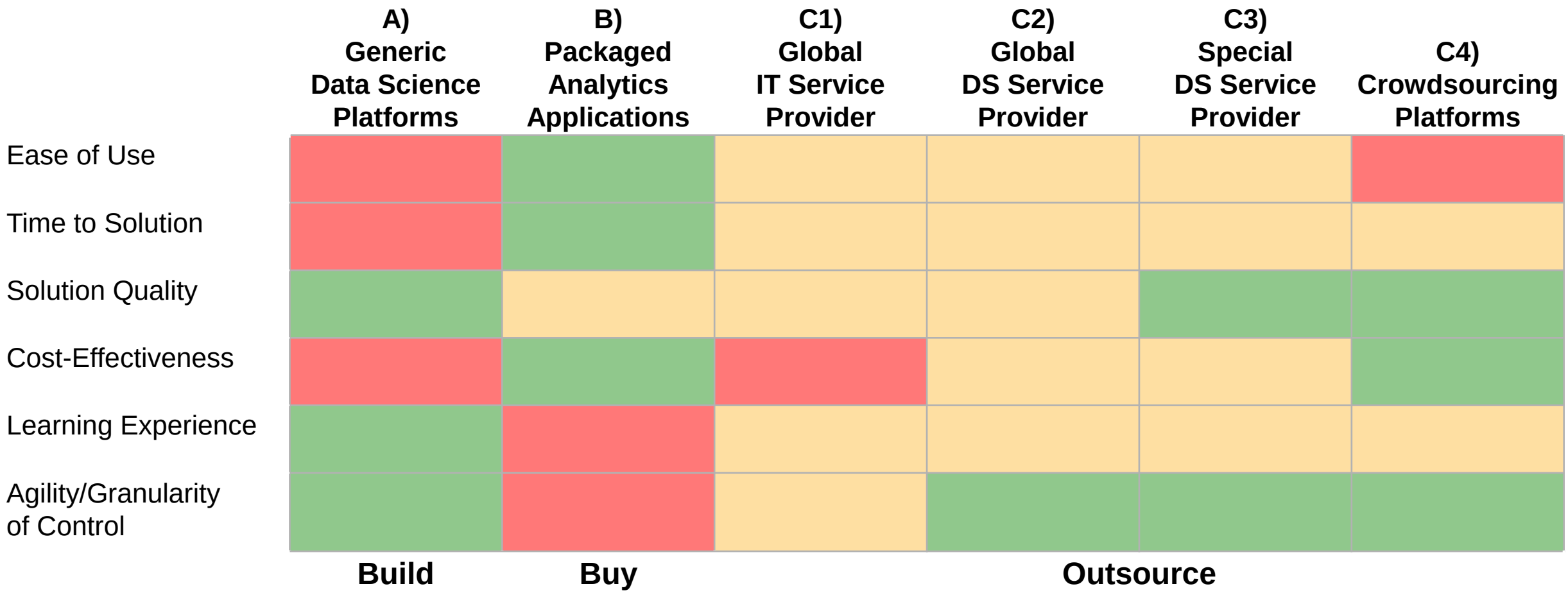
Talent shortage and the danger of automation

Gartner®

Key Issues

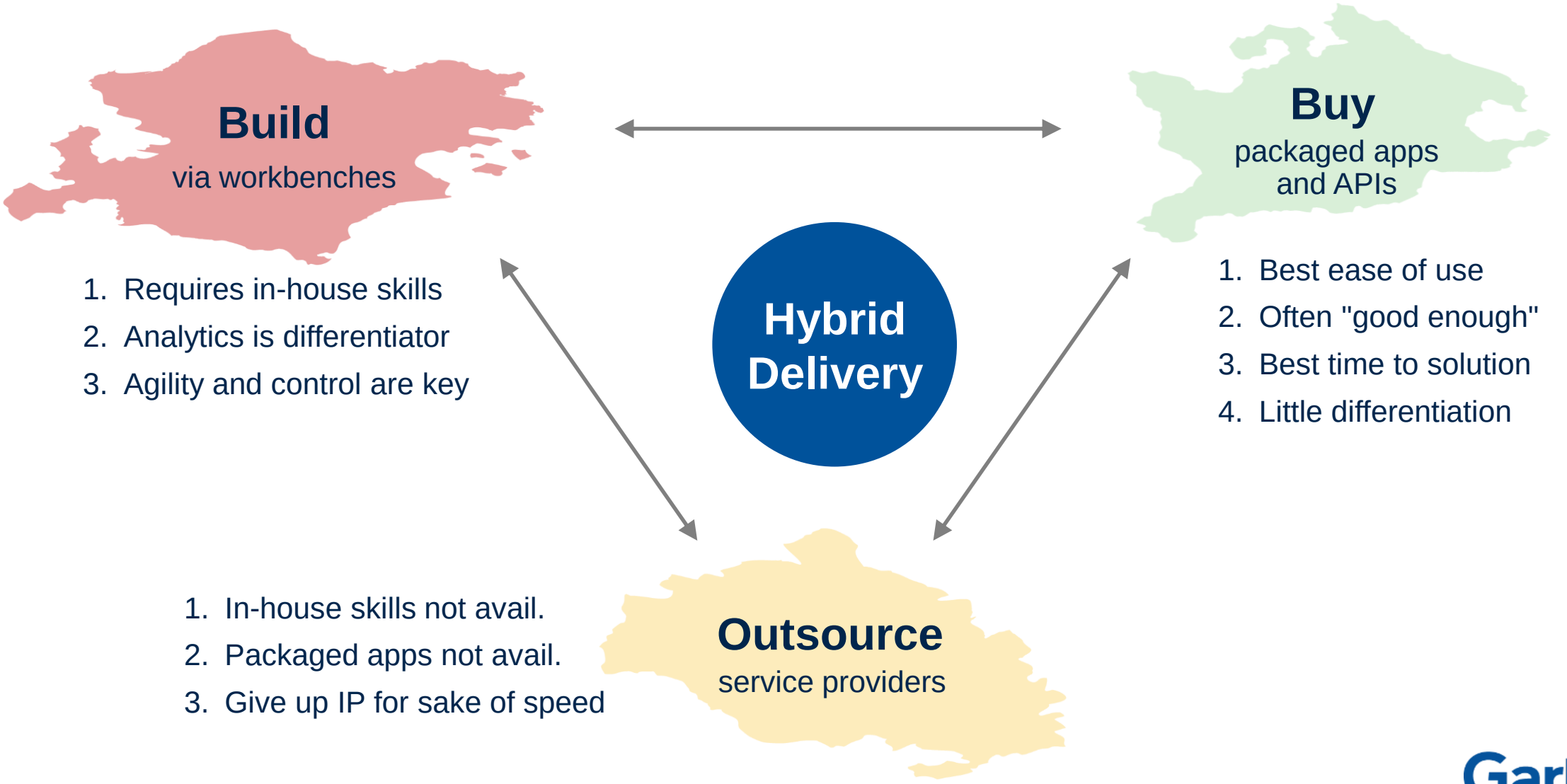
1. How do you navigate all semantics and deal with the shadow of AI?
2. How does it all work?
3. What can you do with data science and machine learning right now?

Build, Buy or Outsource?

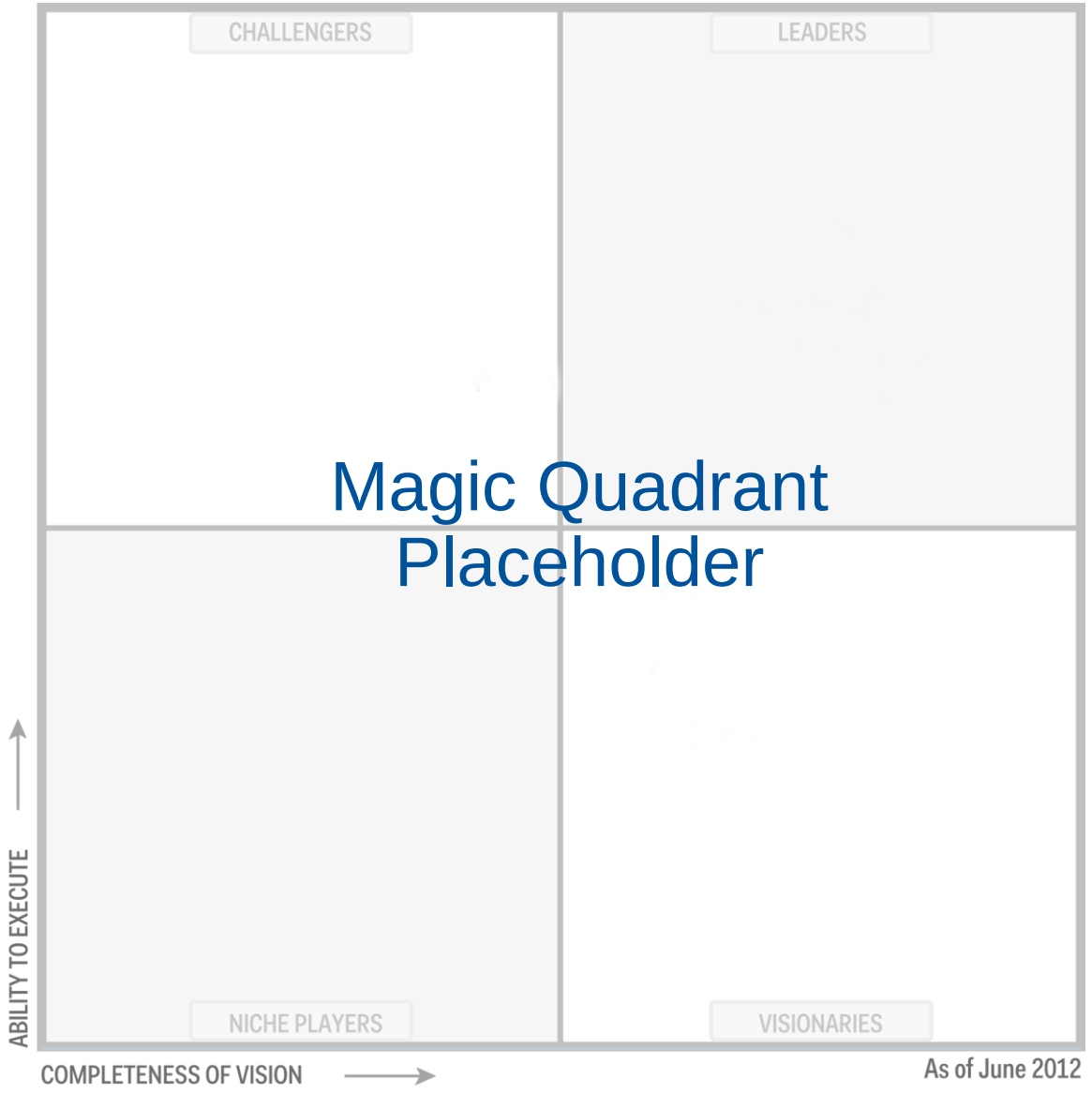


Gartner®

Delivery Models Blurring



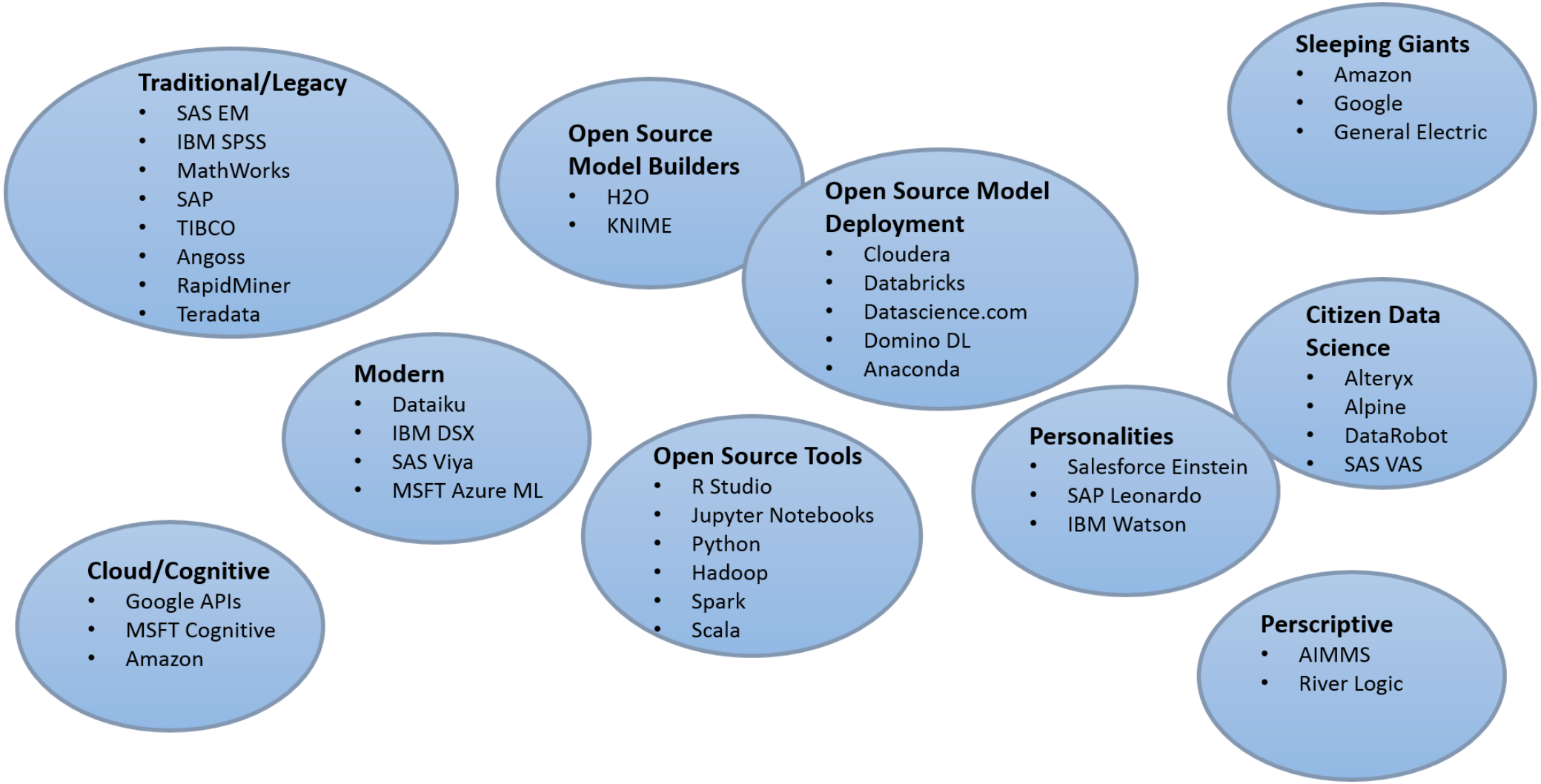
Gartner®



(From "Magic Quadrant for Data Science and Machine Learning Platforms," XX month 20XX)

Gartner®

DS/ML Clusters — Apples to Oranges



Gartner®

Challenges and Pitfalls

Talent Gap

Poor Data Quality

Culture and Territorialism

Scope Creep

Fear and Misunderstanding

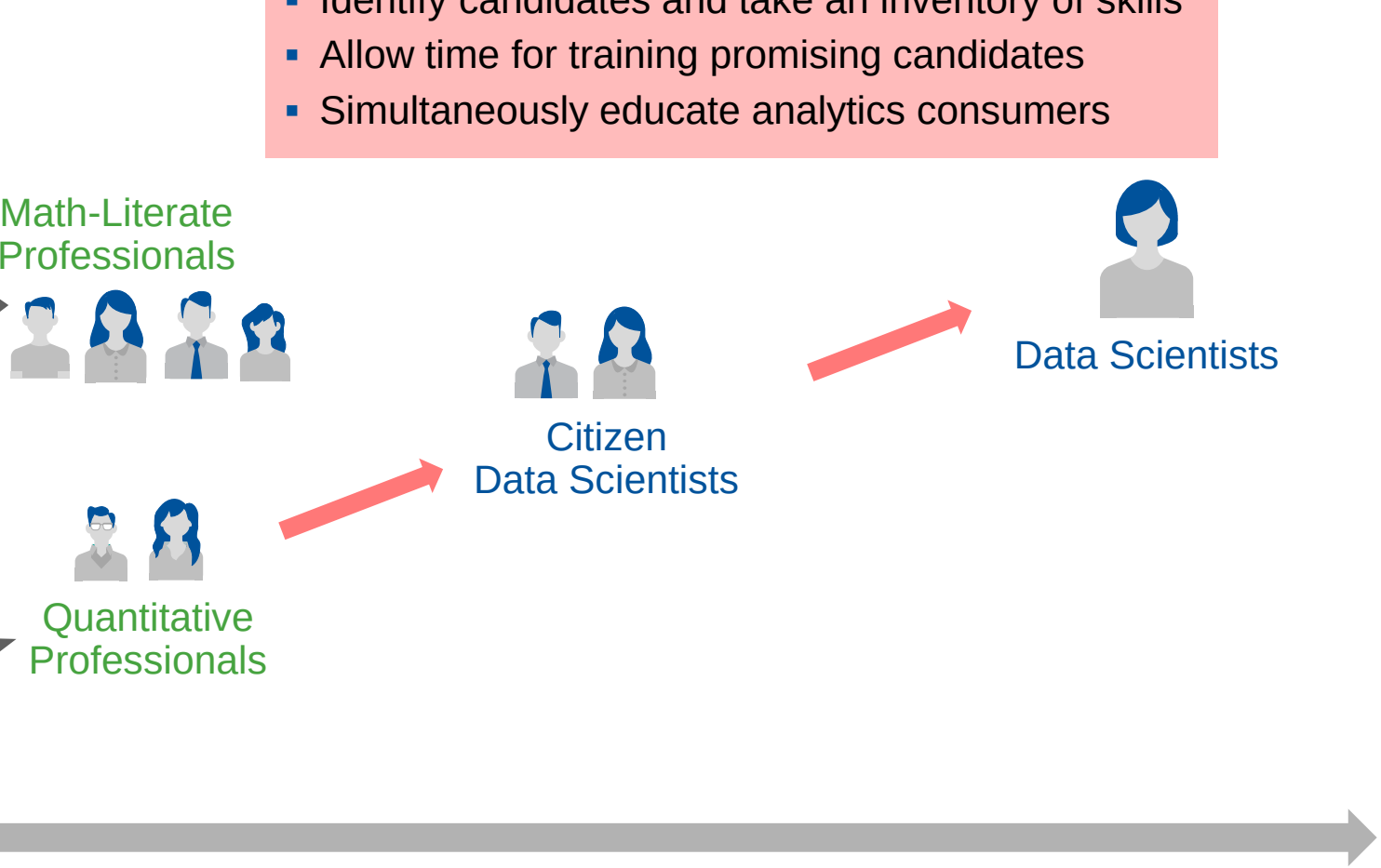
User Inertia and
the Analytical Status Quo

Gartner®

Upskilling

- Physicists
- Chemistrists
- Biologists
- Engineering Disciplines
- Social Scientists
- Computer Scientists
- Statistician
- Operations Researcher
- Mathematicians
- Industrial Engineers
- MBAs
- Astronomers
- Data Analysts
- Actuaries
- Risk Managers
- Control Engineers
- Financial Accountants
- Quality Specialists (Six Sigma)

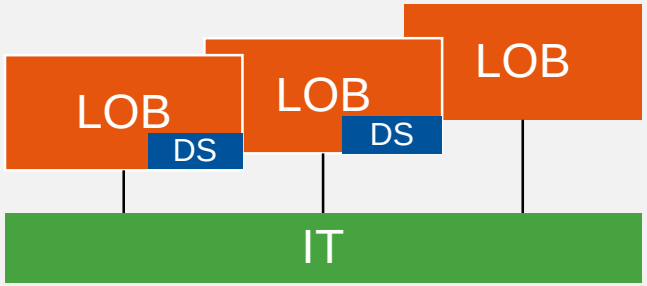
- Cast a wide net
- Identify candidates and take an inventory of skills
- Allow time for training promising candidates
- Simultaneously educate analytics consumers



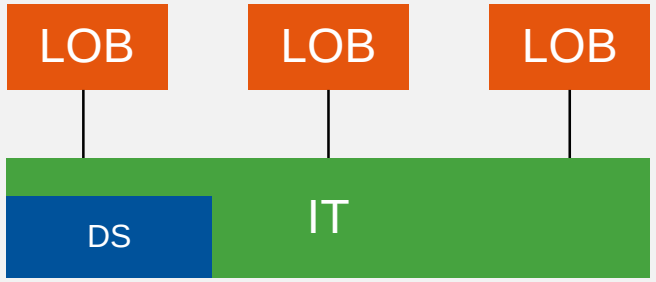
Gartner®

Where to Put the Data Scientists?

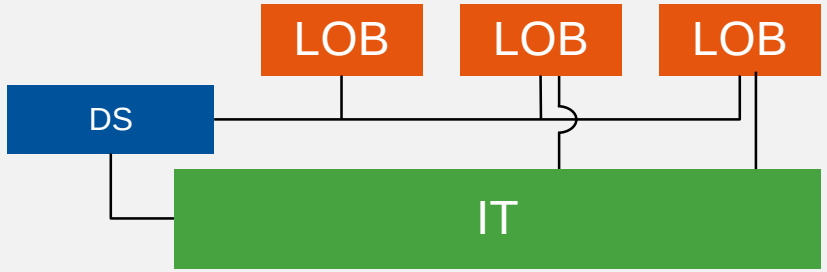
Data Scientists @ LOBs



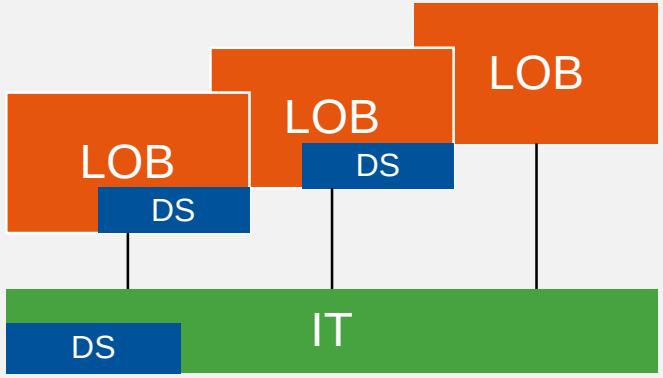
Data Scientists @ IT



Data Scientists as Separate BU

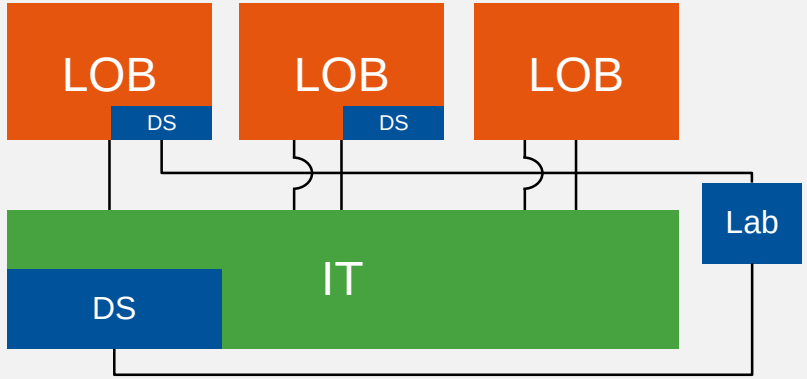


Scattered Experts



- Business Intimacy
- Knowledge Sharing
- Agility
- Cross-Functional View
- Proximity to Process and Data

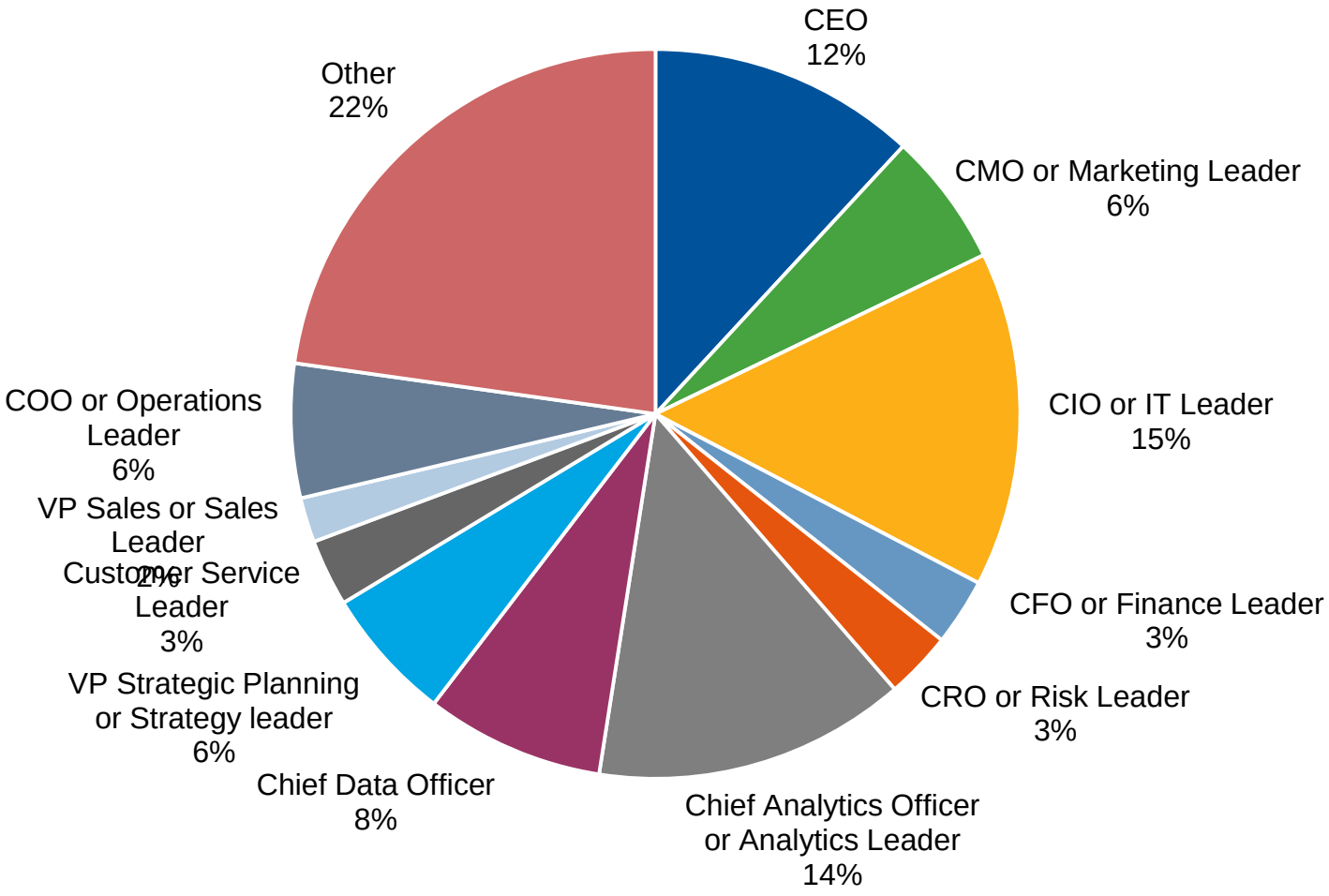
Data Science Lab/CoE



Source: "Organizational Principles for Placing Data Science and Machine Learning Teams," (G00325989)
DS = Data Science; LOB = Line of Business

Gartner®

Data Science Teams are Found Everywhere — Only 15% Report to IT



Recommendations

- ✓ Don't fear the shadow of AI.
- ✓ Cut through the hype, start focused and stay focused.
- ✓ Always look first at packaged applications.
- ✓ Be prepared for significant staffing and communication challenges.
- ✓ Take an inventory of data science skills and support upskilling initiatives.

AI Renaissance or Apocalypse?

Gartner®

Recommended Gartner Research

- **How to Start a Machine-Learning Initiative With Less Anxiety**
  Svetlana Sicular (G00331893)
- **Leading Upskilling Initiatives in Data Science and Machine Learning**
  Peter Krensky, Shubhangi Vashisth and Douglas Laney (G00334219)
- **Innovation Insight for Deep Learning**
  Alexander Linden, Tom Austin and Svetlana Sicular (G00319191)
- **Machine Learning: FAQ From Clients**
  Shubhangi Vashisth, Alexander Linden and Others (G00327948)
- **Hype Hurts: Steering Clear of Dangerous AI Myths**
  Tom Austin, Alexander Linden and Mike Rollings (G00324274)

For information, please contact your Gartner representative.

Gartner®