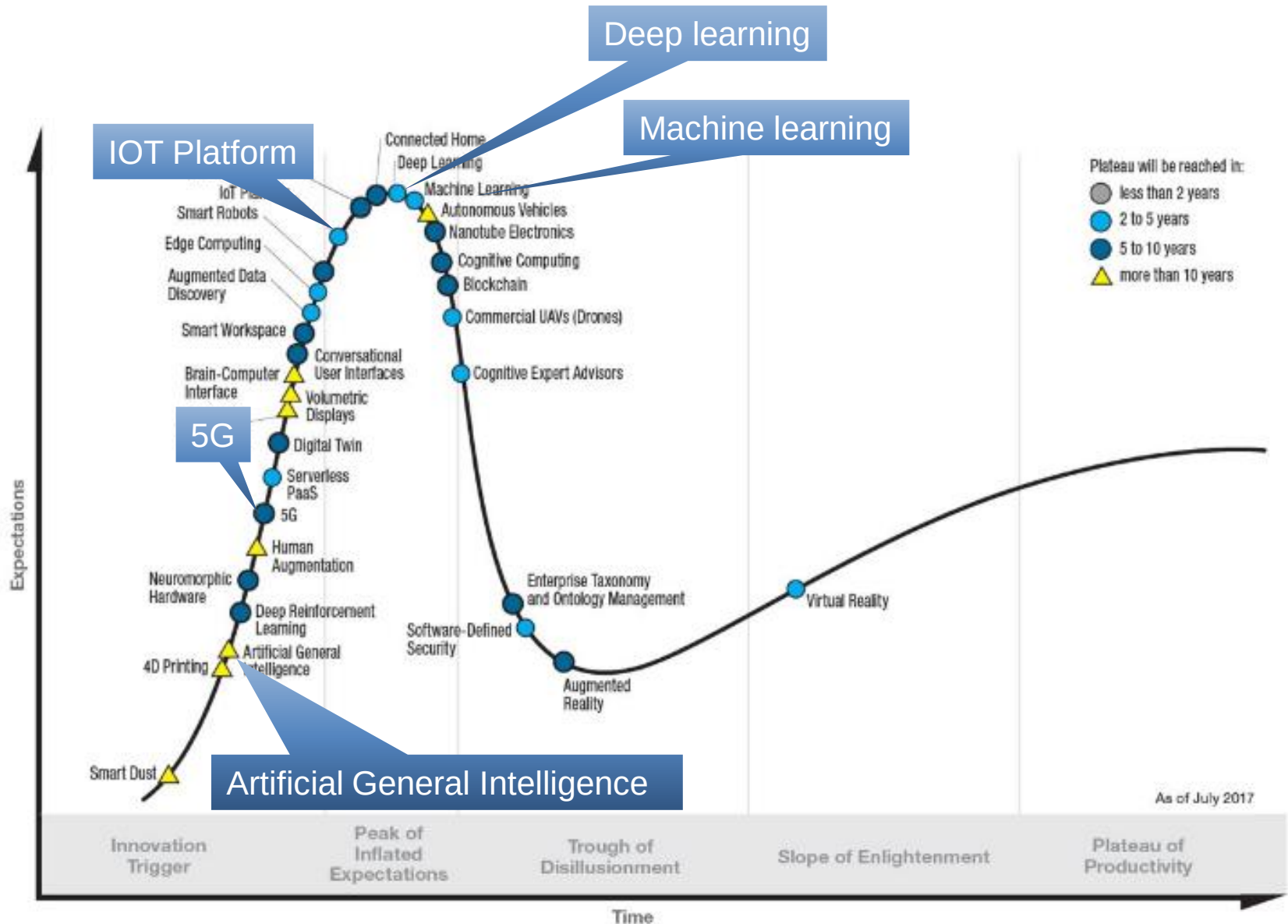


BIG DATA ANALYTICS

By Assoc. Prof. Dr. Tiranee Achalakul
Ministry of Digital Economy and Society



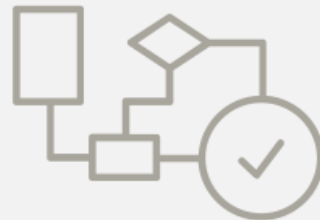
Gartner's hype cycle for Emerging Technologies 2017



What has changed to make digital technology so powerful today?



**NEW
ALGORITHMS**



COMPUTING





An example of the 3 waves in action

Convergence of available technologies and a demanding marketplace



Branch Banking

Back-office automation and secure data processing



ATM

Secure online transaction processing with distributed computing



e-business

Connect existing IT infrastructure with the Web



Cloud

An open, standards-based, dynamic infrastructure

1960s

1970s

1990s

2010s

```
graph TD; PC[Prospective Customer] -- "Ask Questions" --> CC[Call Center]; CC -- "Enrollment" --> EC[Enrolled Customer]; EC -- "Purchase" --> OS[Online Store]; OS -- "Add to Cart" --> C[Cart]; C -- "Payment" --> P[Paid]; P -- "Receipt" --> EC;
```

Cart

PERSONALIZED EXPERIENCE

A lot of data must be collected



Shopping cart

Wish list and Previous purchases

Items rated and reviewed

Geo-location

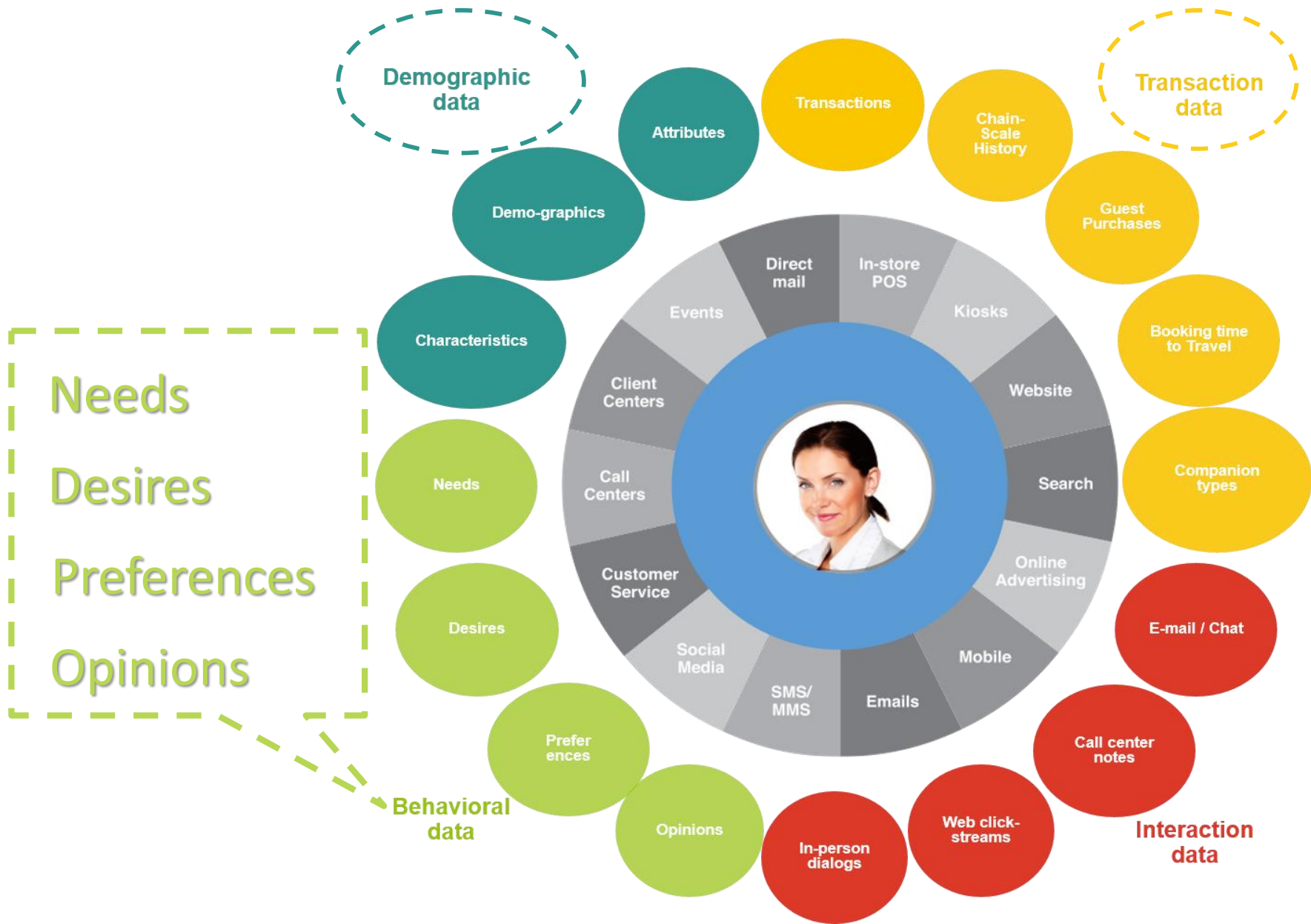
Time-on-site and Duration of views

Links clicked & Text Searched

Telephone inquiries

Responses to marketing materials

Social media posting



Internet of Things

MORE SERVICE FOR STORES
NEW IMPETUS FOR YOUR TURNOVER



T · Systems ·

Technological Roots of IoT

- Embedded devices have existed for years.
- They do lots of (smart) things but still not part of IoT until they go online
- Far greater values with connectivity

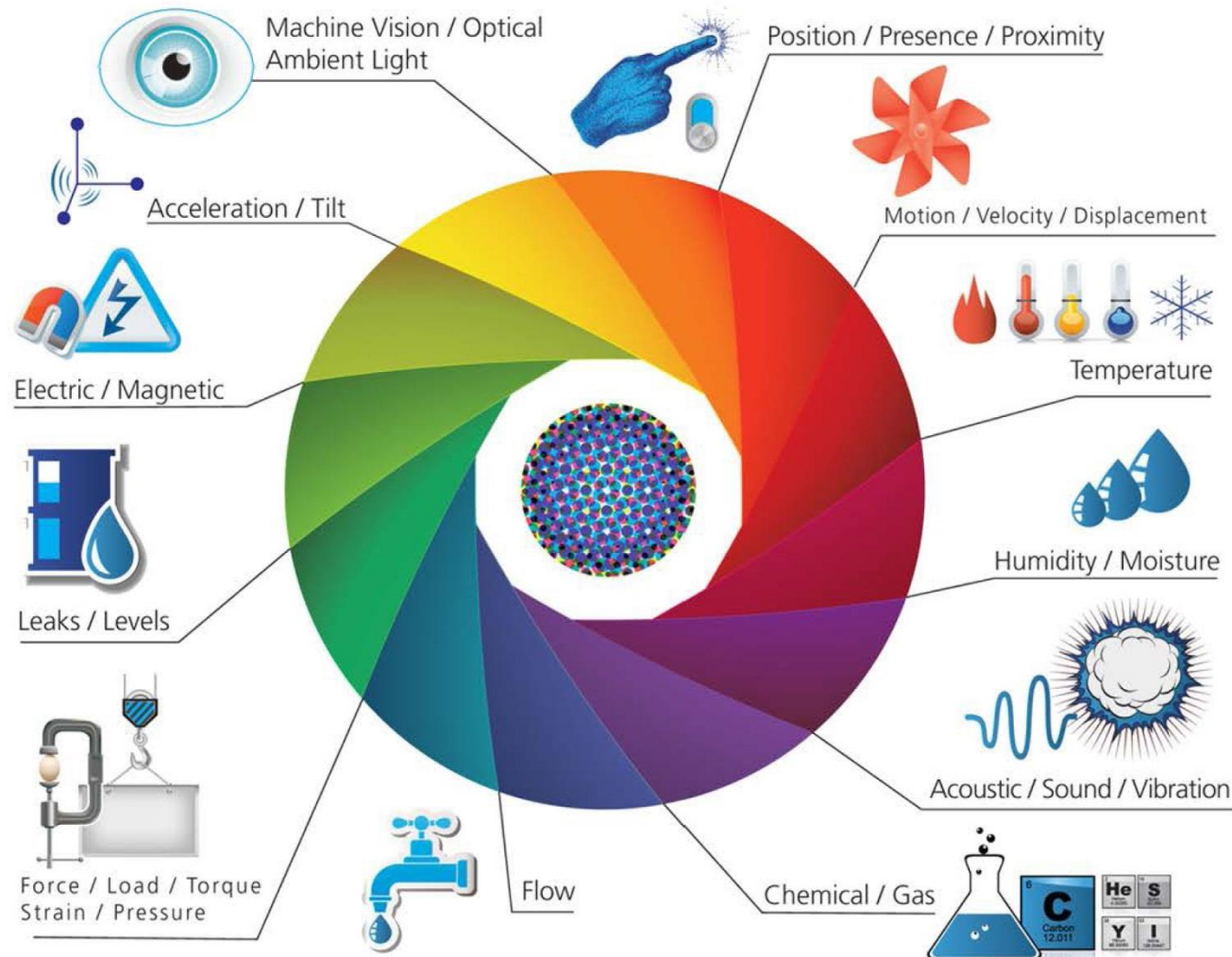
With Connectivity, Nest Learning Thermostat can

- “learns” how and when the user likes to adjust the temperature.
- optimize the house’s temperature for energy efficiency.
- allow utility companies to offer incentives for using less power at peak times



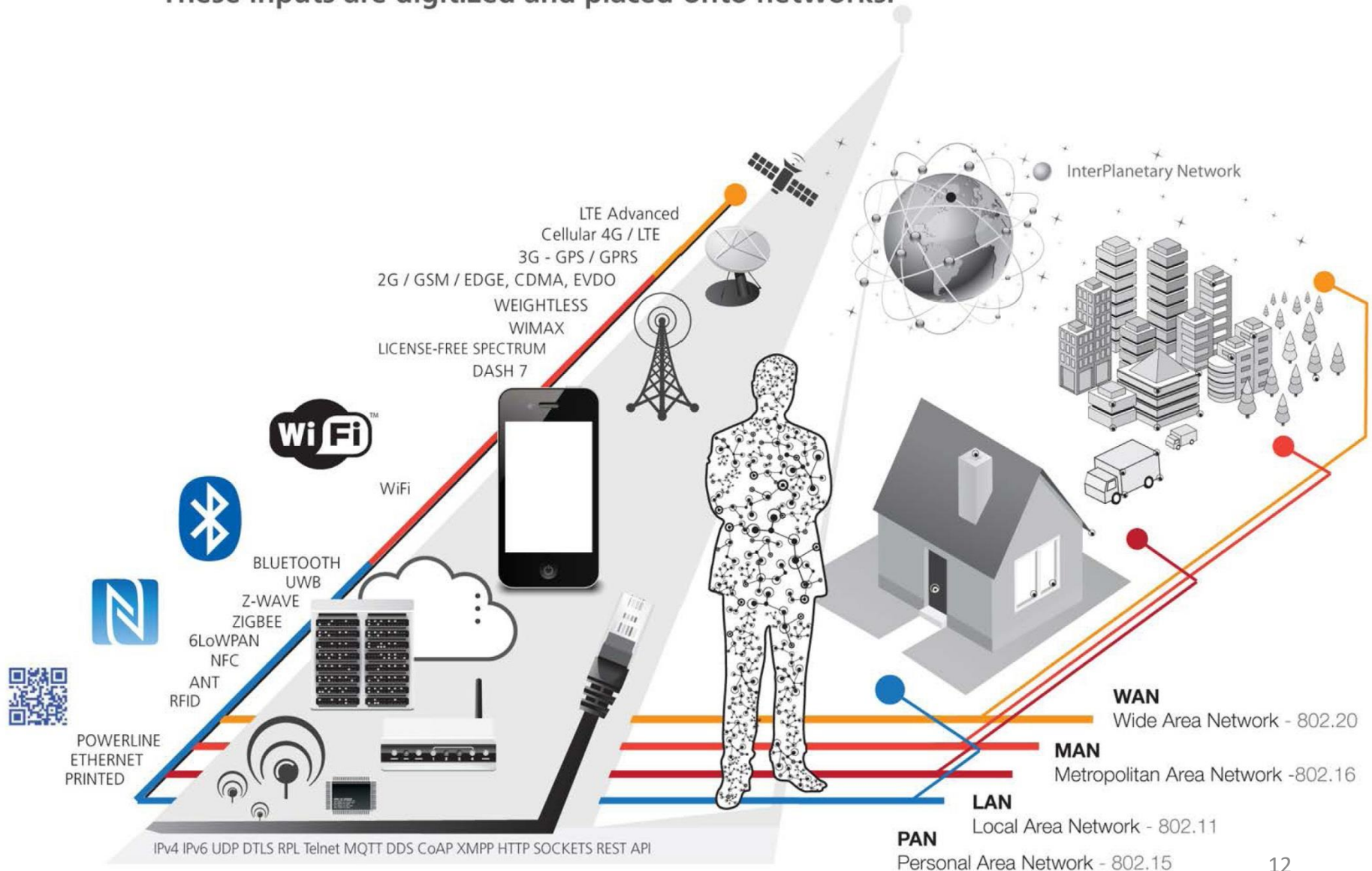
1 SENSORS & ACTUATORS

We are giving our world a digital nervous system. Location data using GPS sensors. Eyes and ears using cameras and microphones, along with sensory organs that can measure everything from temperature to pressure changes.



2 CONNECTIVITY

These inputs are digitized and placed onto networks.



FROM HEAD TO TOE WEARABLE TECHNOLOGY

SHIRT

Conductive thread means a computer is literally built into the fabric of the shirt, providing the processing power for all the other wearable gadgets.

WRISTBAND

A sensor that tracks movement to determine the number of steps taken through the day – 10,000 is ideal – and how much sleep the wearer gets at night.

TROUSERS

Also made with conductive thread, the trousers take the energy generated by movement and use it to power the other gadgets.

GLASSES

Overlays navigation directions and information about points of interest directly on to the wearer's field of vision.

WRISTWATCH

Vibrates when a message arrives and displays it on the watch face. Tells the time too.

HAND

Embedded under the skin is a chip containing medical records, passport data and credit records. Information is transferred by waving the hand over a suitable scanner.

SHOES

GPS chip provides directions using LED lights in each shoe: the left shoe indicates direction, while the right shows distance.



GRAPHIC: JOHN BRADLEY

Wearable Wireless Thermometer



TEMP TRAQ™

Get the most out of your medication by “ingestible sensors”

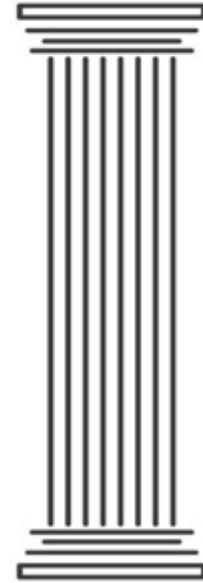
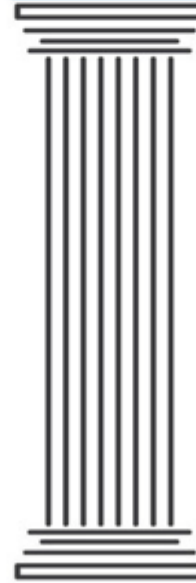
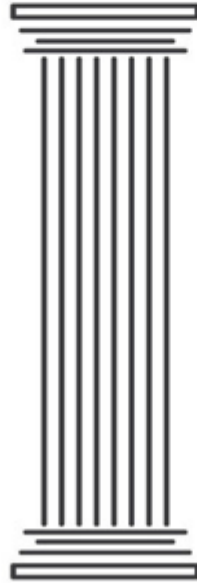


Proteus Digital Health™



BIG Data

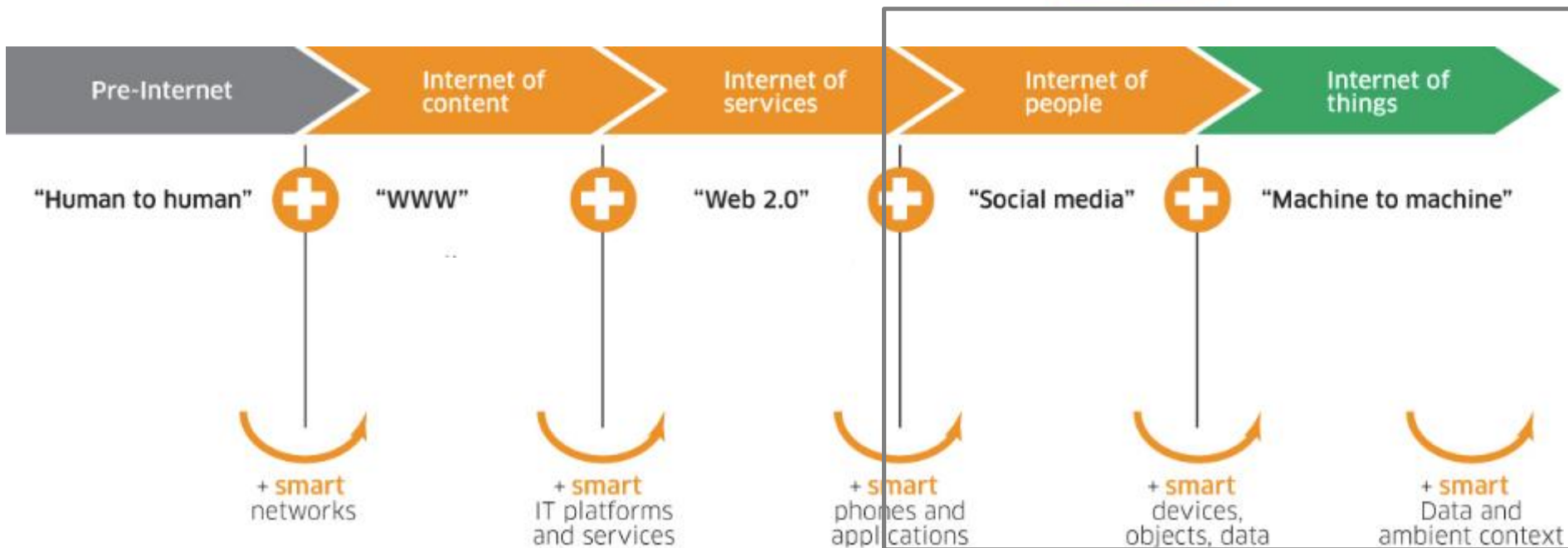
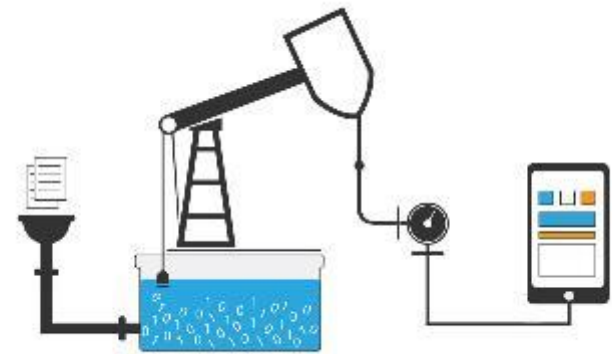
00101011101101010001010110
00101011101101010001010110
00101011101101010001010110



Digital technologies that change our way of lives

DATA IS THE NEW OIL

The rapid development of technology led to the **explosive growth of data** in almost every industry and business area.



Data of great
Volume, Variety, and Velocity

BIG DATA

An umbrella term for all sorts of data



Structured

Unstructured

Hospital records, Citizen data, Insurance claim

Digital data to the government archived paperwork

Social media (needs, trends, and opinions)

Sensor data (heath, weather, geo-location, etc.)

WHERE TO LOOK FOR DATA



Archives

Scanned documents, statements, medical records, e-mails etc..



Docs

XLS, PDF, CSV, HTML, JSON etc.



Business Apps

CRM, ERP systems, HR, project management etc.



Media

Images, video, audio etc.



Social Networks

Twitter, Facebook, Google+, LinkedIn etc.



Public Web

Wikipedia, news, weather, public finance etc



Data Storages

RDBMS, NoSQL, Hadoop, file systems etc.



Machine Log Data

Application logs, event logs, server data, CDRs, clickstream data etc.



Sensor Data

Smart electric meters, medical devices, car sensors, road cameras etc.

Descriptive Analytics

What happened; How many;
how often; where



Business Intelligence Tools (BI)



Data

Warehouse

Extract, Transform & Load (ETL)

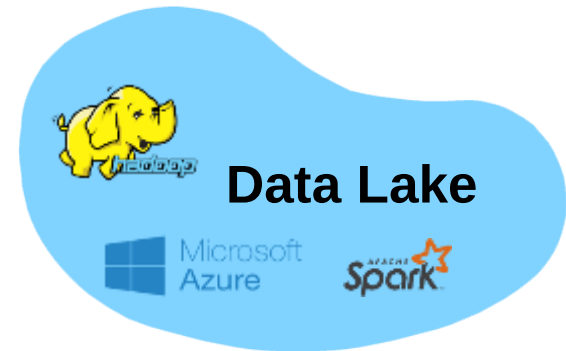


Predictive Analytics

Project what will happen;
possible outcome indication



Developer Environment

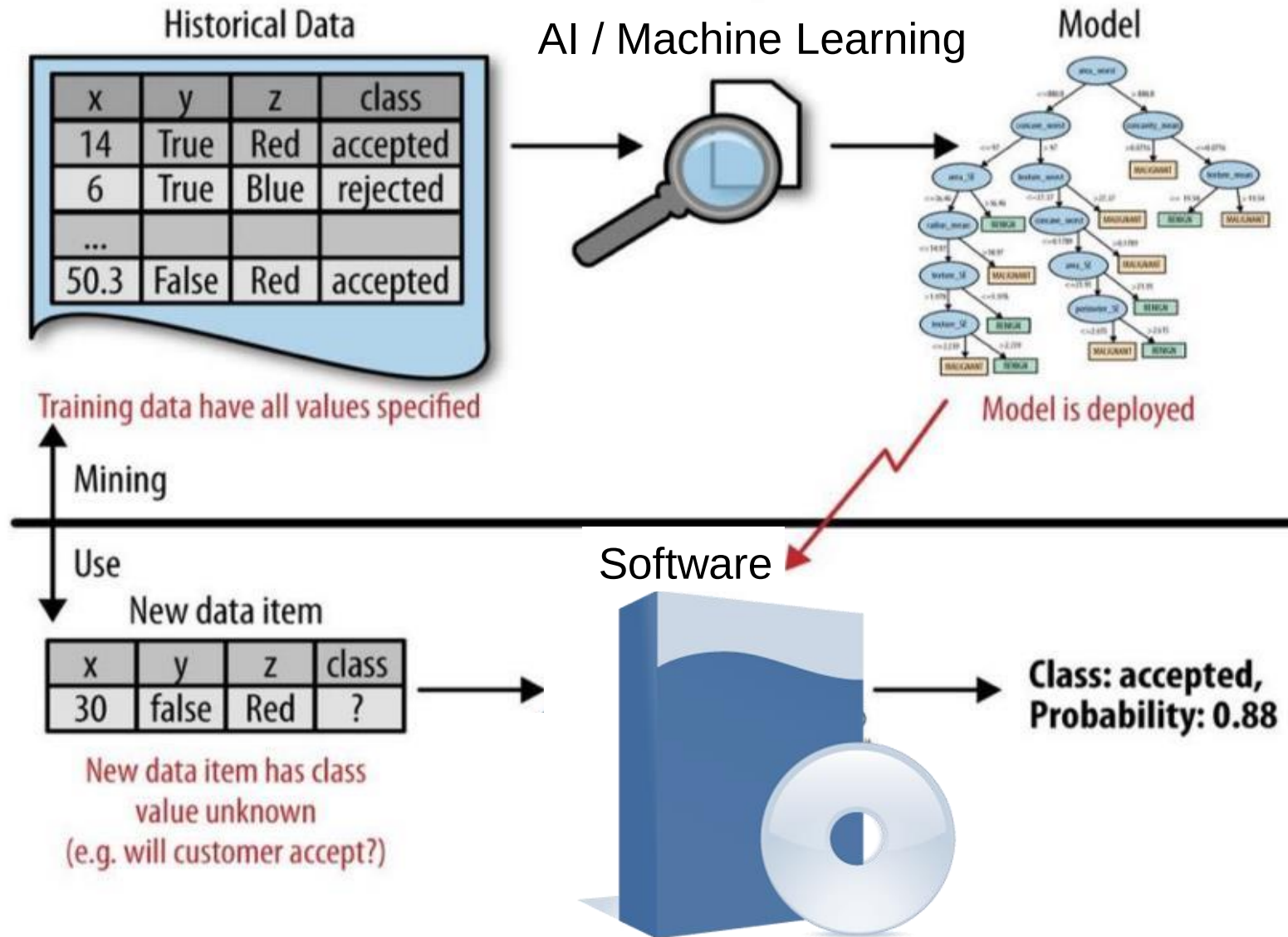


Data Lake

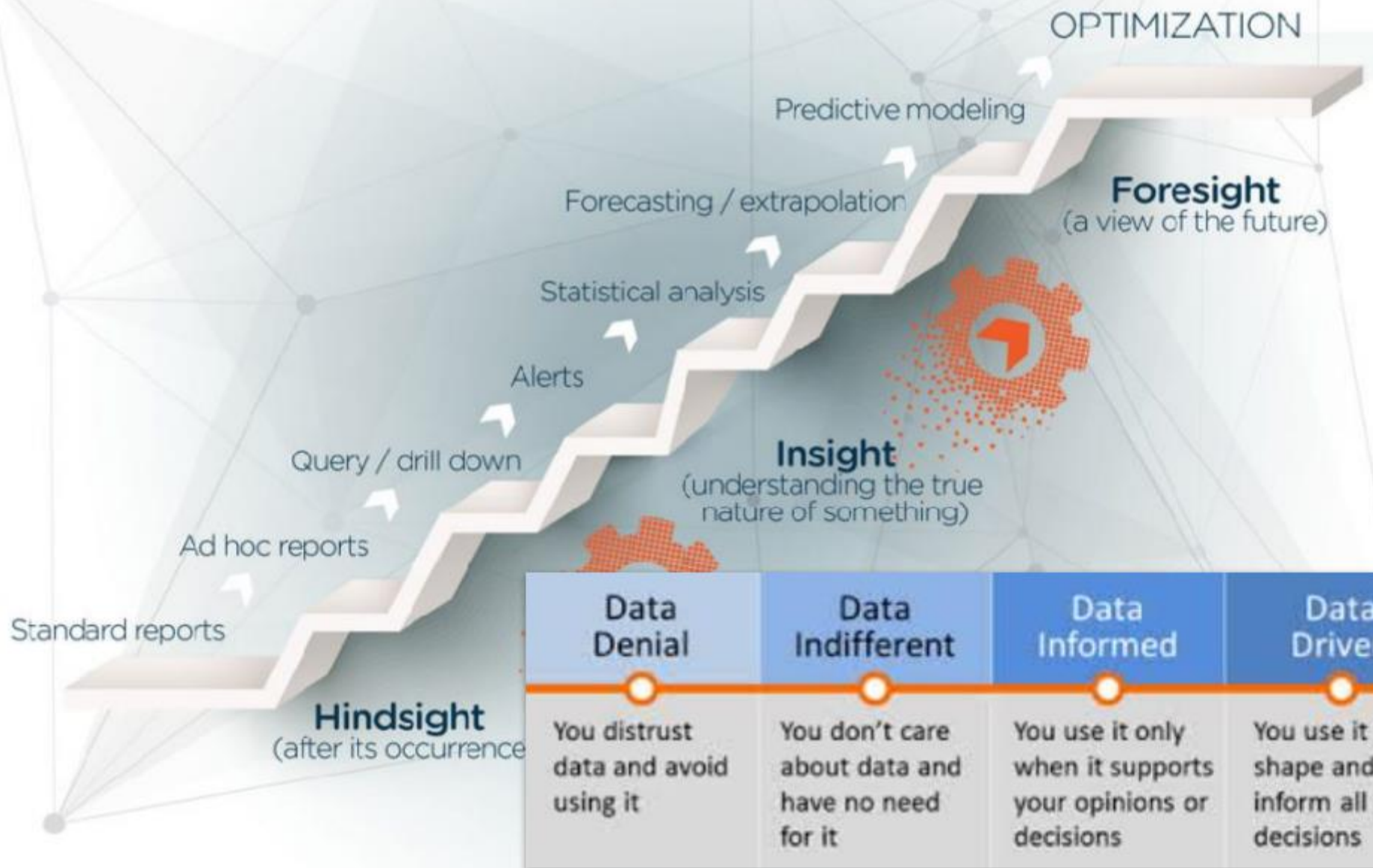
File Copy



PREDICTIVE ANALYTICS



Data Driven Decision Making

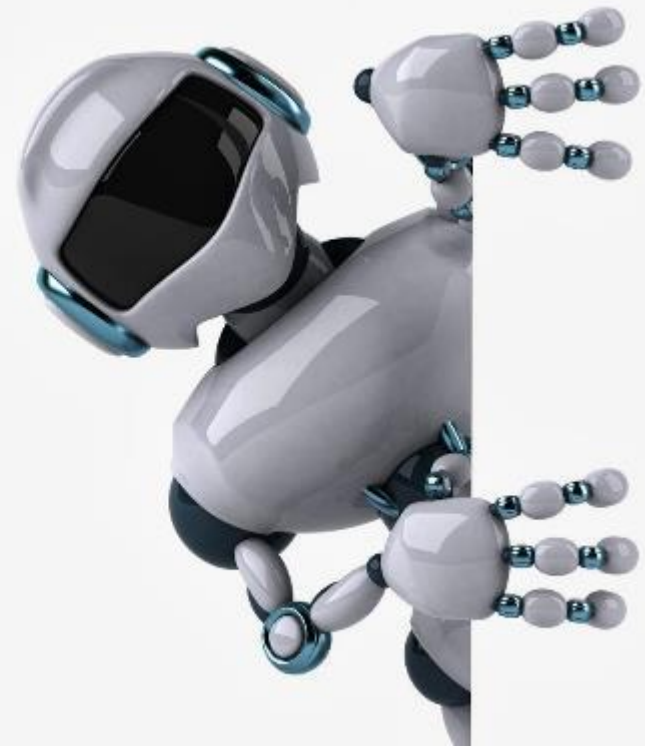


DATA STUDIO



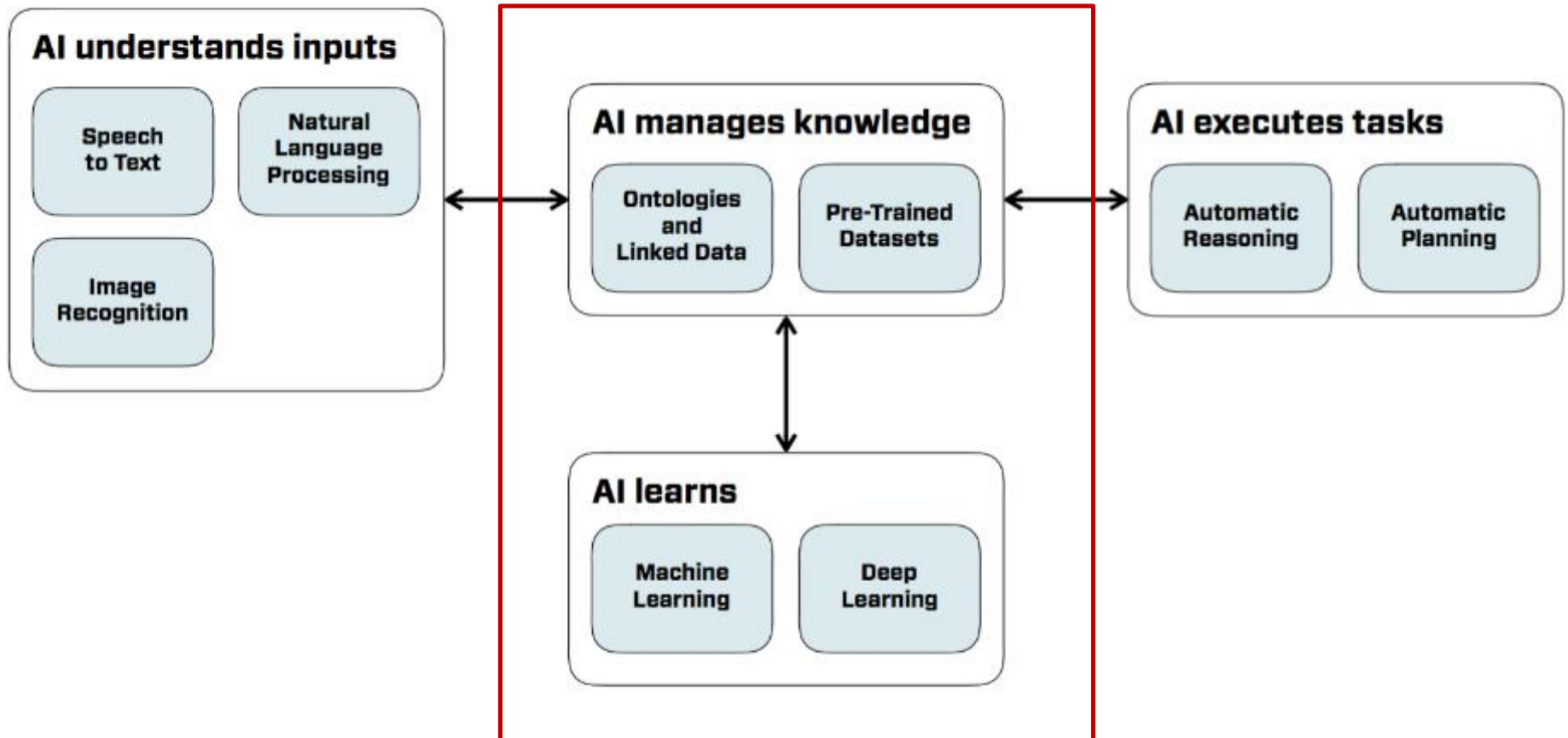
MACHINE LEARNING and AI

Learn from data and make predictions about data by using statistics to develop self learning algorithm





PUTTING AIs TO WORK



Real-time Face Recognition



City Surveillance



Video Analytics for Profiling

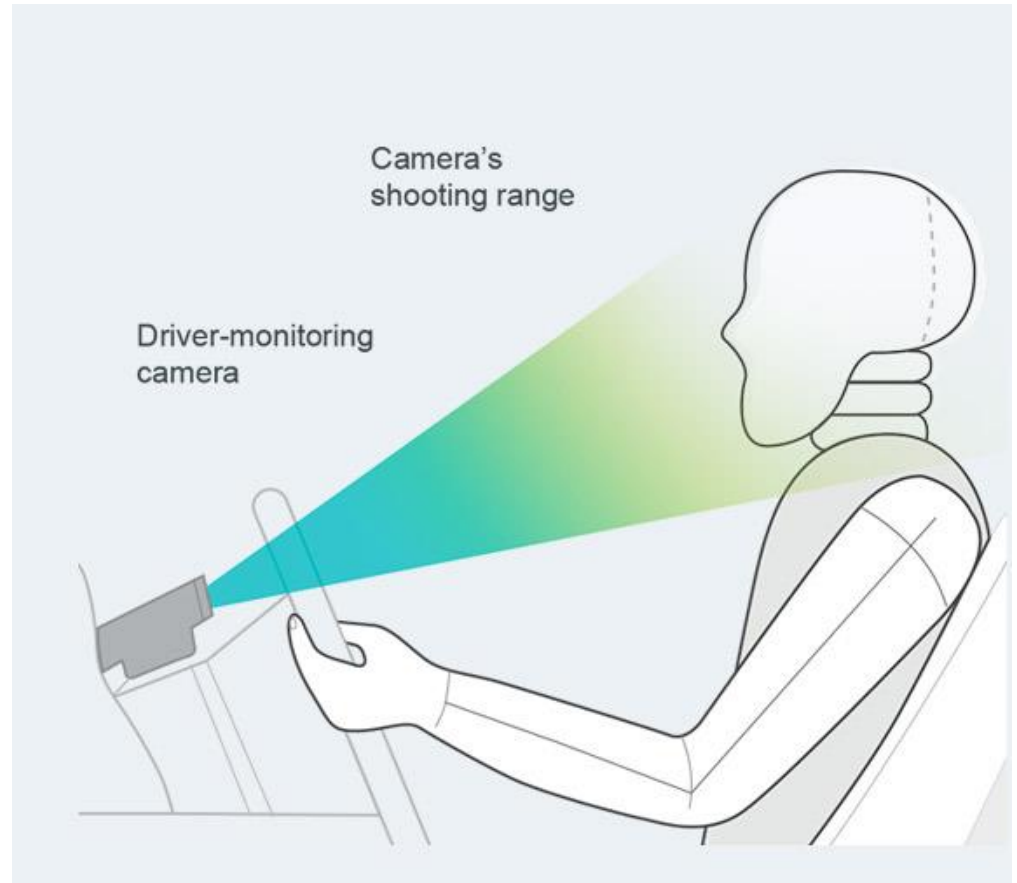


DOZE OFF DETECTION

With Machine Learning

Supervised machine learning

Use twenty second stretches of driving data and the labels are values of visually scored drowsiness. The goal is to predict future dozing off



Three Levels of Intelligence

Artificial Narrow Intelligence

Specialized in
one specific
area.

Artificial General Intelligence

Specialized in
all areas.

Artificial Super Intelligence

Smarter than
human in every
way.

Three Levels of Intelligence

Artificial Narrow Intelligence

Specialized in
one specific
area.

Artificial General Intelligence

Specialized in
all areas.

Artificial Super Intelligence

Smarter than
human in
every way.

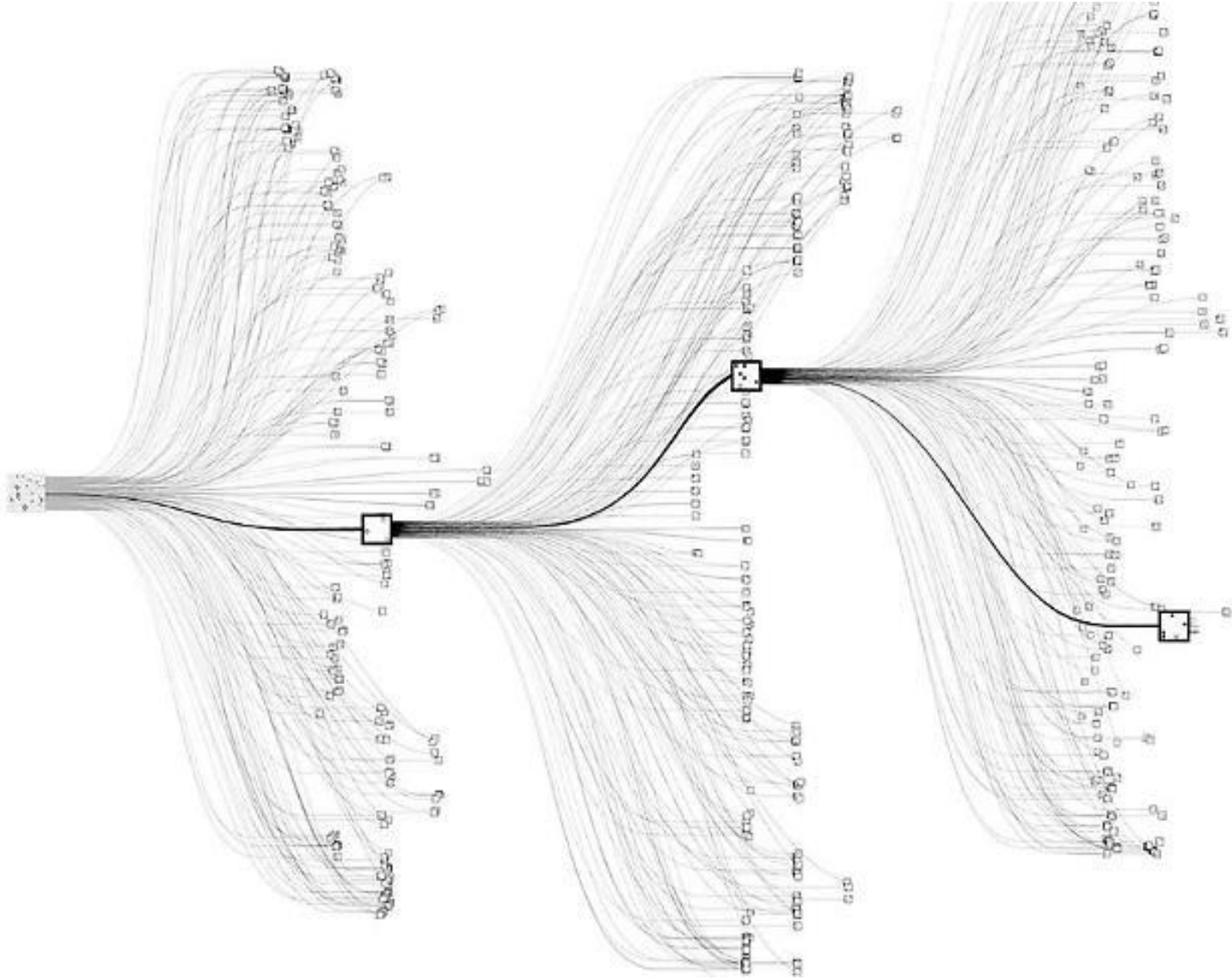
Work in Progress

Singularity

ALPHA GO



ALPHA GO



ALPHA GO ZERO



A Step Towards Super Intelligence



David (Hit Theater in 2001)



Sophia (Activated in 2015)

SOPHIA

by Hanson Robotics



- A service robot designed to care for the elderly, disabilities, children with special needs.
- Built based on Audrey Hepburn and Hanson's wife
- 62 facial expressions and “Character Engine AI”, personality software

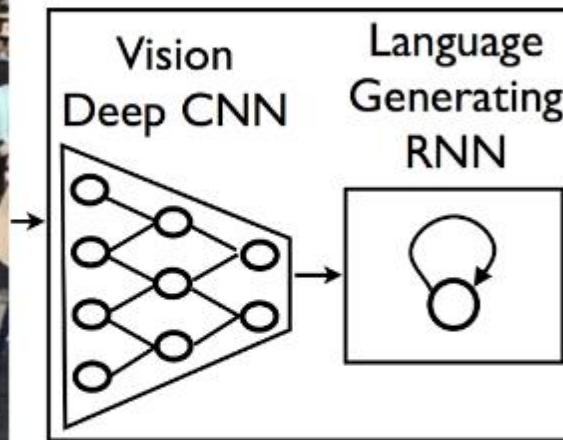
DEEP LEARNING

Deep Learning is the Most Exciting Breakthrough in
Modern Machine Learning

- Learn complex features from big images
- Learn complex action plans for reinforcement learning
- Learn to create and generate
- Learn complex unstructured data such as language, videos, etc.

State of the Art Deep Learning

Generate text from images



A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.

Multi-modal
Processing

State of the Art Deep Learning

Generate high-resolution images based on text

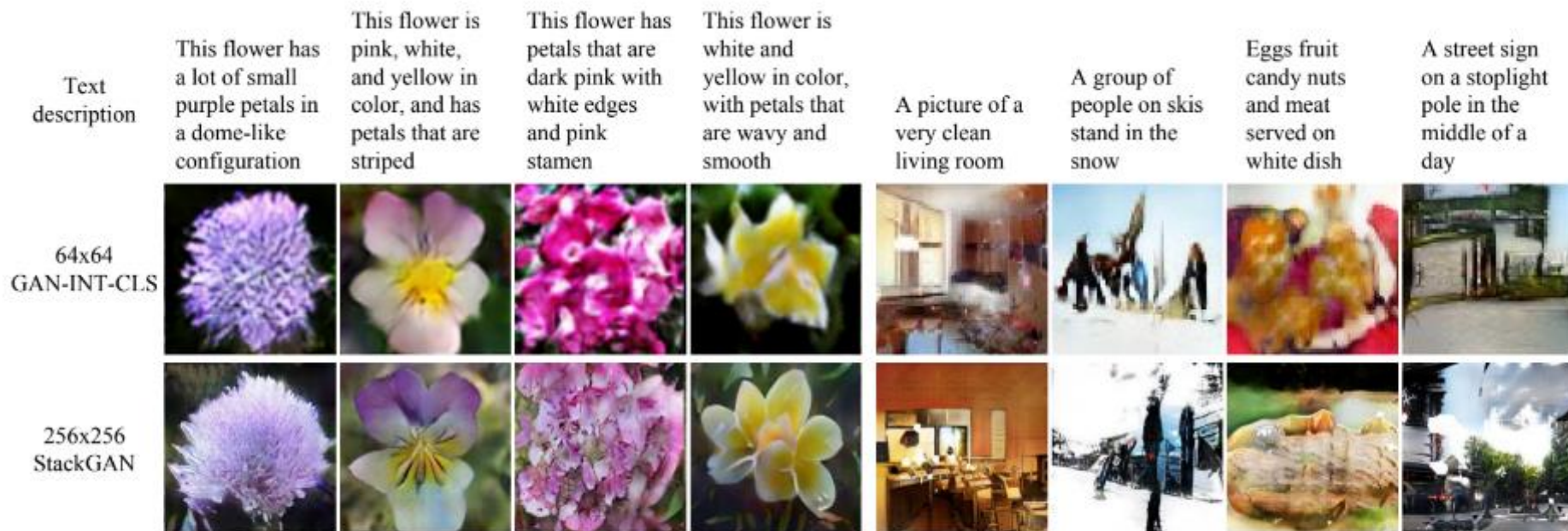


Figure 4. Example results by our StackGAN and GAN-INT-CLS [26] conditioned on text descriptions from Oxford-102 test set (leftmost four columns) and COCO validation set (rightmost four columns).

Deep Neural Networks

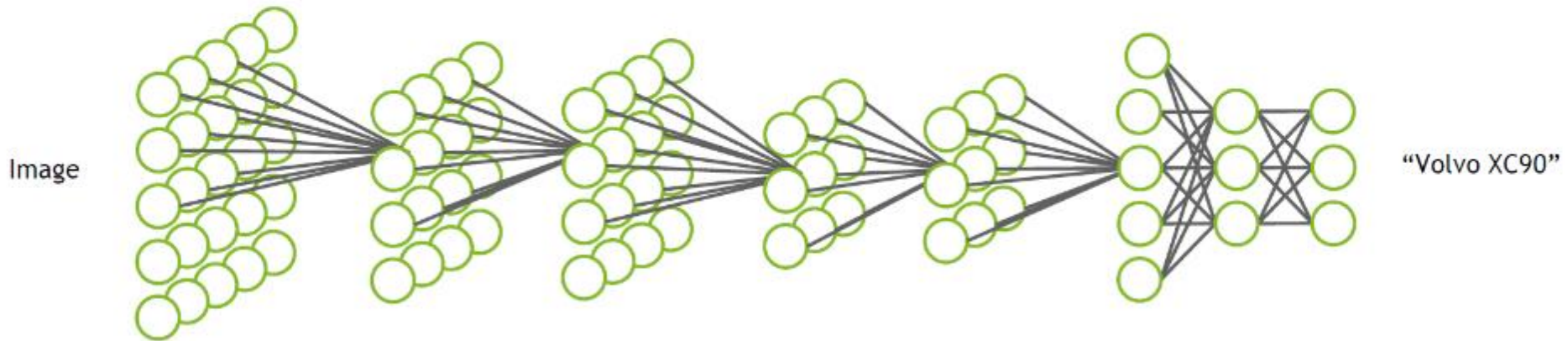
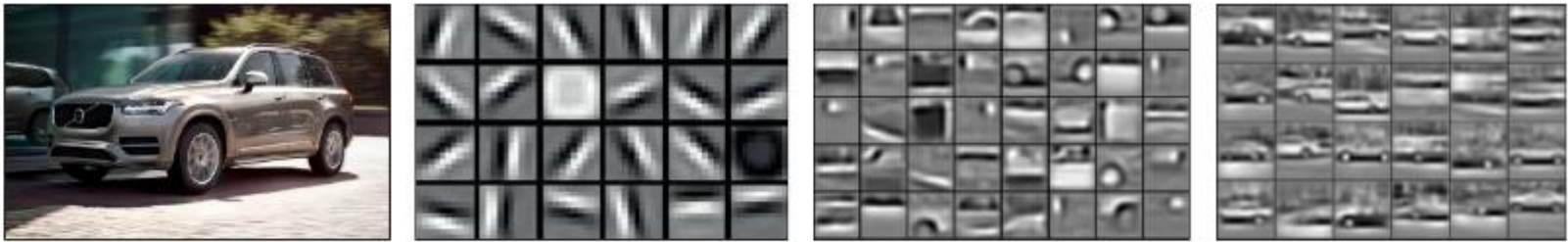
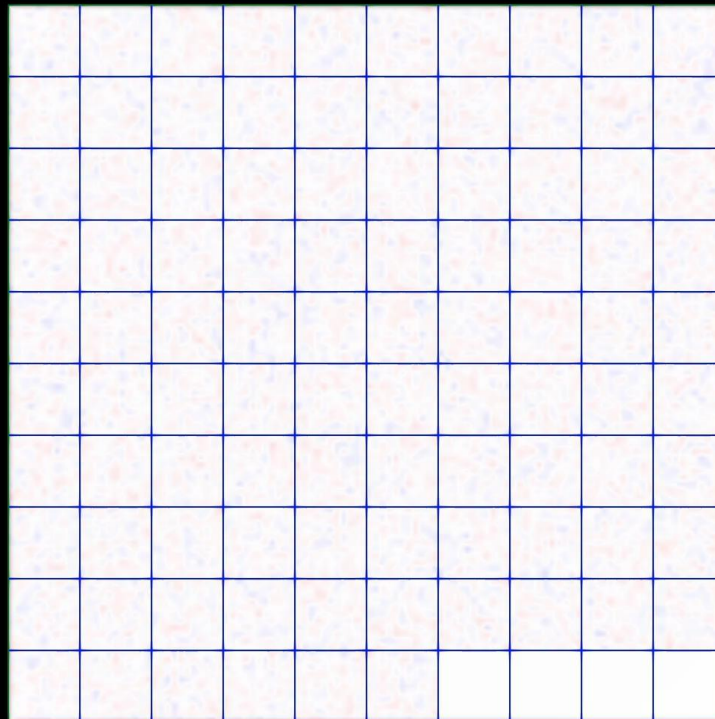


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011, Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

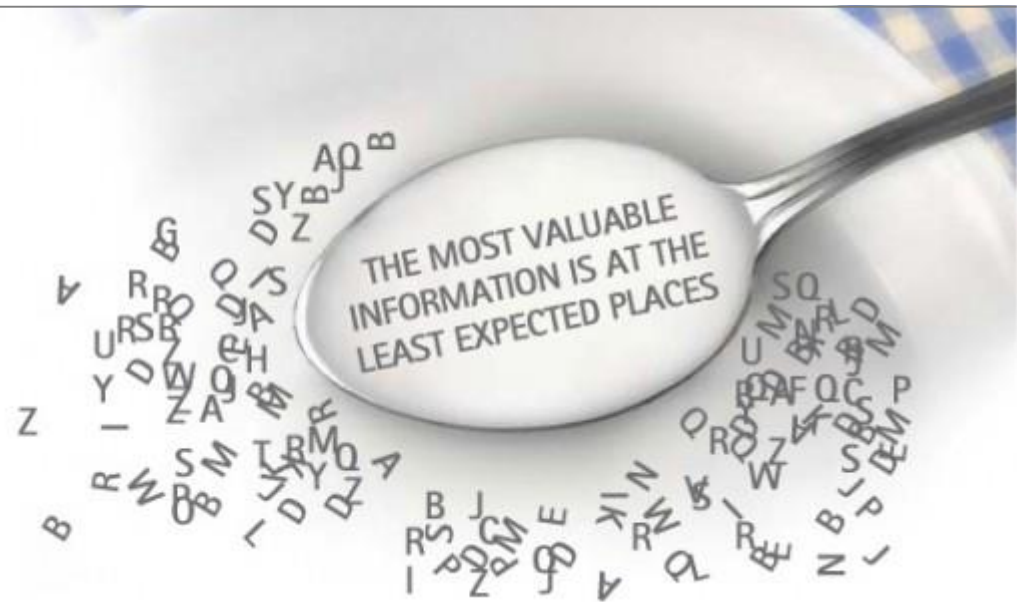
HOW DOES AI LEARN

Network iteration: 166/450026



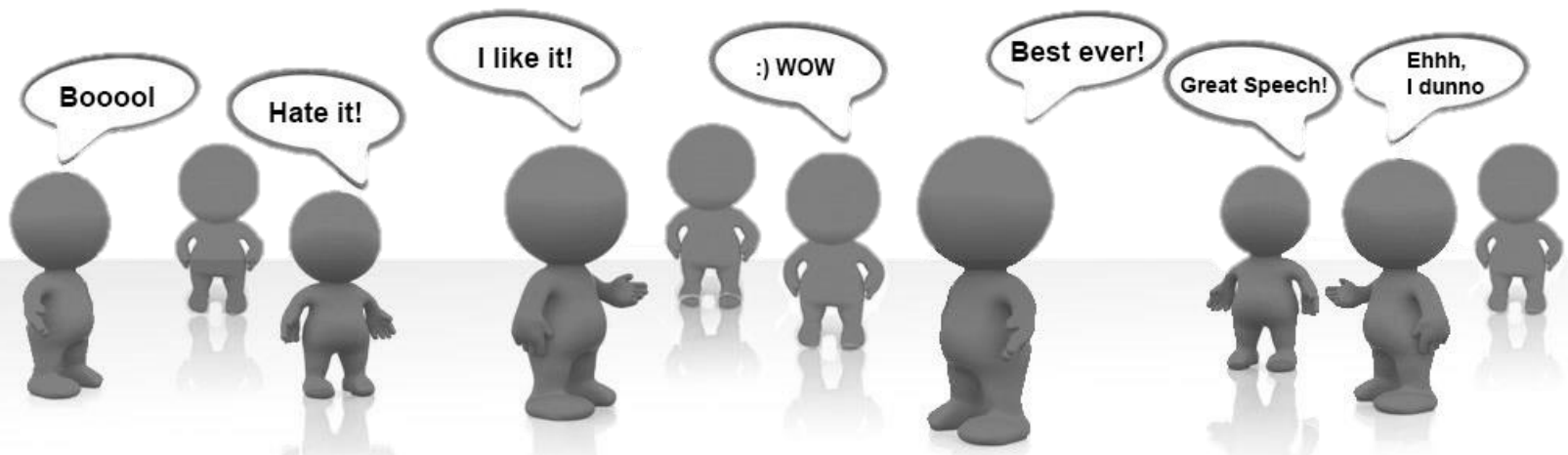
TEXT MINING AND NLP

Deriving high-quality information from text by devising of patterns and trends through means such as statistical pattern learning.



SENTIMENT ANALYSIS

- Use text analytics to key citizen concerns and sentiments
- Assess current situations to create a set of keywords to gather social media post related to crimes.
- Determine the sentiment with respect to topics or the overall contextual polarity of a post/comment related to crimes.



Happiness Index

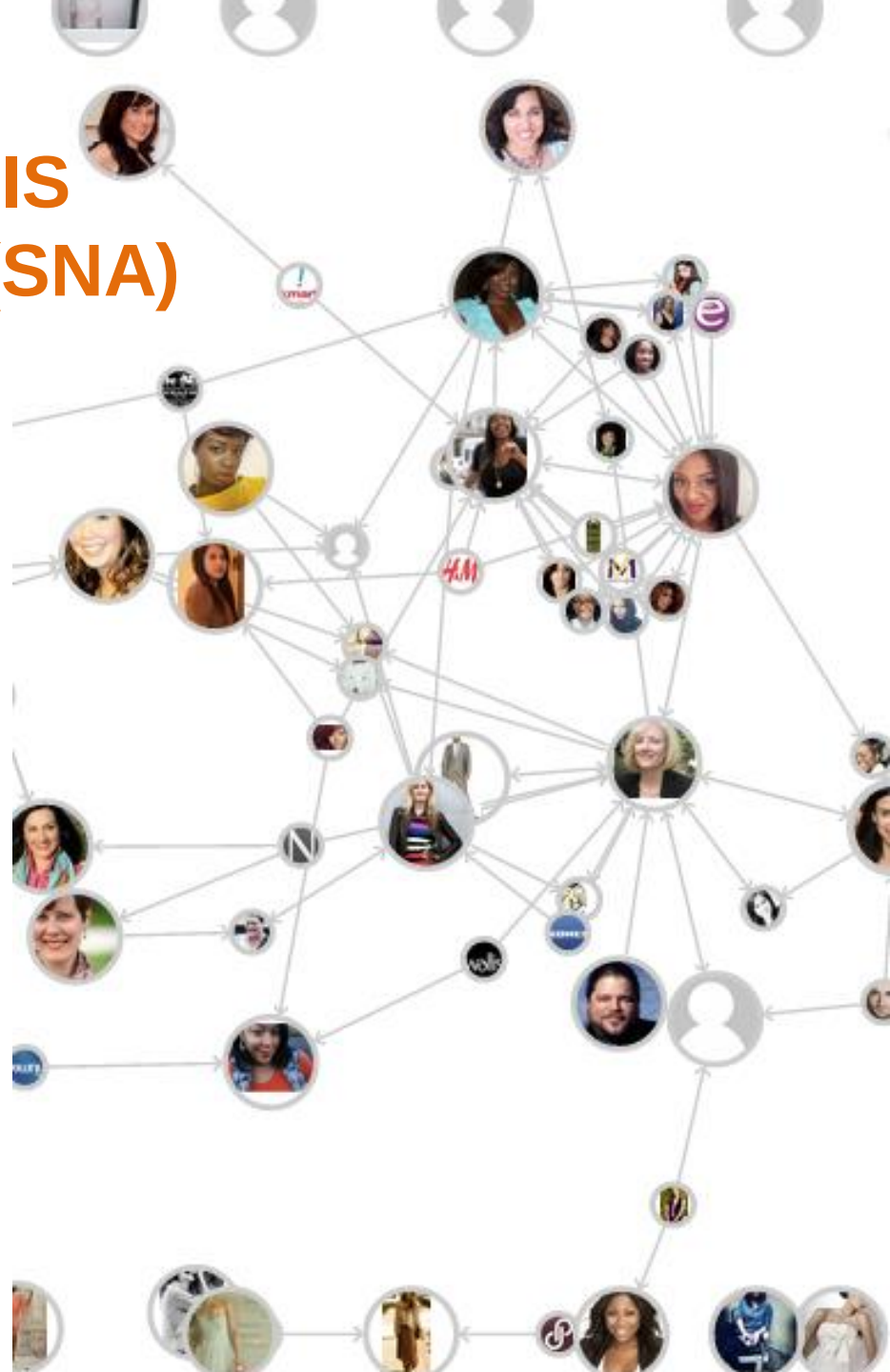


- Use Social media to create a happiness index
- Pilot project: Collected 15,000 text records for the thirty most populous cities in the US
- Parsed text was utilized to calculate happiness scores with happiness index dictionary
- Examine the relationships between the index and real world phenomena including population, crime rate, and climate

INFLUENCER ANALYSIS

Social Network Analysis (SNA)

- SNA investigates social structures through the use of network and graph theory.
- Friends network can be analyzed
- An influencer is an individual who has above-average impact on a specific niche process.
- On the social network, an influencer can be referred to the most shaping a discussion about a topic.



VOICE OF CUSTOMERS

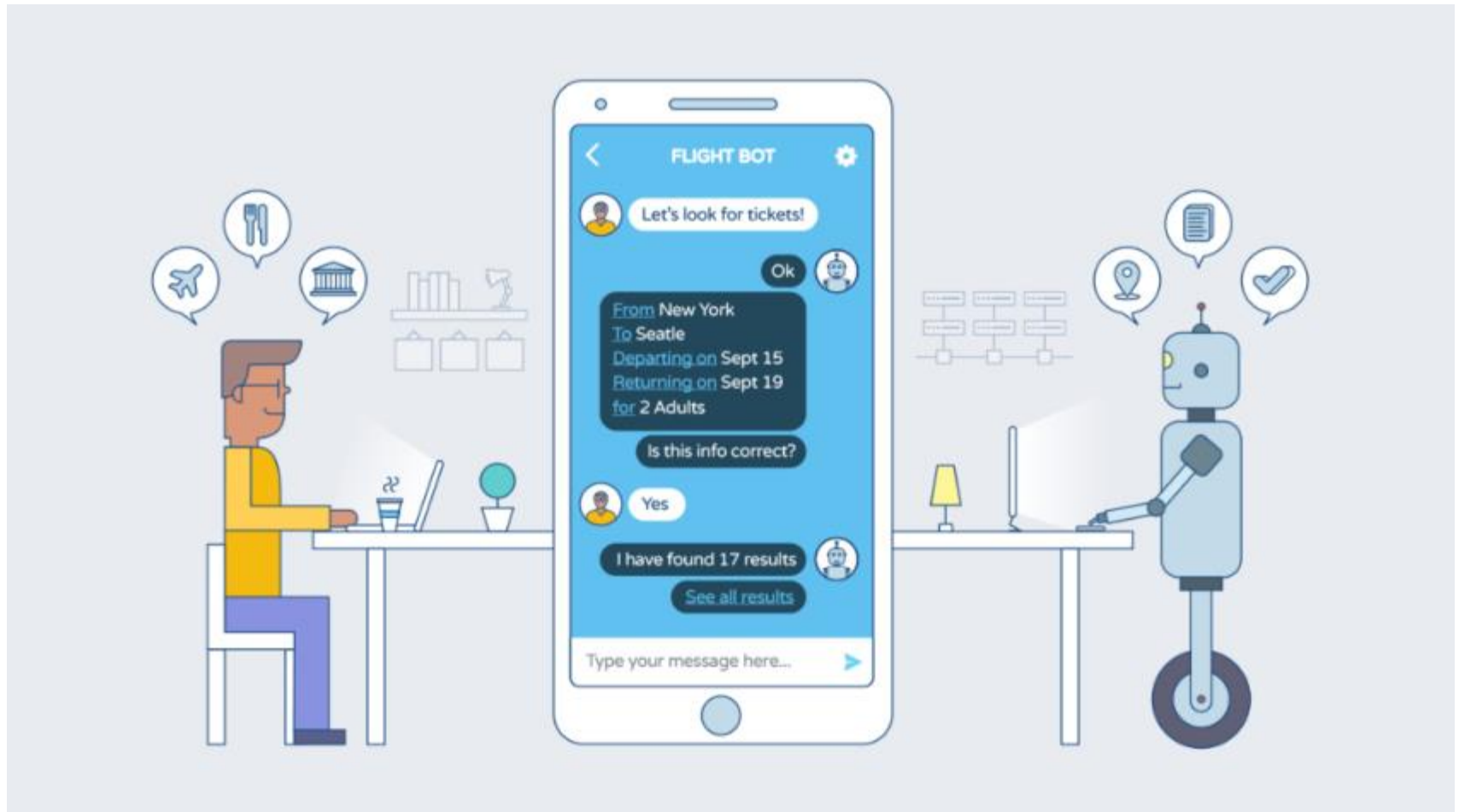
Call Center, Social Network, Front Offices



Knows feedback sentiment, Keep track of behavioral trend, Real-time anomaly detection, Knows where your target audiences are, Find out who are the influencers,

CHATBOT

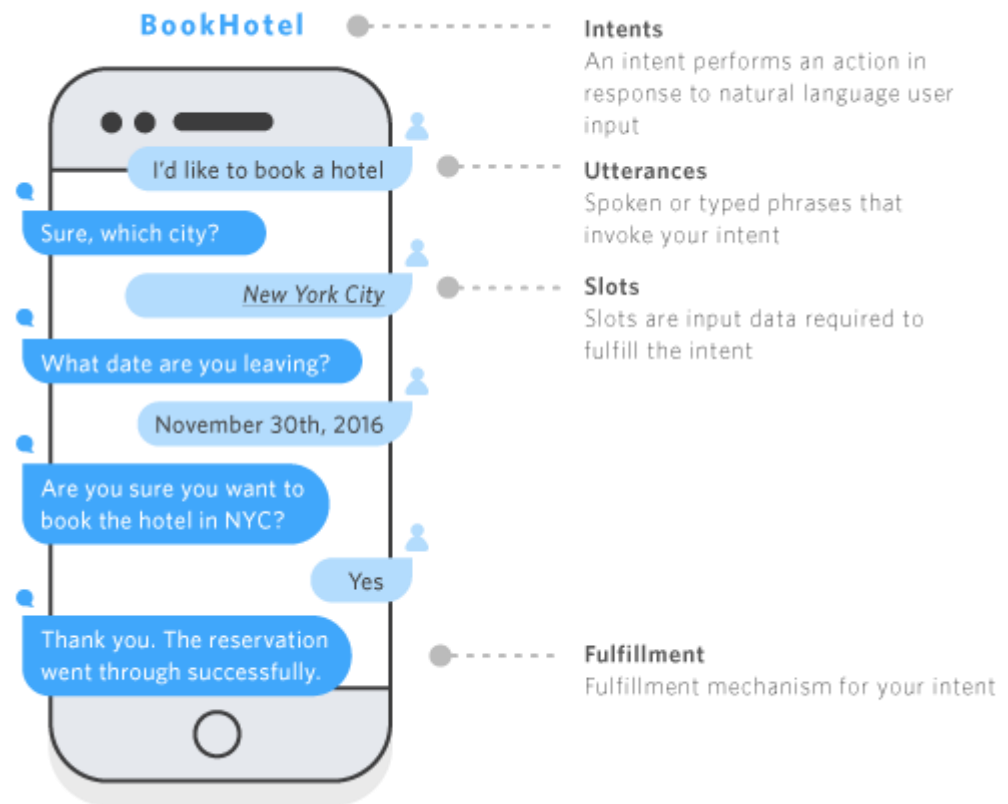
Customer Services



INTENT CLASSIFICATION

Text Classifier: associates an input sentence with a “conversational intent”

- The input “how are you?” is classified as an intent associated with a response such as “I’m good” or “I am well.”
- Three kinds of text classifier
 - Pattern matchers (Brute force or rule based)
 - Algorithms
 - Machine Learning





Whatt time does teh checkup
cilnic open on Mnoday?

Spell Correction

Word Segmentation

POS Tagging

Entity Recognition

Intent Classification



What time does the checkup
clinic open on Monday?

Spell Correction

Word Segmentation

POS Tagging

Entity Recognition

Intent Classification



What | time | does | the | checkup |
clinic | open | on | Monday?

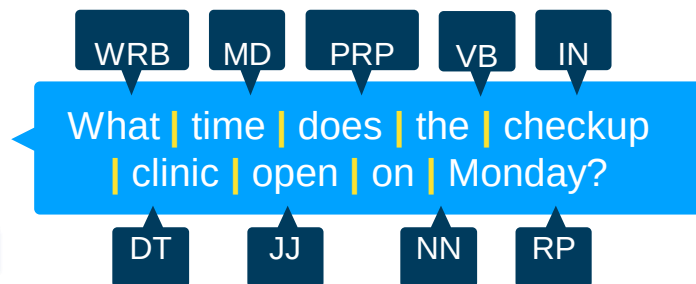
Spell Correction

Word Segmentation

POS Tagging

Entity Recognition

Intent Classification



Spell Correction

Word Segmentation

POS Tagging

Entity Recognition

Intent Classification



What | time | does | the | checkup
| clinic | open | on | Monday?

Date

Spell Correction

Word Segmentation

POS Tagging

Entity Recognition

Intent Classification



What | time | does | the | checkup
| clinic | open | on | Monday?



Working hours
and contact

Coverages

Check up

Waiting Time

Navigation

Spell Correction

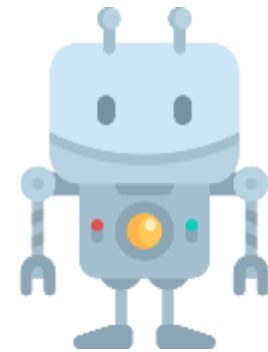
Word Segmentation

POS Tagging

Entity Recognition

Intent Classification

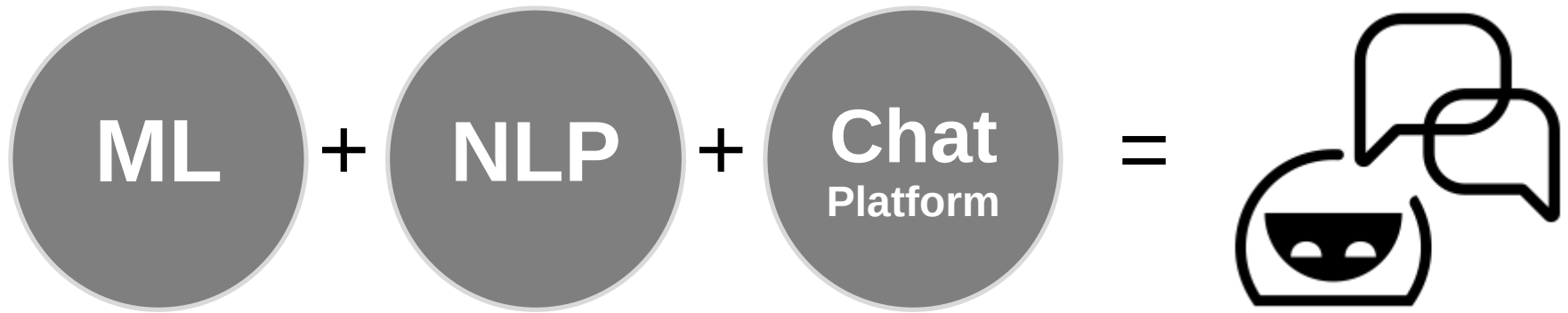
Check up clinic is open on
Monday from 07.00 to 16.00



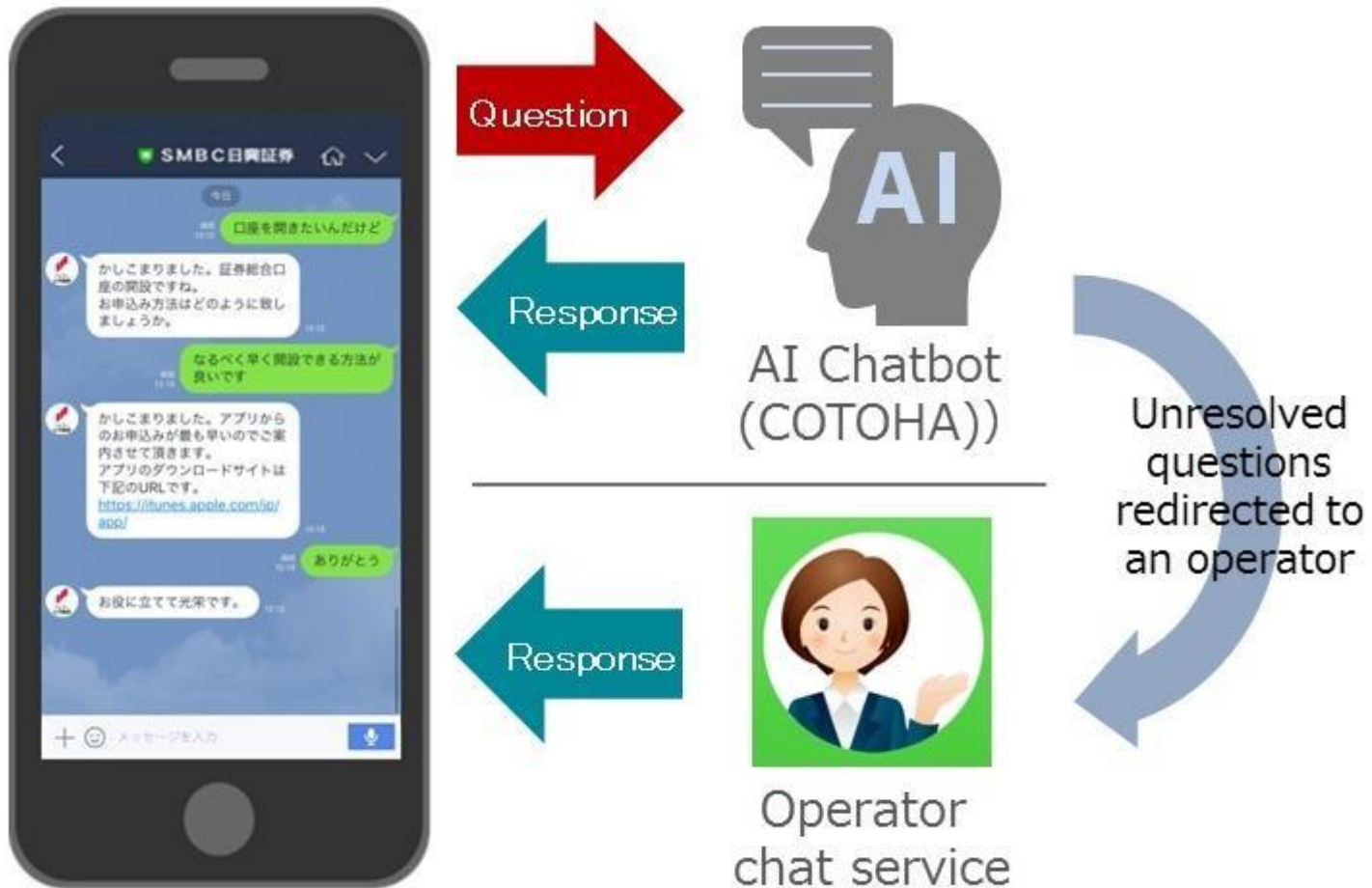
Bot Framework

Chatbots have revolutionized the customer service space

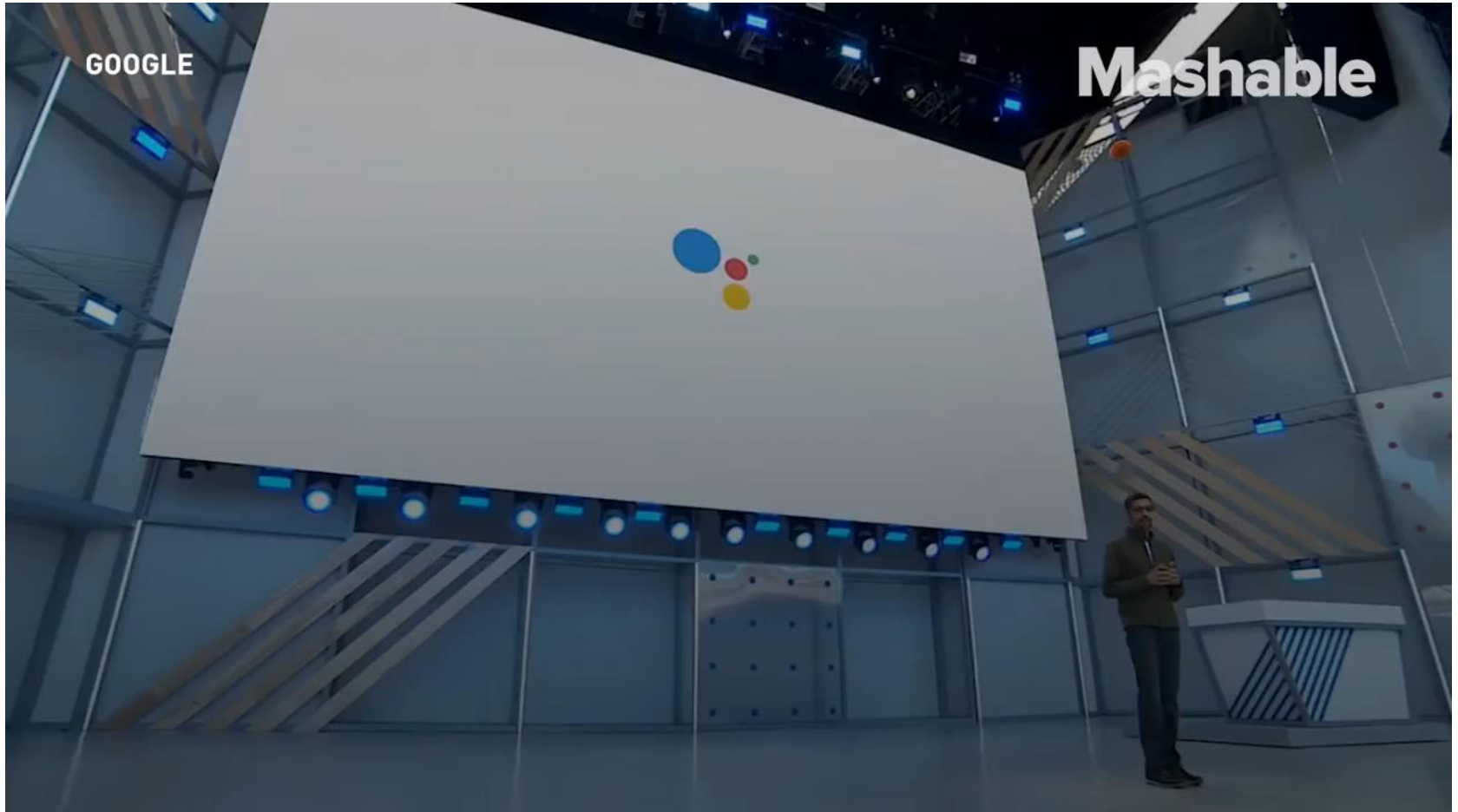
Chatbot = Conversational interface powered by AI



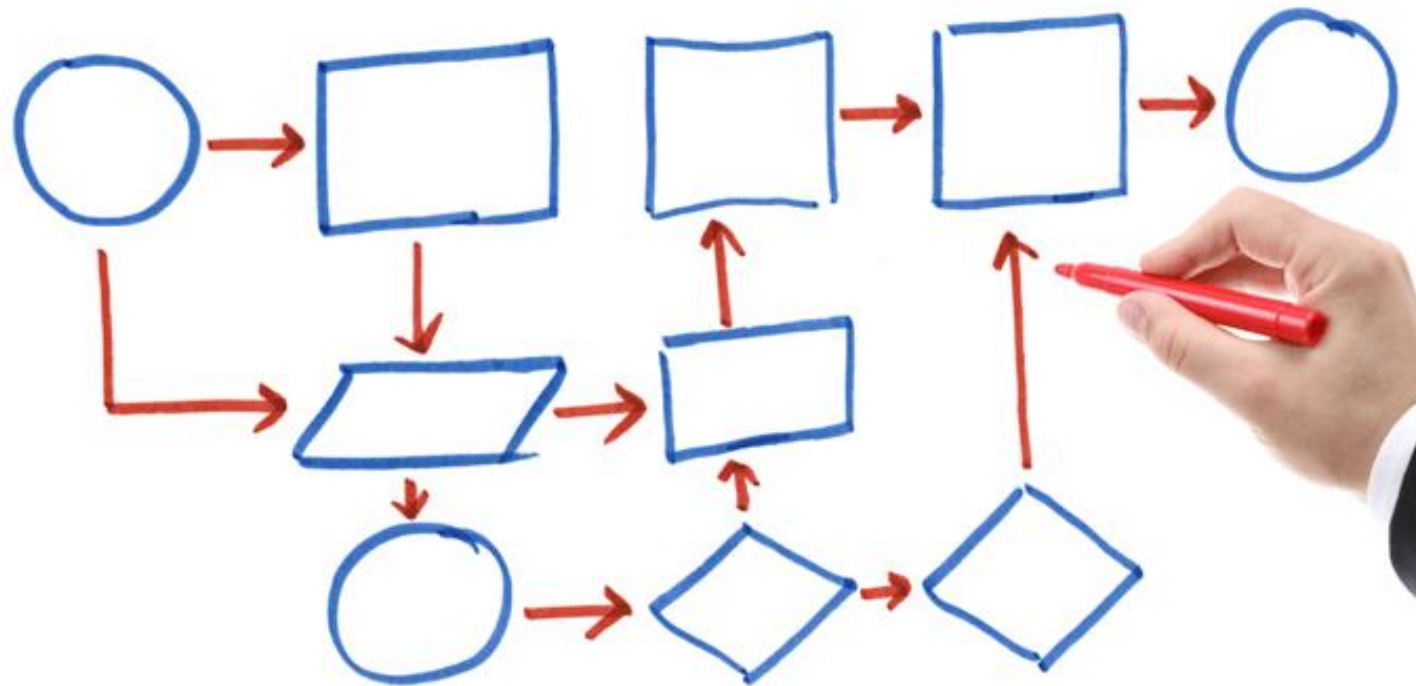
HUMAN HANDOVER



Google Assistant !



BIG DATA PROCESS



Project Inception

Data Scientists และคณะทำงาน Big Data ของแต่ละทีมงานร่วมกันกำหนดโจทย์ที่เหมาะสมและตั้งโครงการพัฒนาและทดสอบโมเดลคณิตศาสตร์นำร่องที่เหมาะสม

Problem Analysis

ทีม Data Scientists สืบหาข้อมูลที่มีอยู่ในปัจจุบันตามโจทย์นำร่องที่กำหนด เพื่อประเมินความพร้อมและแปลงจากความต้องการในเชิงปัญหาให้เป็นข้อกำหนดในเชิงข้อมูลและระบบ

Exploratory Data Analysis

สำรวจการกระจายตัวของข้อมูลเพื่อทำความเข้าใจข้อมูล และหาความสัมพันธ์ระหว่างตัวแปรในข้อมูล ในขั้นตอนนี้ทีมงานจะต้องการตัวอย่างข้อมูลจริง ระบุข้อมูลที่ต้องการเพิ่มเติมและเริ่มเตรียมข้อมูลจริง

Predictive Analytic

ทีมงานนำข้อมูลที่จัดเตรียมในเบื้องต้นมาใช้ในการสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อการทำนาย โดยใช้เทคนิคและอัลกอริทึมต่างๆ และทดสอบความแม่นยำของโมเดลคณิตศาสตร์

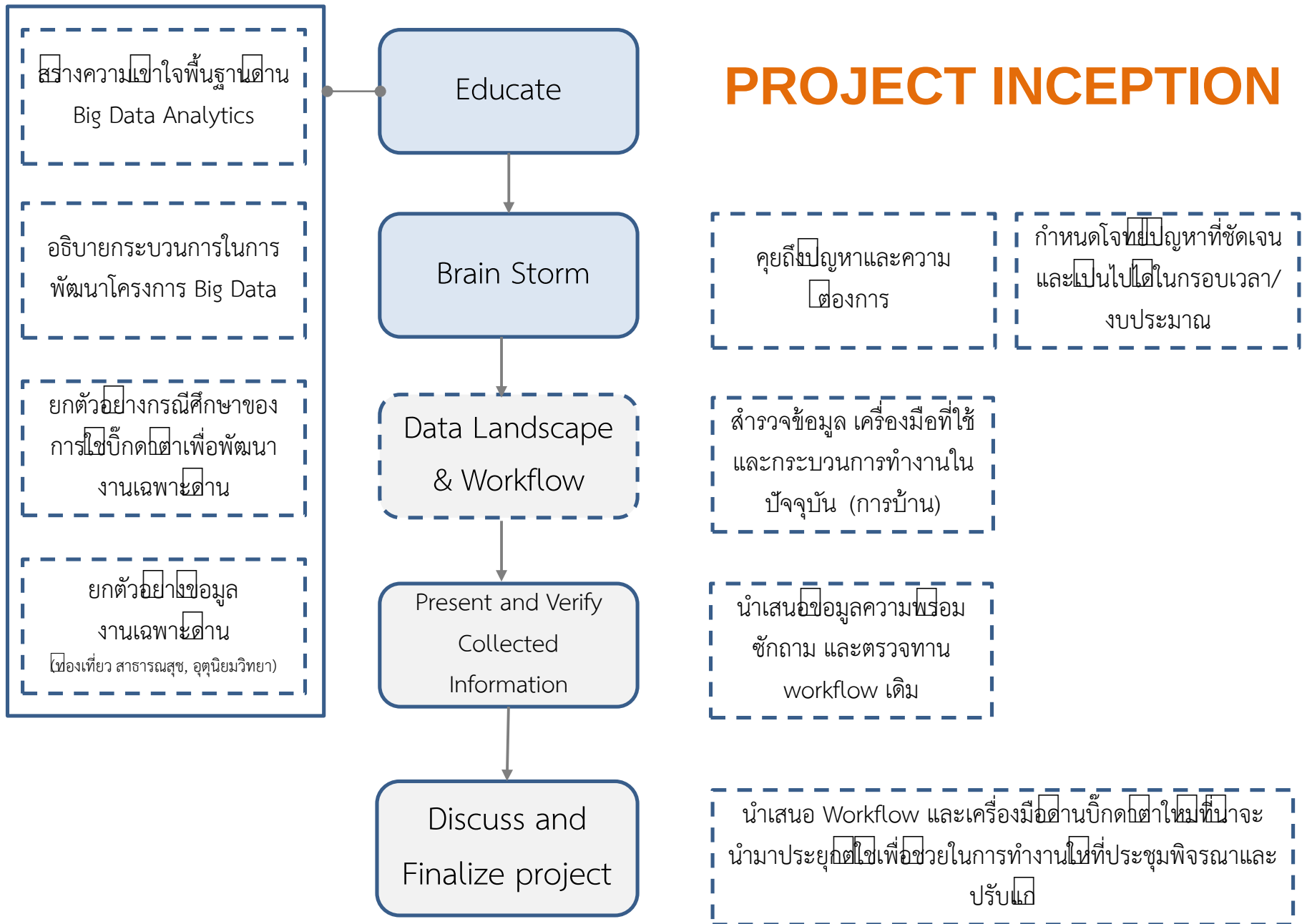
Visualization Dashboard

ออกแบบวิธีการแสดงผลโดยเลือกมิติของข้อมูลที่เหมาะสมบน Interactive Dashboard เพื่อให้นักคณะทำงานทดลองใช้และสื่อสารกับทีมผู้บริหาร และผู้ปฏิบัติ ให้สามารถนำเอาความเข้าใจดังกล่าวไปแปลงเป็นแผนการพัฒนาค่อยๆ

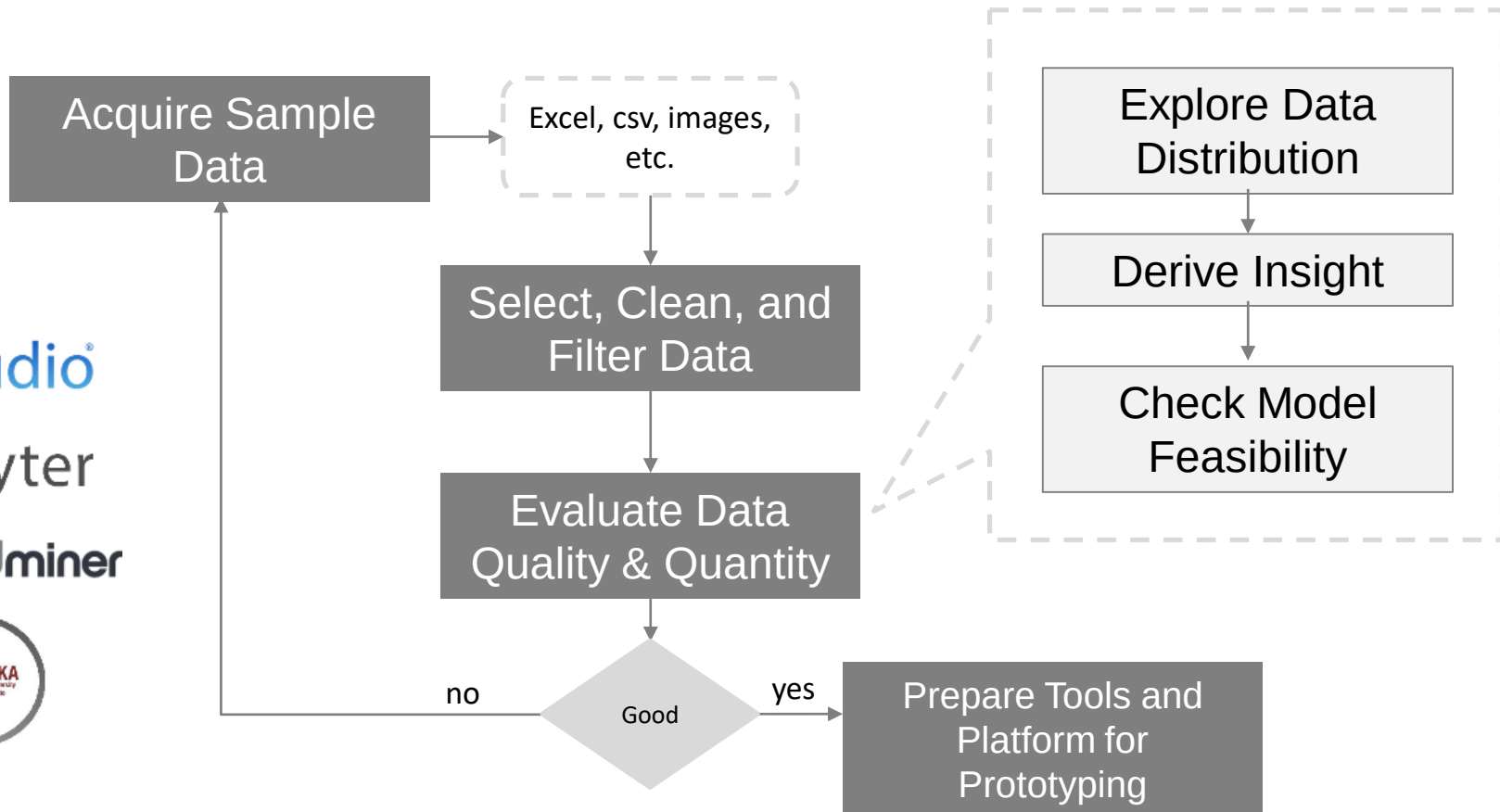
Implementation & Deployment

หลังจากผลลัพธ์เป็นที่พอใจแล้ว นักพัฒนาระบบเริ่มพัฒนาโปรแกรมตามรูปแบบของโมเดลคณิตศาสตร์ที่วางไว้ และตั้งค่าให้โปรแกรมให้ประมวลผลโมเดลแบบอัตโนมัติตามความถี่ที่วางแผนไว้ จากนั้นติดตั้งระบบซอฟต์แวร์เพื่อการใช้งานจริง

PROJECT INCEPTION



Exploratory Data Analysis



PREDICTIVE ANALYTICS



Data Mining

The Computational process of discovering patterns in large data sets involving methods at the intersection of statistics, machine learning, and database systems.



Text Analytics

The process of deriving high-quality information from **text**. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning.



Machine Learning / Deep Learning

The science of getting computers to learn from data without having to be explicitly programmed by humans. Machine model can teach themselves to grow and change when exposed to new data.



Big Data Technology

Technology designed to manage and process extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

EXAMPLE TASKS

- **Regression**

produce a model that, given an individual, estimates the value of the particular variable specific to that individual.

- ✓ How much will a given housing unit cost ?

- **Clustering**

group individuals in a population by their similarity (not driven by any specific purpose).

- ✓ Do our customers form natural groups or segments?

- **Co-occurrence grouping**

find associations between entities based on transactions involving them.

- ✓ What items are commonly purchased together?

- **Profiling**

characterize the typical behavior of an individual, group, or population.

- ✓ What is the typical service usage of this customer segment ?

- ✓ Used to establish behavior norms for anomaly detection (fraud detection)

CREDIT SCORING



- First Union Bank deployed a value predicting system that assigns green / yellow / red flag to each customer, based on their predicted lifetime value.
- Service representatives were instructed to waive fee for green customers, and not waive for red customers. For yellow customers, they can make their own judgement.
- This strategy generated over \$100 million in incremental revenue.



ANSWERING BUSINESS QUESTIONS

- Who are the most profitable customers?
 - A straightforward database query, if “profitable” can be defined clearly.
- Is there really a difference between the profitable customers and the average customer?
 - Statistical Hypothesis testing
- But who really are these customers? Can I characterize them?
 - Automated pattern finding
- Will some new customer be profitable ? How much revenue can I expect?
 - Predictive model of profitability

Project Inception

Data Scientists และคณะทำงาน Big Data ของแต่ละทีมงานร่วมกันกำหนดโจทย์ที่เหมาะสมและตั้งโครงการพัฒนาและทดสอบโมเดลคณิตศาสตร์นำร่องที่เหมาะสม

Problem Analysis

ทีม Data Scientists สืบหาข้อมูลที่มีอยู่ในปัจจุบันตามโจทย์นำร่องที่กำหนด เพื่อประเมินความพร้อมและแปลงจากความต้องการในเชิงปัญหาให้เป็นข้อกำหนดในเชิงข้อมูลและระบบ

Exploratory Data Analysis

สำรวจการกระจายตัวของข้อมูลเพื่อทำความเข้าใจข้อมูล และหาความสัมพันธ์ระหว่างตัวแปรในข้อมูล ในขั้นตอนนี้ทีมงานจะต้องการตัวอย่างข้อมูลจริง ระบุข้อมูลที่ต้องการเพิ่มเติมและเริ่มเตรียมข้อมูลจริง

Predictive Analytic

ทีมงานนำข้อมูลที่จัดเตรียมในเบื้องต้นมาใช้ในการสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อการทำนาย โดยใช้เทคนิคและอัลกอริทึมต่างๆ และทดสอบความแม่นยำของโมเดลคณิตศาสตร์

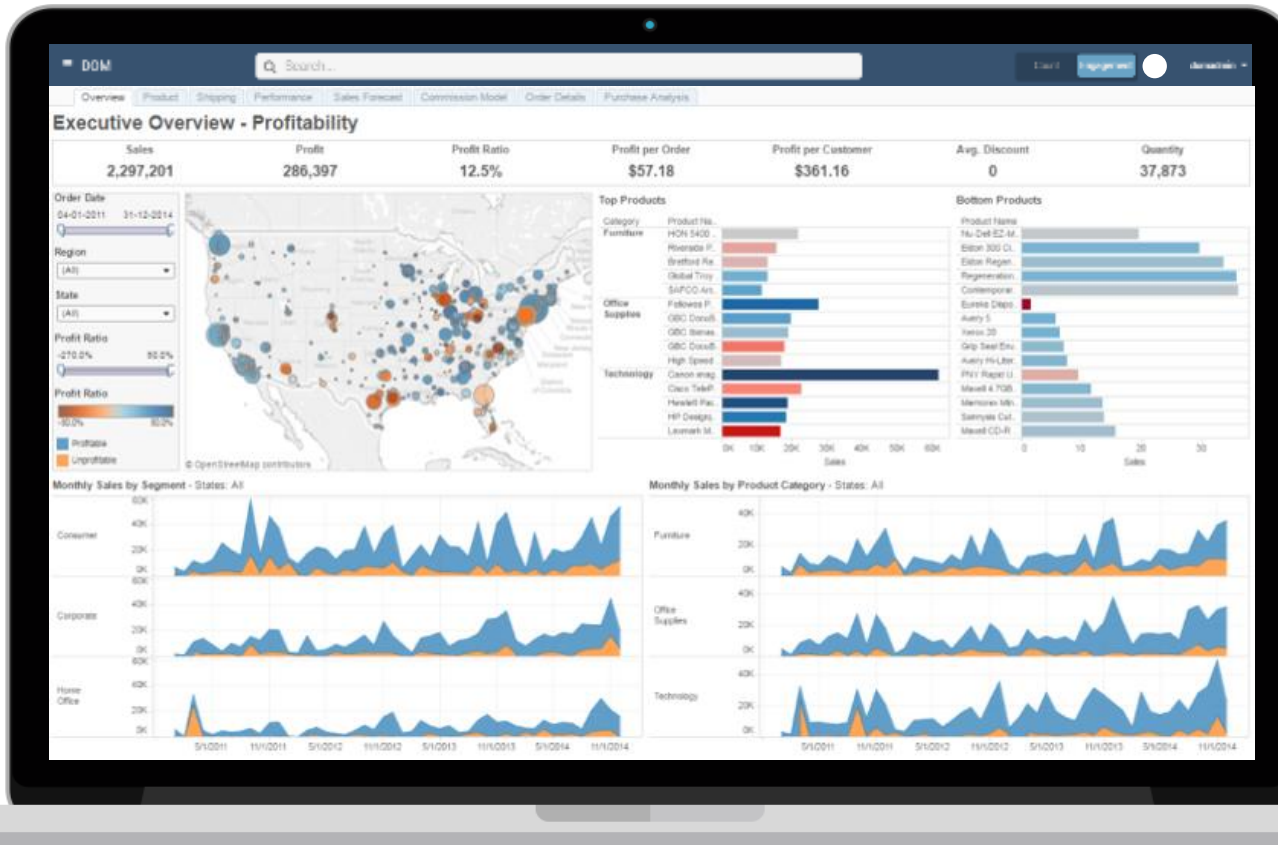
Visualization Dashboard

ออกแบบวิธีการแสดงผลโดยเลือกมิติของข้อมูลที่เหมาะสมบน Interactive Dashboard เพื่อให้นักคณะทำงานทดลองใช้และสื่อสารกับทีมผู้บริหาร และผู้ปฏิบัติ ให้สามารถนำเอาความเข้าใจดังกล่าวไปแปลงเป็นแผนการพัฒนาค่อยๆ

Implementation & Deployment

หลังจากผลลัพธ์เป็นที่พอใจแล้ว นักพัฒนาระบบเริ่มพัฒนาโปรแกรมตามรูปแบบของโมเดลคณิตศาสตร์ที่วางไว้ และตั้งค่าให้โปรแกรมให้ประมวลผลโมเดลแบบอัตโนมัติตามความถี่ที่วางแผนไว้ จากนั้นติดตั้งระบบซอฟต์แวร์เพื่อการใช้งานจริง

Visualization Dashboard



etc.

Implementation and Deployment

Data Input

Messaging,
and Web Services



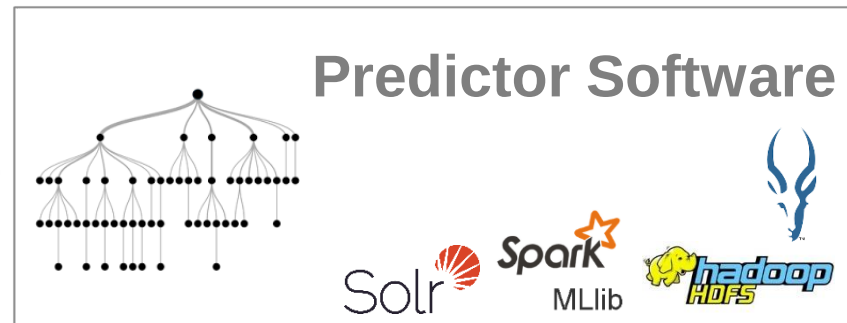
EDW, OLAP



Social Media, Weblogs



Machine Devices, Sensors



Visualization



THE ANALYTIC CAPABILITY

Big decisions need better analytics

Data science provides fact-based, math-based decision support (*Data-Driven Decision*)



Descriptive Analytics

tell you what happened;
how many; how often;
where critical events occur.



Predictive Analytics

project what will happen
next; provide indications
of possible outcomes if
trends continue.



Prescriptive Analytics

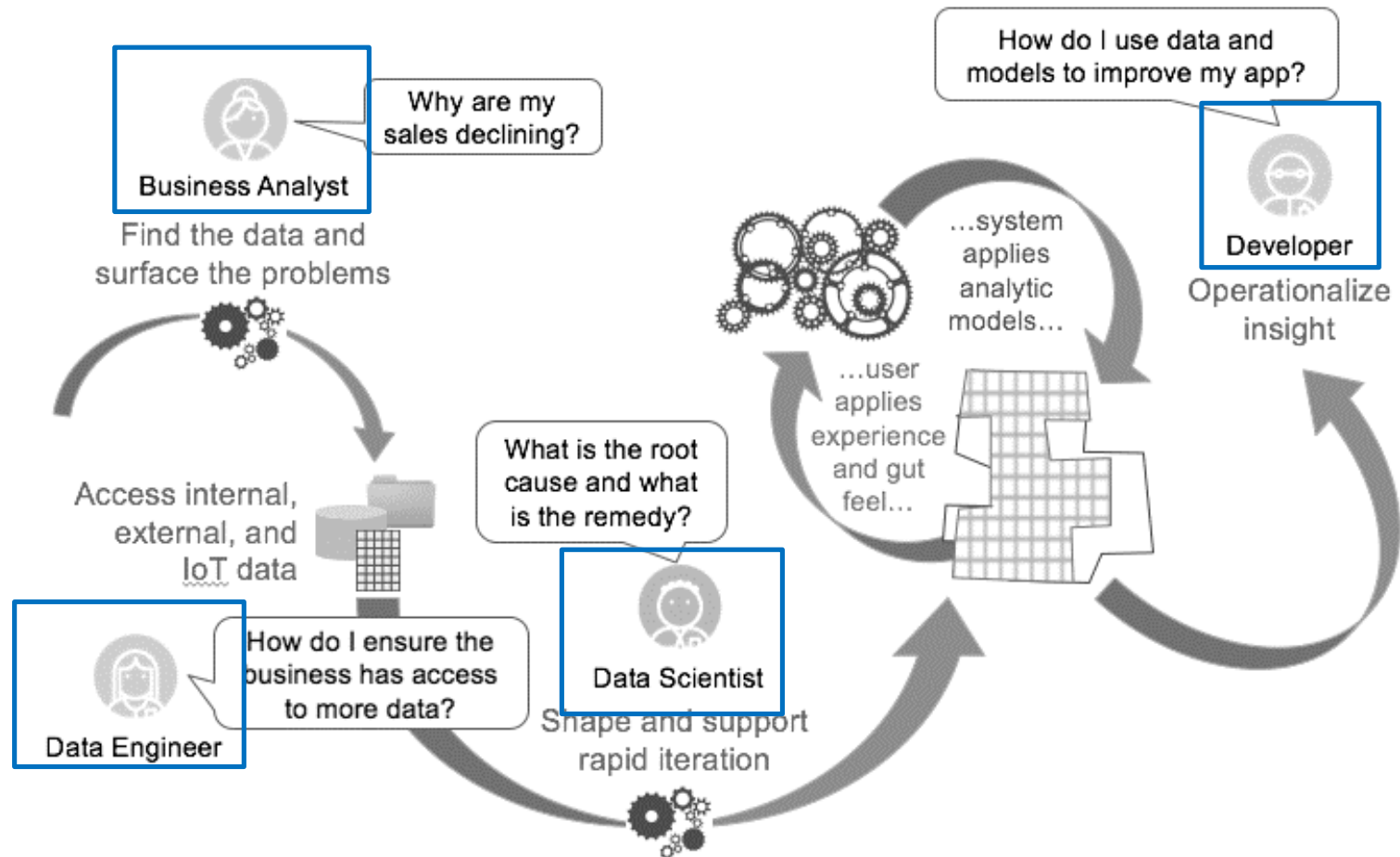
Synthesizes big data to
make predictions and then
suggests decision options to
take advantage of the
predictions

1

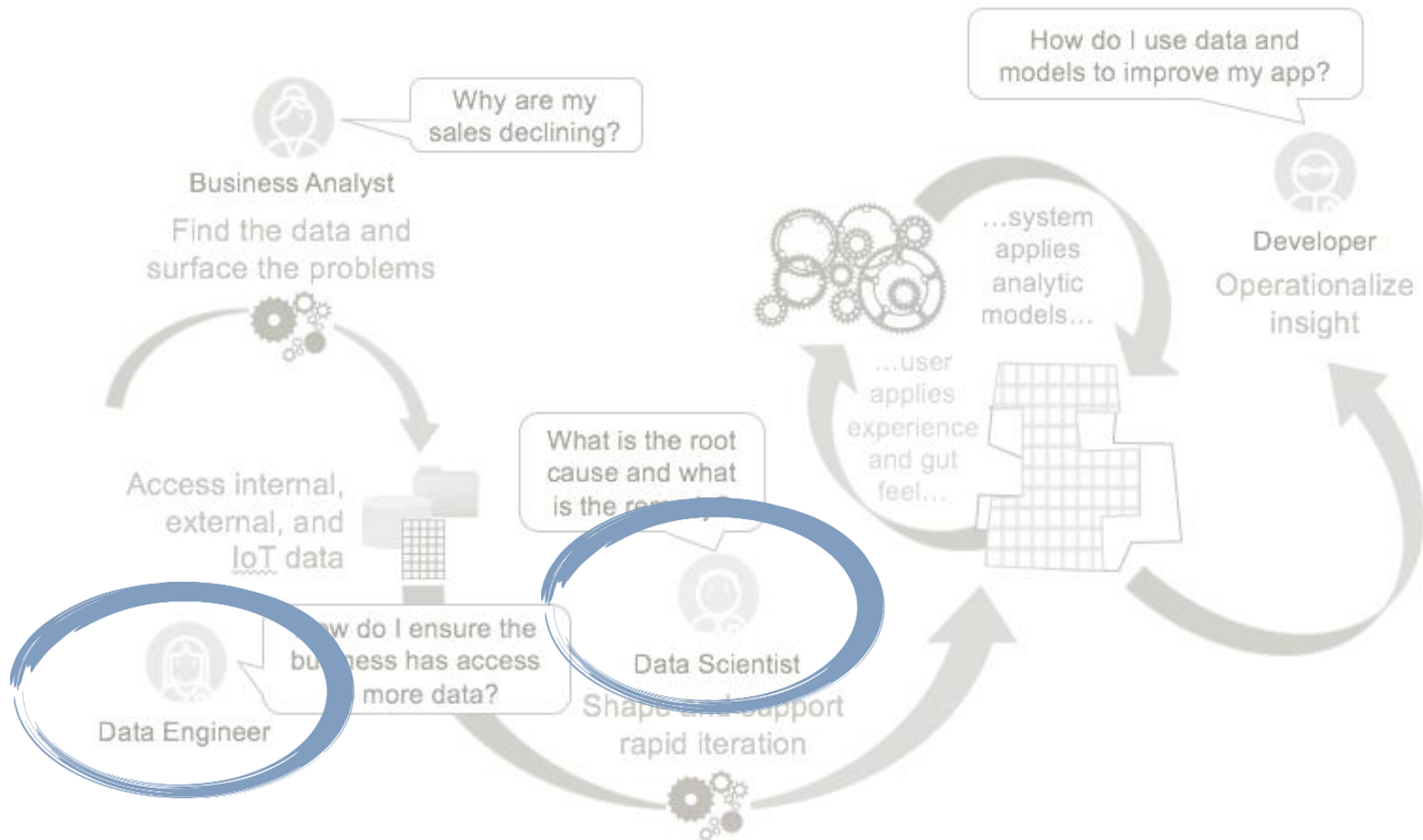
2

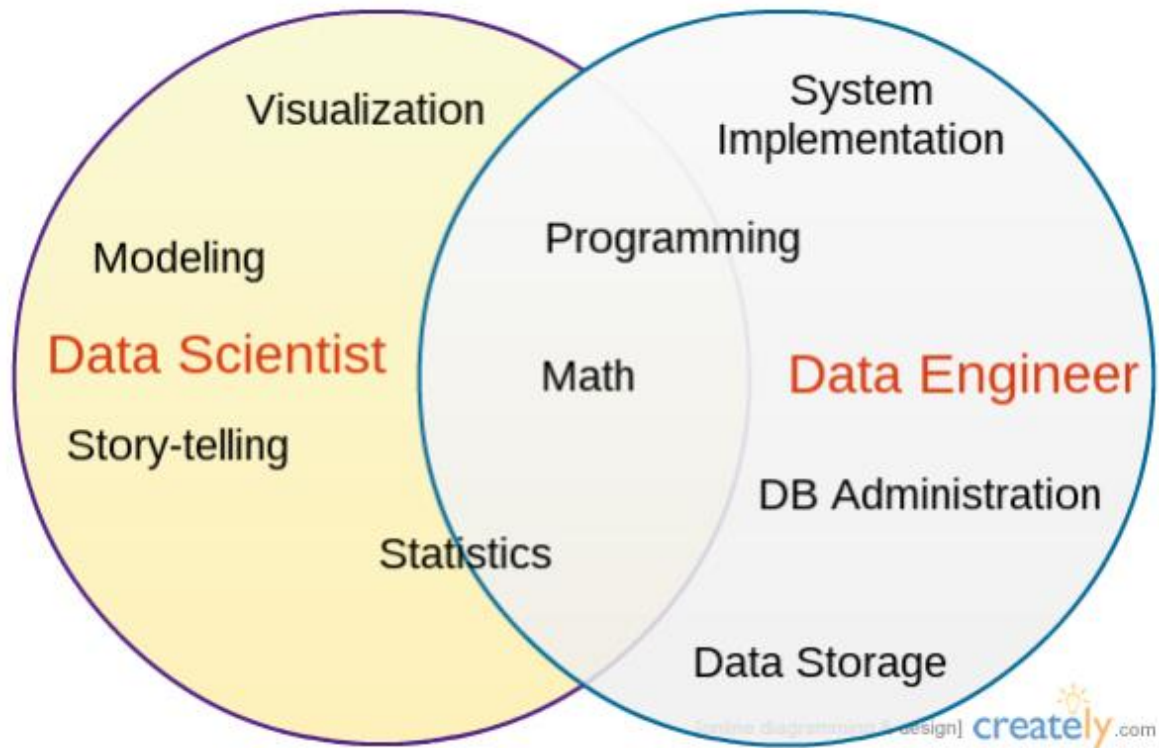
3

The 4 Personas for Data Analytics



The 4 Personas for Data Analytics





Statistics
Machine learning
Information retrieval
Signal processing
Data visualization
Databases

Big data platform and tools
Data modeling and ETL tools
Data warehousing solutions
Data APIs
Clouds
High performance computing



Some examples of data-driven projects

BIG DATA EXAMPLES

WITH BIG DATA, WE CAN

- Gauge sentiments of the general public towards services
- Make it easier to maintain vehicle fleet
- Find more efficient routes to deliver
- Quicken the last mile shipping
- Make sure drivers are complying with traffic law



The DHL Story

POSSIBLE DATA SOURCES

The DHL Story

- Traditional enterprise data from operational systems
- Traffic & weather data from sensors, monitors and forecast systems
- Vehicle diagnostics, driving patterns, and location information
- Financial business forecasts
- Advertising response data
- Website browsing pattern data
- Social media data



PROCESS OPTIMIZATION

- **Goal:** Streamline operational processes: mail and parcel processing and delivery
- **Predictive analytics:** predict parcel volumes and better determine delivery staff and vehicle requirements through analysis of correlations between external factors such as weather, flu epidemics, Google trends, and shipment data.
- **Predictive maintenance:** Equip new electric vehicles with sensors that track vehicles' vital statistics.

TRANSPORTATION

Traffic Prediction
Personalized Insurance
Connected Car

AFC Tap-in & tap-out



Event ticketing



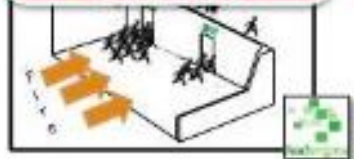
Crowd Counting



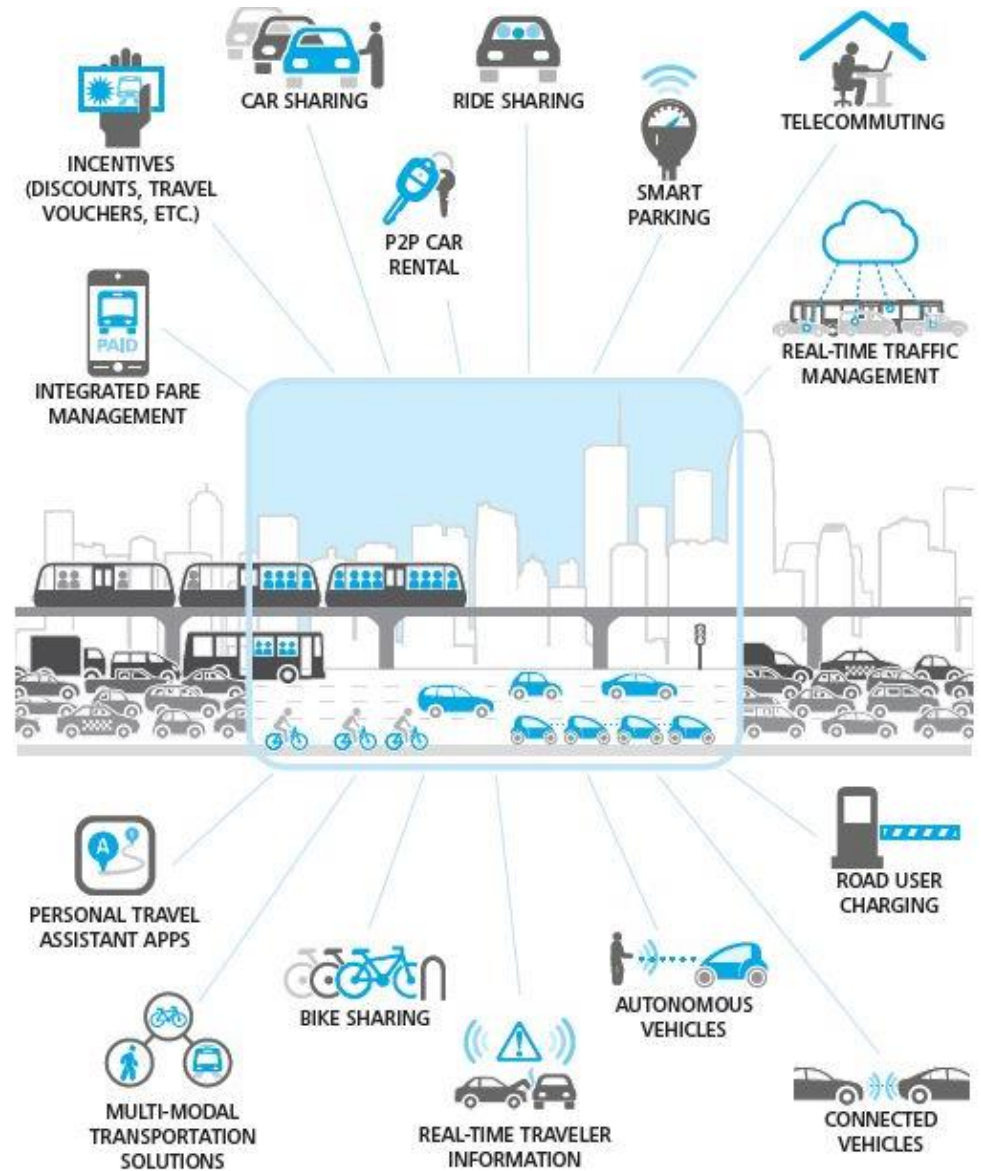
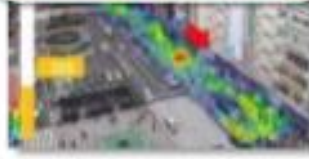
Transport App



Sensor-based Crowd Detection



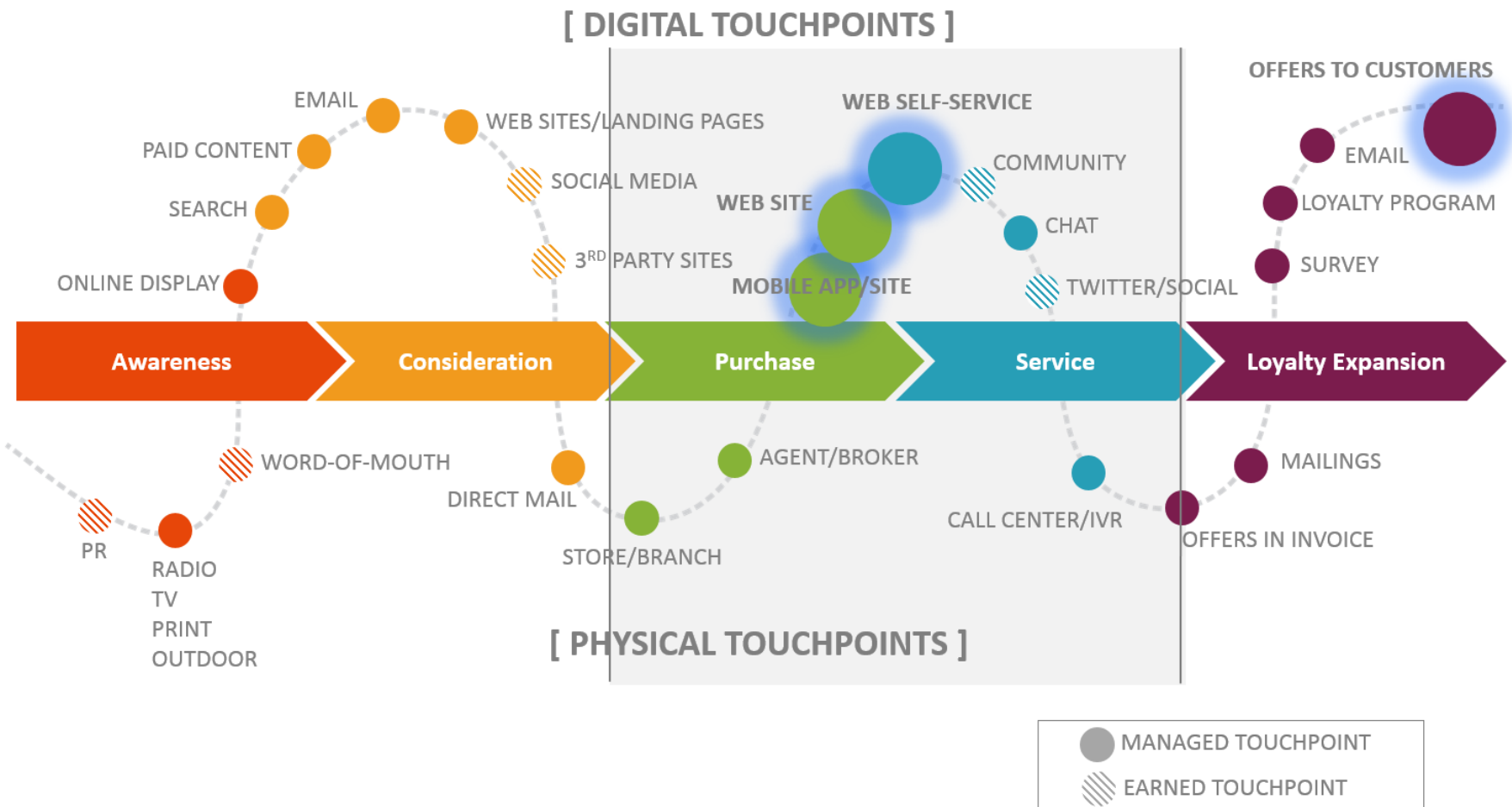
Video-based Crowd Behavior Analysis



Vehicle Allocation Optimization

- Goal: Maximize availability and minimize transition
- Predict quantity requirements by locations
- Factors:
 - Previous month vehicle locations
 - Previous months requirements of vehicle
 - Location constraints
 - Age of vehicle
 - Scheduled maintenance
- Is it better to keep vehicles in central storage or spread among operating locations
- 30 % increase in tasks being met and 40% decrease in transitions

CUSTOMER TOUCHPOINTS



BIG DATA AND MARKETING



Targeted market



Campaign combinations
for effective up-selling



Recommendations
for cross-selling



Predicting
customer
churn

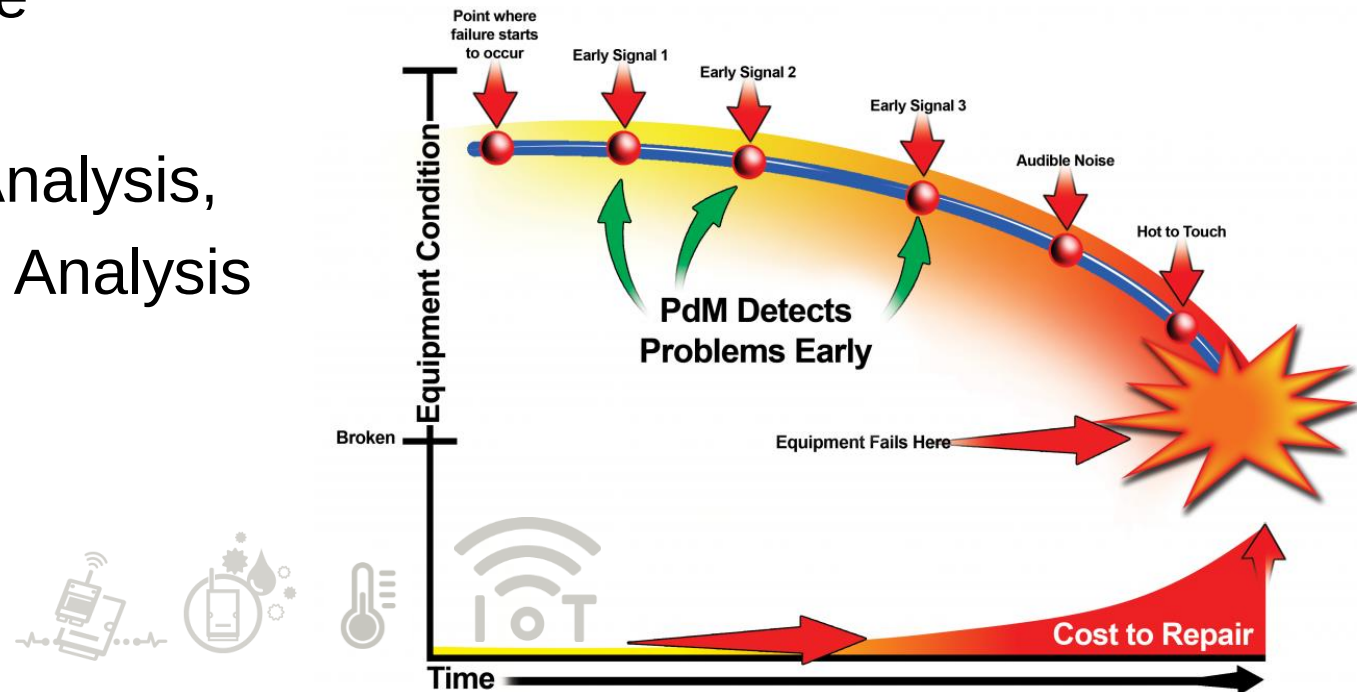


Customer behavior analysis: the key to marketing in this digital era is contacting customers just

- When they wish to be reached
- When they are in the right location
- Then, engaging them with personalized real-time offers.

PREDICTIVE MAINTENANCE (PdM)

- Determine the condition of in-service equipment in order to predict when the maintenance should be performed
- Cost savings over routine or time-based preventive maintenance
- Fault Tree Analysis,
- Time Series Analysis



PEOPLE ANALYTIC

- Predictive analytics is used for talent acquisition, retention, placement, promotion, compensation, or workforce and succession planning.
- Analyzing the skills and attributes of high performers in the present, then build a template with quality hiring factors for future hires.
- Non-traditional data gathering sources
 - Social media channels where prospective candidates usually leave their digital *'thought prints'*.
- Statistical analysis of productivity and turnover
 - The data showed that old indicators (such as GPA and education) were far less critical to performance and retention. Factors like experience is much more important.

Bersin by Deloitte

Talent Analytics Maturity Model®

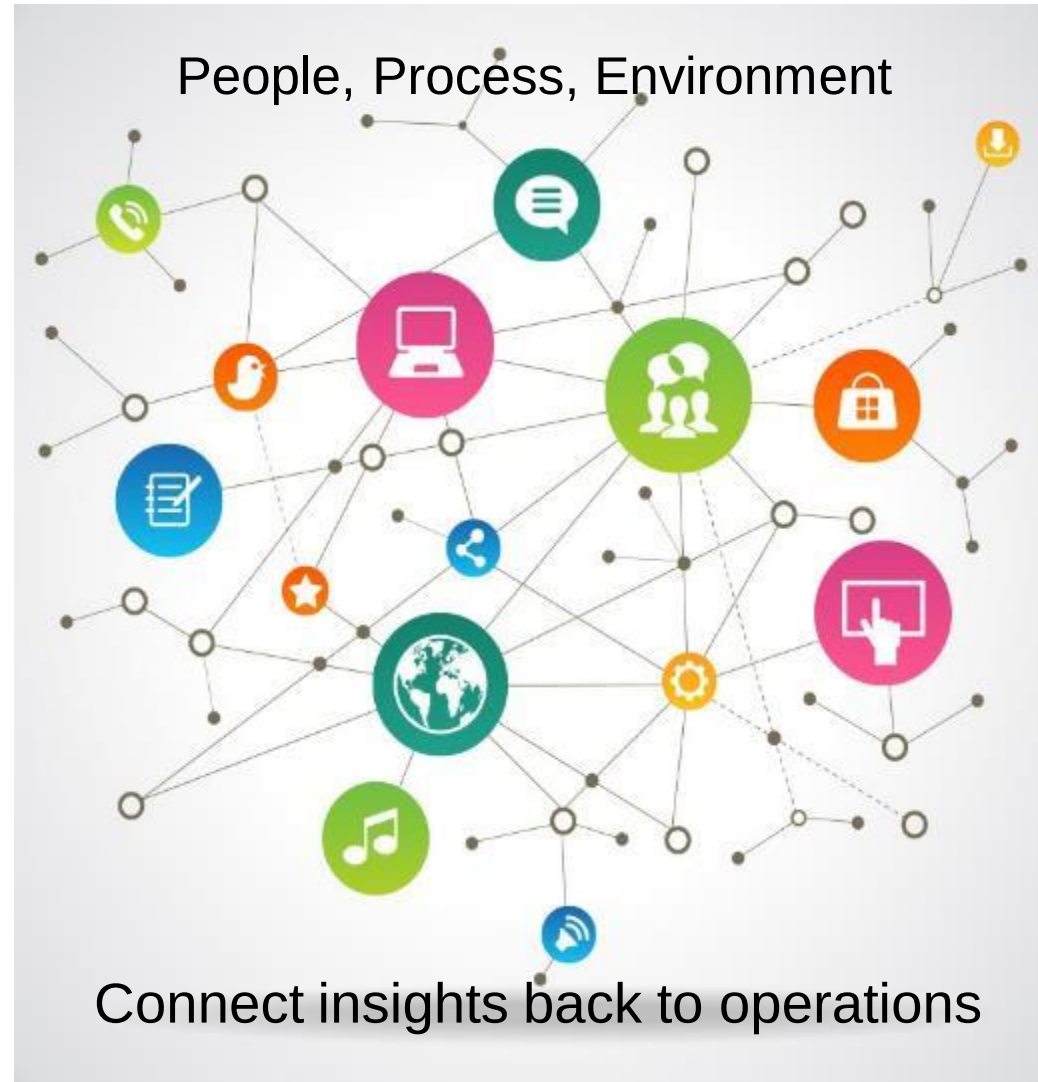


Ref: Forbe





Analytics is the process of **capturing**, **interpreting** and **communicating** useful information for better decision making





GET READY FOR THE WORLD OF DATA

THE OBSTACLES

- The absence of data and data gathering tools
- Existing data quality (consistency, accuracy, completeness, conformity)
- Lack of concept understanding in business problem formulation
- Lack of data scientists / analysts
- Data sharing within and across organizations
- Maintainability after initiatives

DATA, HOW TO OBTAIN MORE

- Investment in mobile, AI, and conversational platforms
- Investment in the use of IoTs
- Create a seamless customer cross-channel experience
- Explore social media sentiment and other unstructured data
- Integrate data silos (Administration, CRM, billing, compliances, etc.)



BIG DATA SANDBOX

- Sandbox allows an organization to realize its actual investment value in big data.
- It is a developmental platform used to explore an organization's information sets.
 - Encourage collaboration and interaction
 - Encourage learning by doing
 - Build a data pool
- Build or rent one ?



HR PREPERATION

- HR should identify current and future human resources needs for an organization to become a data-driven one.
- HR should come up with a plan to create
 - A new salary model for data scientist hiring
 - An incentive program and new opportunities for existing staffs
 - A multi-disciplinary task force



THE ROADMAP

