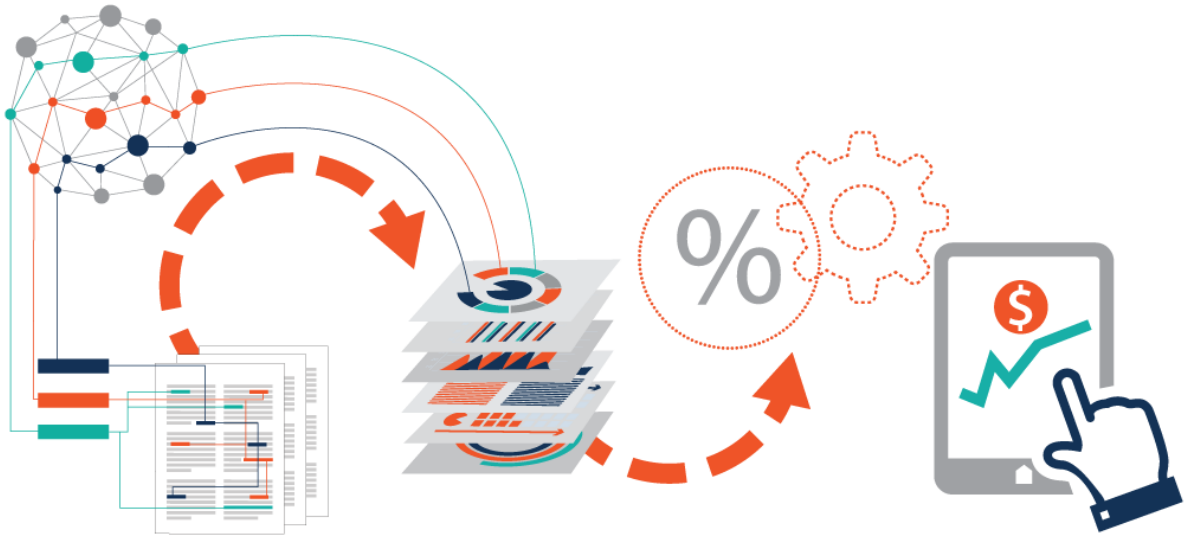
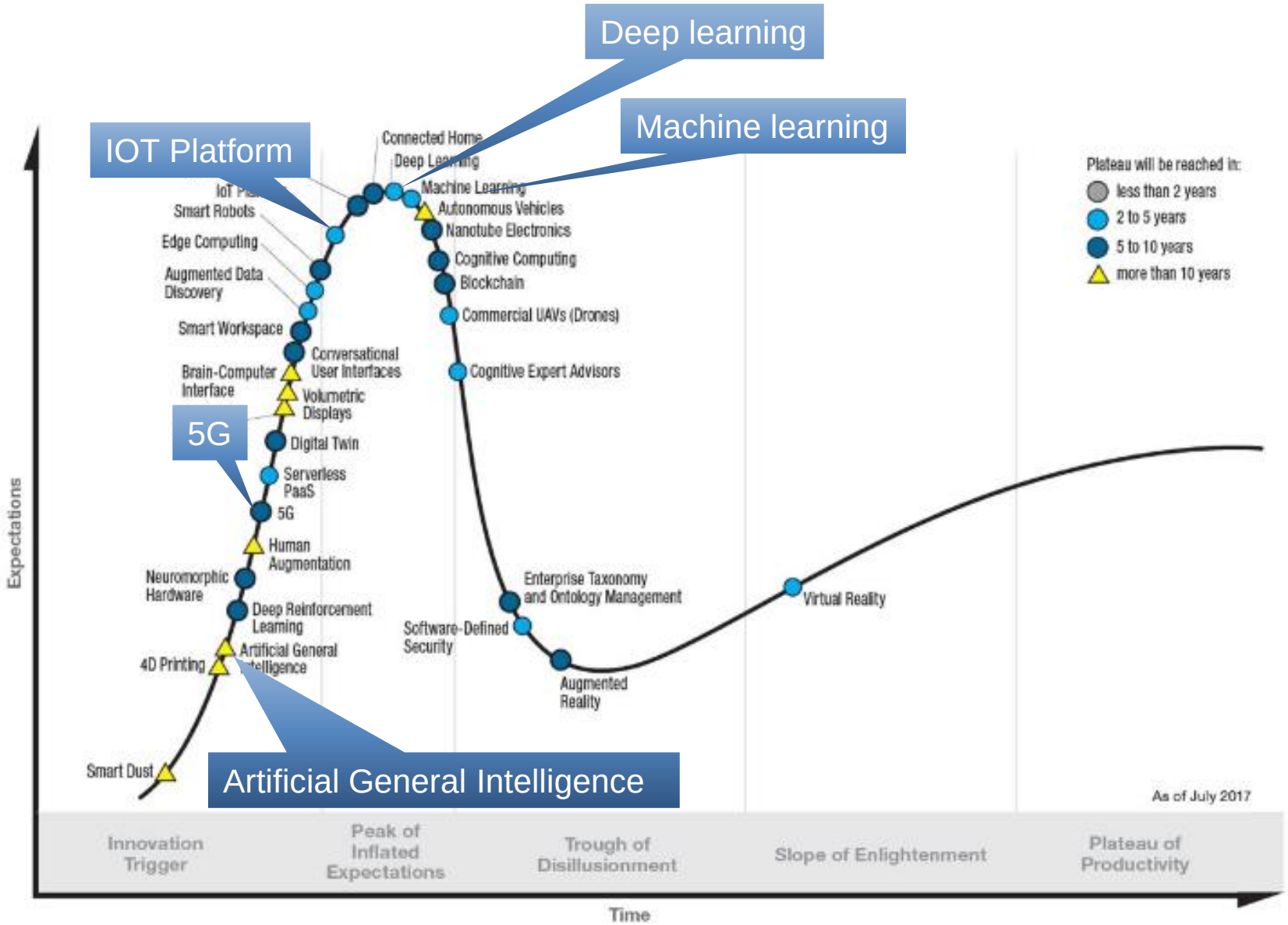
Big Data Analytics Overview

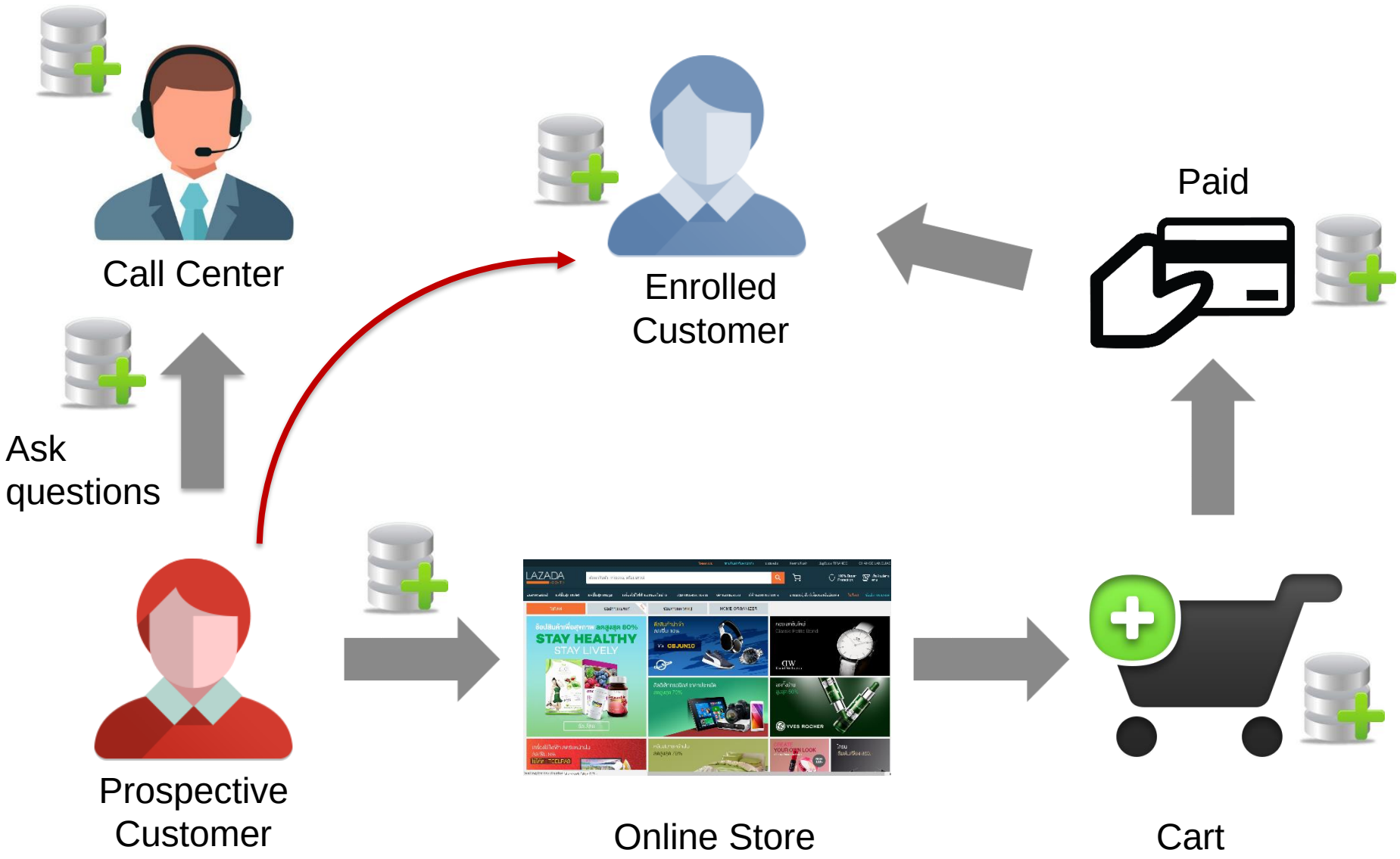
By Assoc. Prof. Dr. Tiranee Achalakul
Ministry of Digital Economy and Society



Gartner's hype cycle for Emerging Technologies 2017



What has changed to make digital tech so useful today?


Call Center
Enrolled Customer
Paid
Ask questions
Prospective Customer
LAZADA
STAY HEALTHY
STAY LIVELY
Online Store
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

CUSTOMER 360

Credit: www.t-systems.com

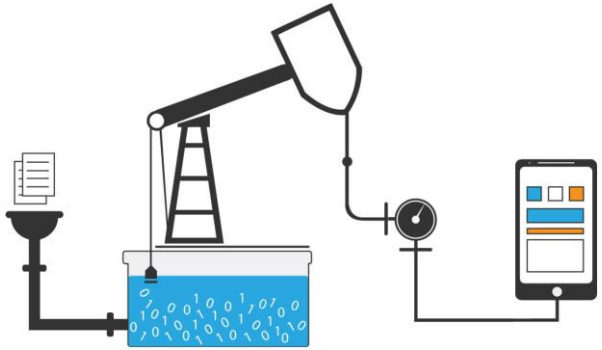
E-Wallet

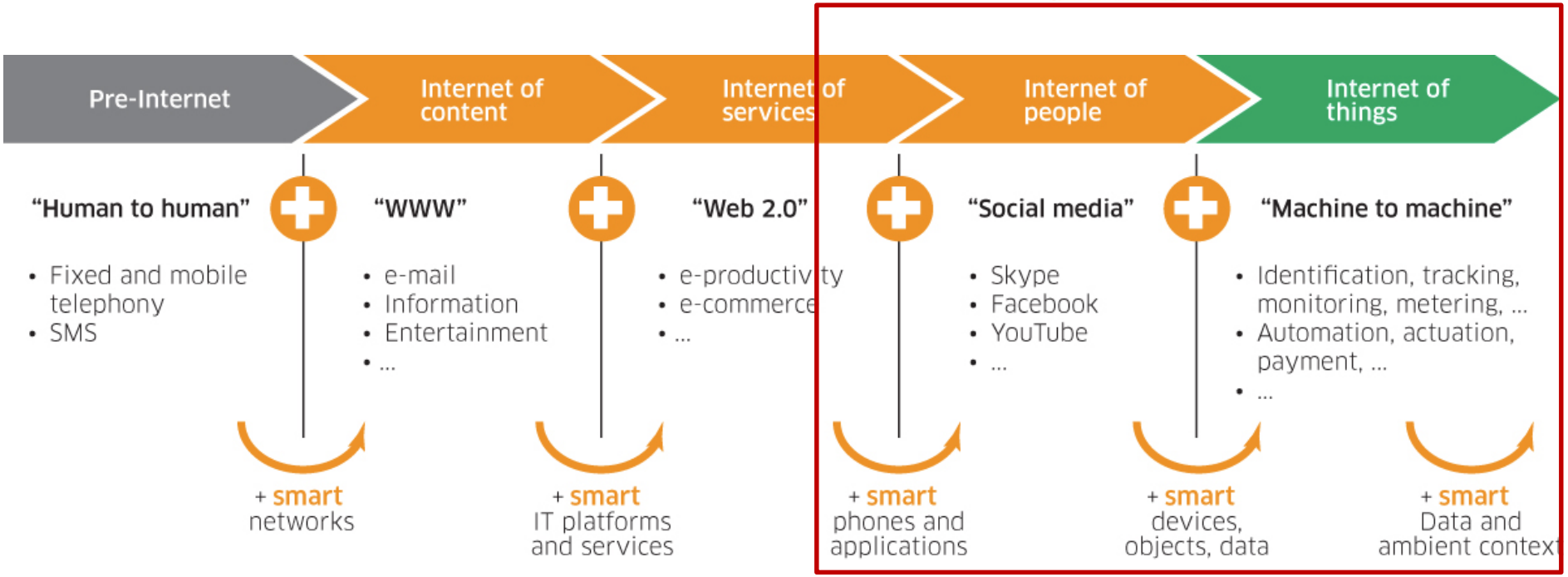
Pokémon
GO
Lv 9
Lynn96

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

Tourists information, spending patterns, insurance claims

Digital data to the government archived paperwork

Social media (needs, trends, preferences, and opinions)

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

WHERE TO LOOK FOR DATA

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

Media
Images, video, audio etc.

Data Storages
RDBMS, NoSQL, Hadoop, file systems etc.

Docs
XLS, PDF, CSV, HTML, JSON etc.

Social Networks
Twitter, Facebook, Google+, LinkedIn etc.

Machine Log Data
Application logs, event logs, server data, CDRs, clickstream data etc.

Business Apps
CRM, ERP systems, HR, project management etc.

Public Web
Wikipedia, news, weather, public finance etc

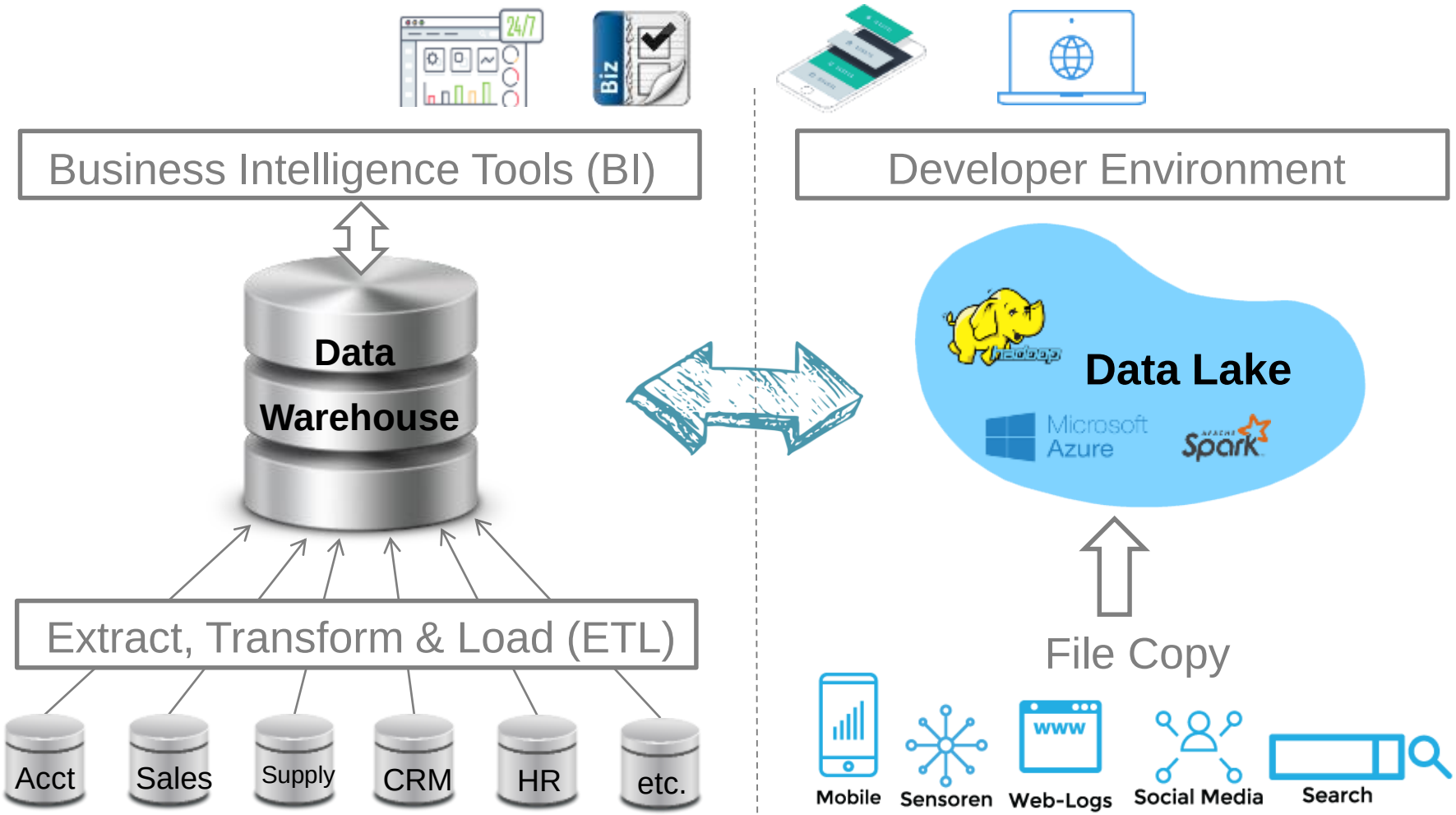
Sensor Data
Smart electric meters, medical devices, car sensors, road cameras etc.

Descriptive Analytics

What happened; How many;
how often; where

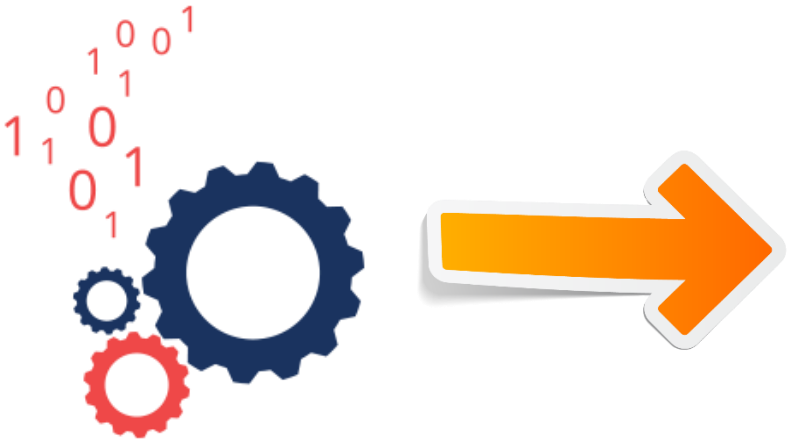
Predictive Analytics

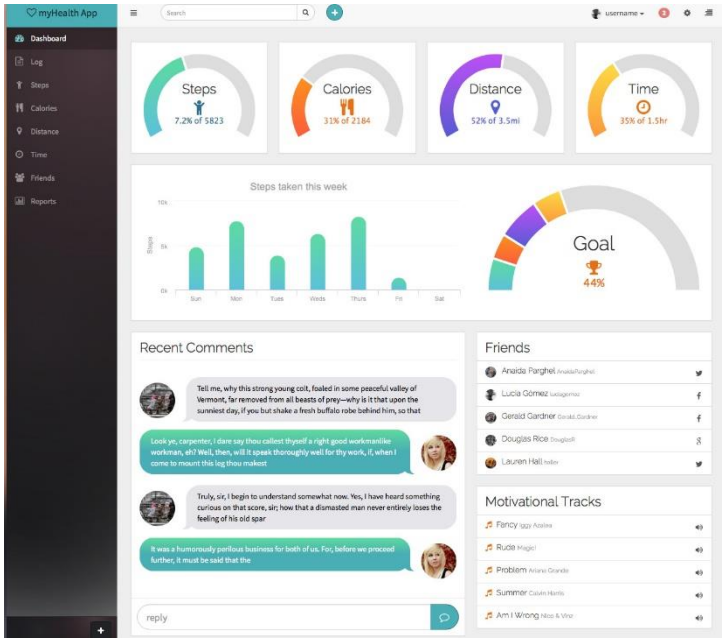
Project what will happen;
possible outcome indication



BIG DATA ANALYTICS

A set of fundamental concepts/principles that underlie techniques for extracting useful knowledge from large datasets containing a variety of data types. To uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful business information

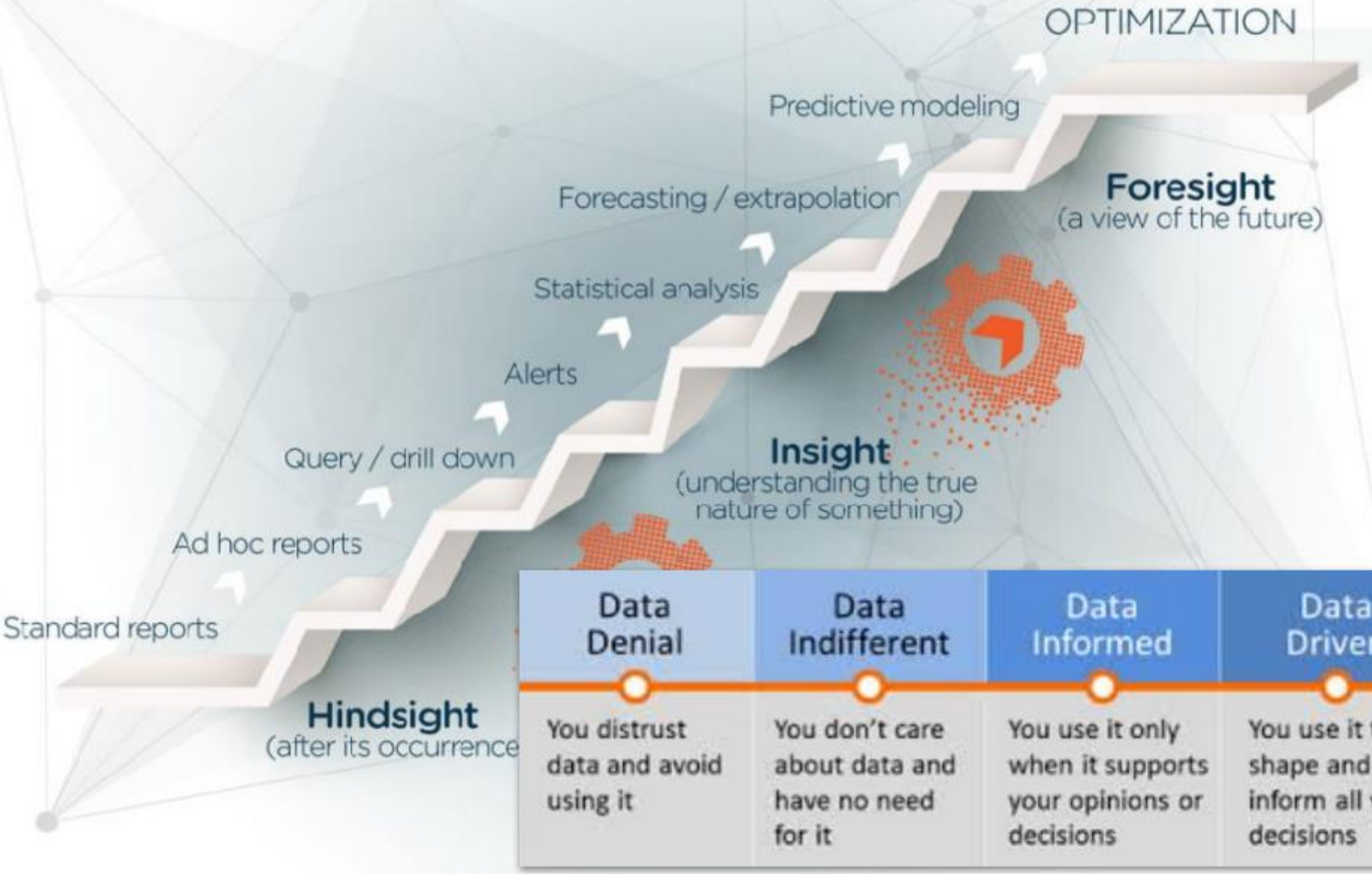




CREDIT SCORING

- 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.

Data Driven Decision Making



BIG DATA STUDIO

TOOLS AND APPLICATIONS

- An internet tool for internet crawling, searching, data aggregation, data analysis, data visualization, data extraction and image analysis.
- Intelligence Algorithm can be developed to
  - Smart search
  - Anticipate specific incident
  - Alert the authorities on the mention of concepts such as terrorism, bombs, guns, etc. across various information sources including CCTV.
  - Track people in the building

BIG DATA PLATFORM

Data Input

Messaging, and Web Services

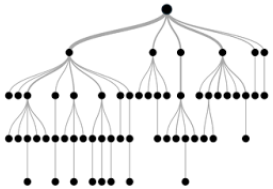
EDW, OLAP

Social Media, Weblogs

Machine Devices, Sensors

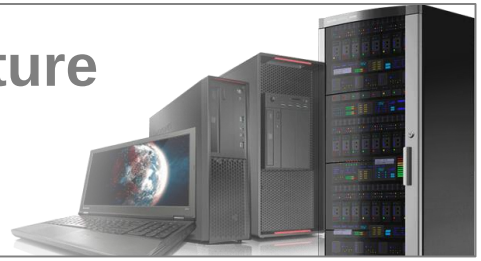
Tools and Data Services

Predictor Software



Solr Spark MLlib hadoop HDFS

IT Infrastructure



Visualization

Project Inception

Data Scientists และคณะทำงาน Big Data ของแต่ละกลุ่มงาน รวมกำหนดโจทย์ที่เหมาะสมและตั้งโครงการพัฒนาและทดสอบโมเดลคณิตศาสตร์นำร่องที่เหมาะสม

Problem Analysis

ทีม Data Scientists สำรวจข้อมูลที่มีอยู่ในปัจจุบันตามโจทย์นำร่องที่กำหนด เพื่อประเมินความพร้อมและแปลงจากความต้องการในเชิงปัญหาให้เป็นข้อกำหนดในเชิงข้อมูลและระบบ

Exploratory Data Analysis

สำรวจการกระจายตัวของข้อมูลเพื่อทำความเข้าใจข้อมูล และหาความสัมพันธ์ระหว่างตัวแปรในข้อมูล ในขั้นตอนนี้ทีมงานจะต้องการตัวอย่างข้อมูลจริง ระบุข้อมูลที่ต้องการเพิ่มเติมและเริ่มเตรียมข้อมูลจริง

Predictive Analytic

ทีมงานนำข้อมูลที่จัดเตรียมในเบื้องต้นมาใช้ในการสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อการทำนาย โดยใช้เทคนิคและอัลกอริธึ่มต่างๆ และทดสอบความแม่นยำของโมเดลคณิตศาสตร์

Visualization Dashboard

ออกแบบวิธีการแสดงผลโดยเลือกมิติของข้อมูลที่เหมาะสมบน Interactive Dashboard เพื่อให้คณะทำงานทดลองใช้และสื่อสารกับทีมผู้บริหาร และผู้ปฏิบัติ ให้สามารถนำเอาความเข้าใจดังกล่าวไปแปลงเป็นแผนการพัฒนาต่อยอด

Implementation & Deployment

หลังจากผลลัพธ์เป็นที่พอใจแล้ว นักพัฒนาระบบเริ่มพัฒนาโปรแกรมตามรูปแบบของโมเดลคณิตศาสตร์ที่วางไว้ และตั้งค่าให้โปรแกรมให้ประมวลผลโมเดลแบบอัตโนมัติตามความถี่ที่วางแผนไว้ จากนั้นติดตั้งระบบซอฟท์แวร์เพื่อการใช้งานจริง

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.

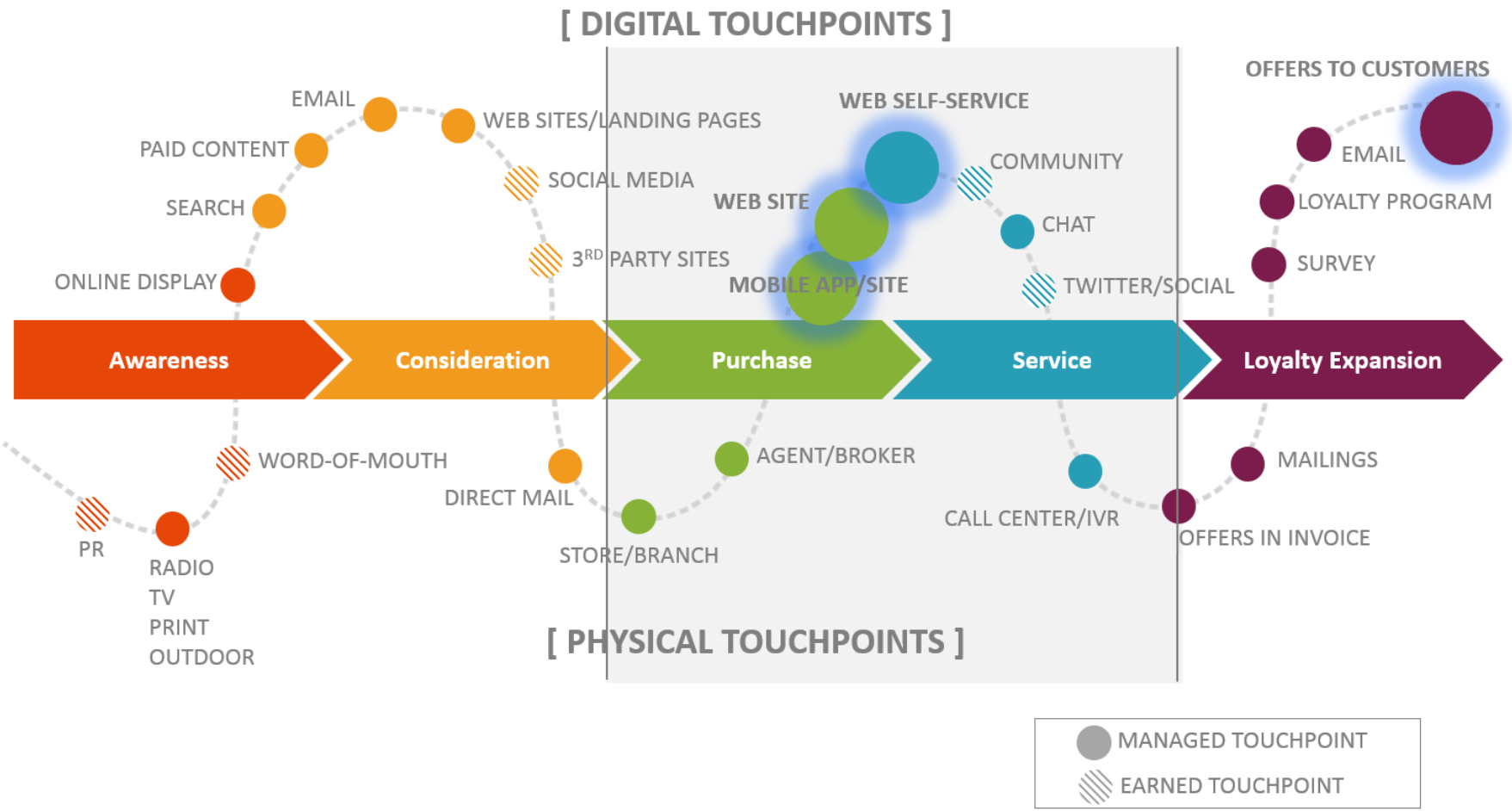
Big Data Technology

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)

CUSTOMER TOUCHPOINTS



Digital marketing-Mckinsey

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.

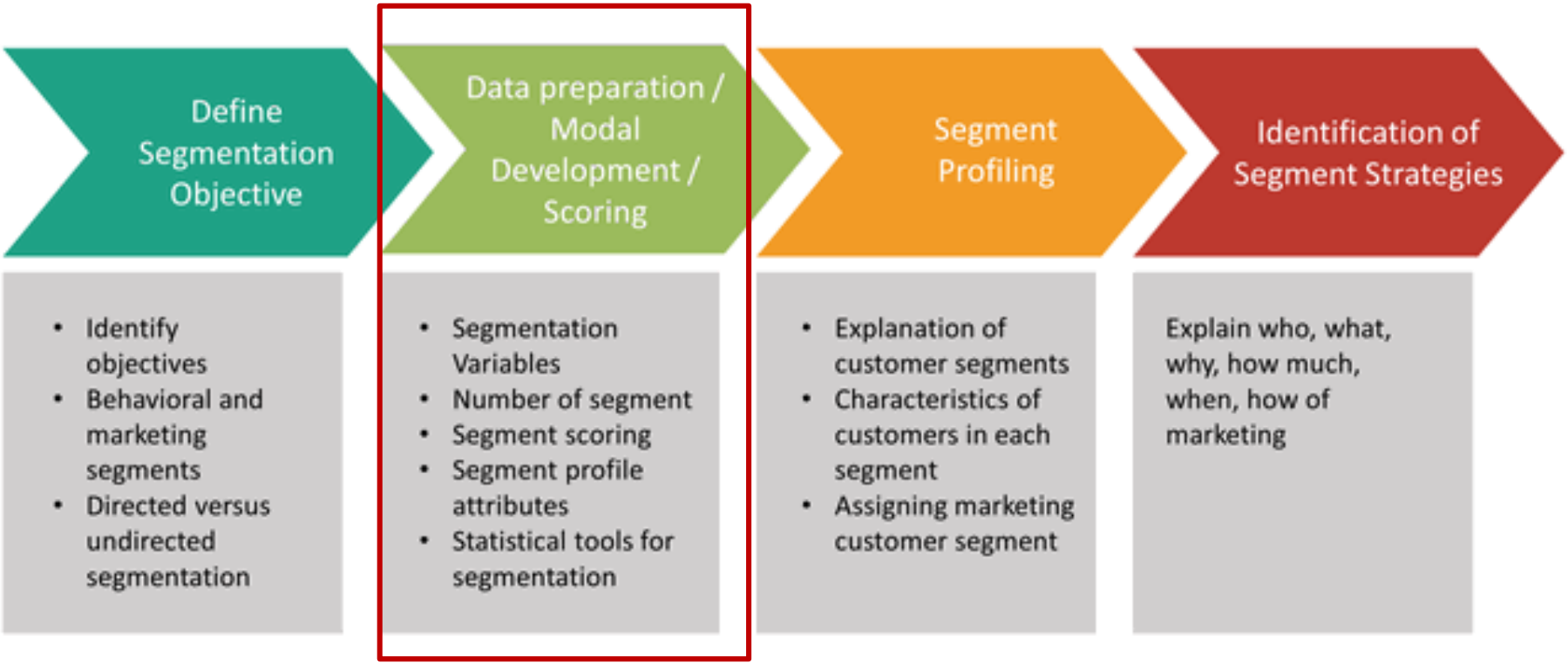
CAMPAIGN ANALYTICS

Tracking customer interactions (Google Analytics)

Linking marketing activities and conversions across systems

Linking social media marketing insights

CUSTOMER SEGMENTATION



- The better you understand, who are the customers at your store, the easier it is to tailor to your their needs.
- WHAT you sell and HOW you sell it.

PREDICTIVE and PERSONALIZED Omni-Channel Offers

Cross-channel data collection with unified data and seamless UE

INTEGRATE

- WEB BEHAVIORS
- IN-STORE BEHAVIORS
- CALL CENTER
- USAGE BEHAVIOR
- DEMOGRAPHICS

Generating customer profiles to ...

PREDICT

- PRODUCT PREFERENCE
- PURCHASE PROPENSITY
- NEED FOR SUPPORT

Product

Starting with a range of products AIS has to offer across different channels.

Customer

Develop a profile of customers (demographic and behaviors across channels).

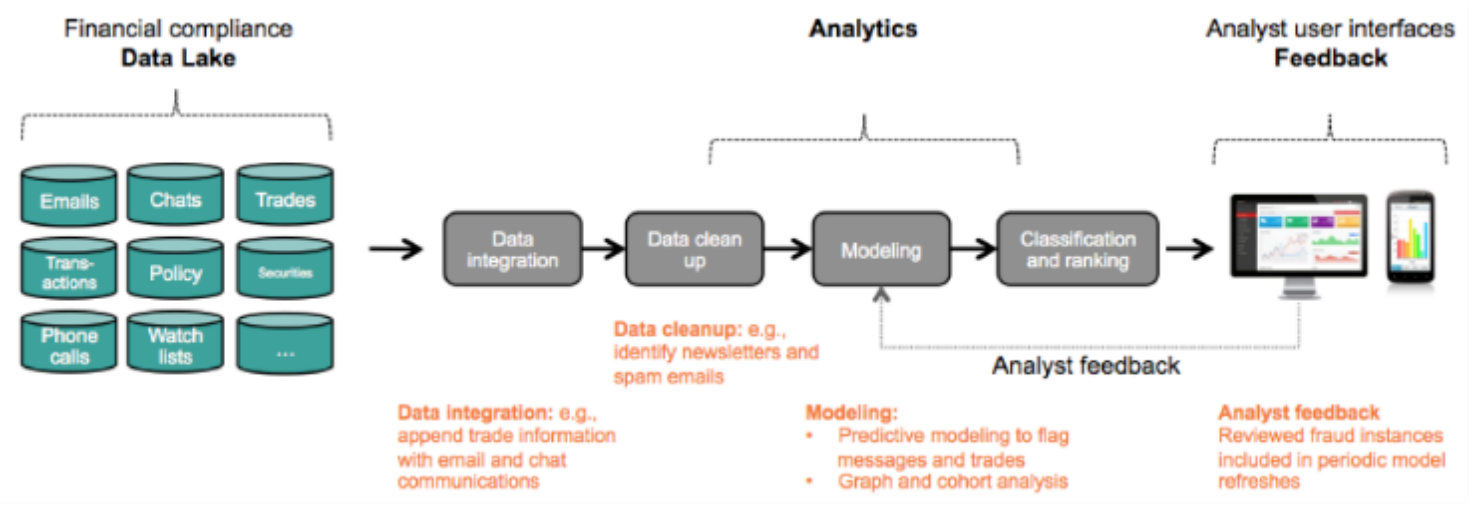
Make a model of factors that predict propensity to buy certain products.

Recommendation

Make recommendations to customers (with little pushes).

Convert likely sales to real sales across retail stores and call center.

FINANCIAL COMPLIANCE

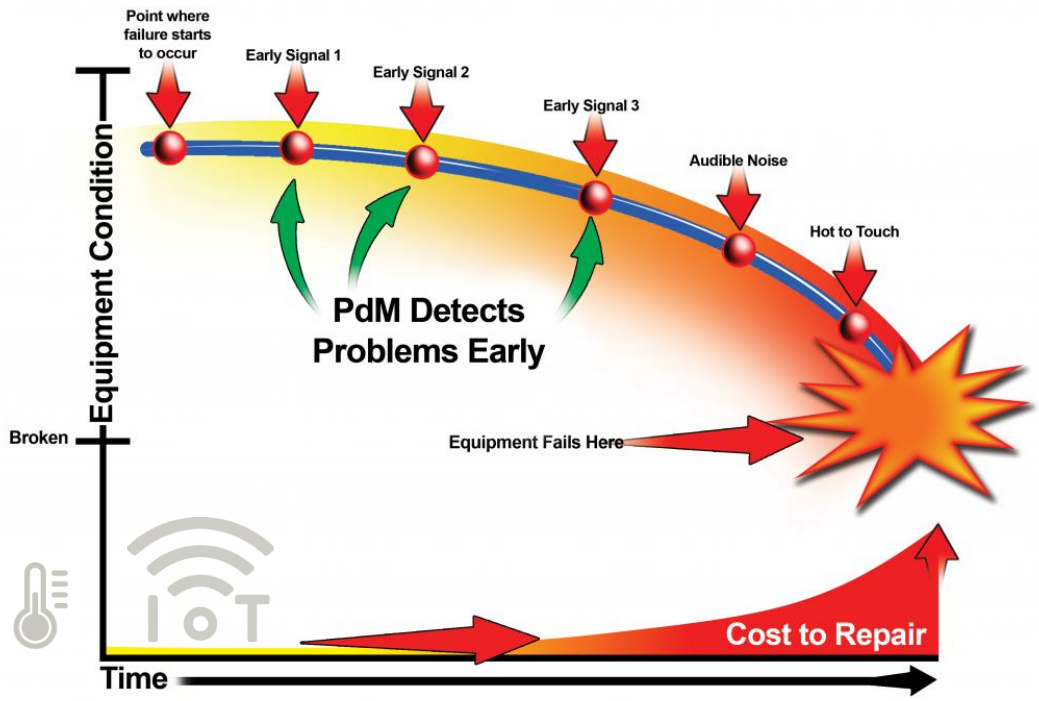


- **Integrate –** Feed in the trade information from email and chat
- **Classification** - To identify irrelevant messages such as automated notifications, or newsletters.
- **Graph analysis** - To build communication profiles of individuals. This technique is often used in security analytics and malware detection to identify anomalous behavior. Graph analysis can establish hot spots of fraudulent activity based on who is talking to whom.
- **Text analytics** - To identify the language behind fraud, determine the sentiment and certainty in the language of a trader before and after executing trades.

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

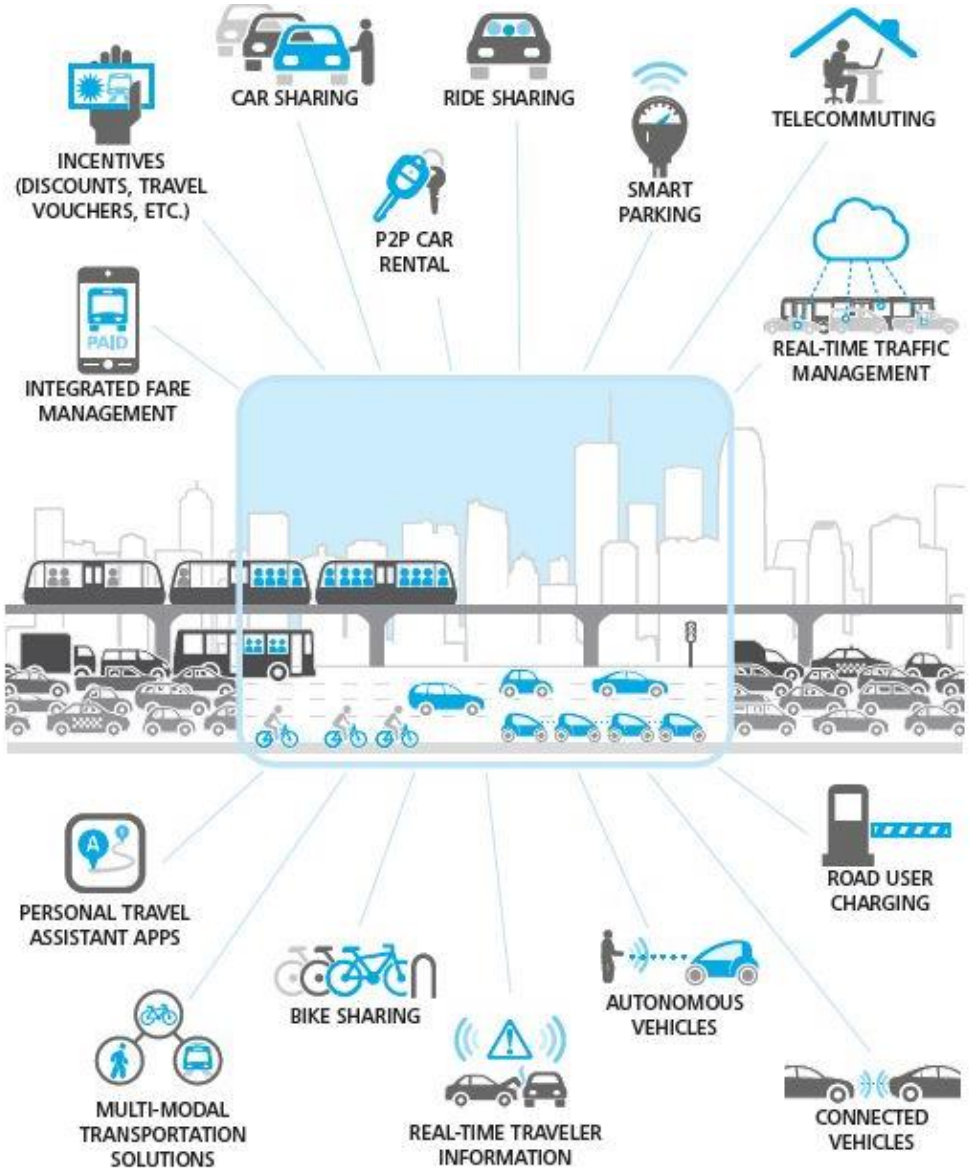


PRECISION AGRICULTURE

- Optimize farming decisions in order to maximize yields.
- Farmers can make proactive decisions based on future conditions
  - when to plant, fertilize and harvest crops
- Adopt wireless, cloud-connected systems, and place sensors throughout the fields
  - Provide real-time monitoring: measure temperature and humidity of the soil and air
  - Take pictures of fields using satellite imagery and robotic drones. The images over time show crop maturity.
  - Predictive weather modeling show pinpoint conditions 24-48 hours in advance
  - Automate everyday agriculture operations
  - Provide data analysis for smart decision making (day-to-day, season-to-season)

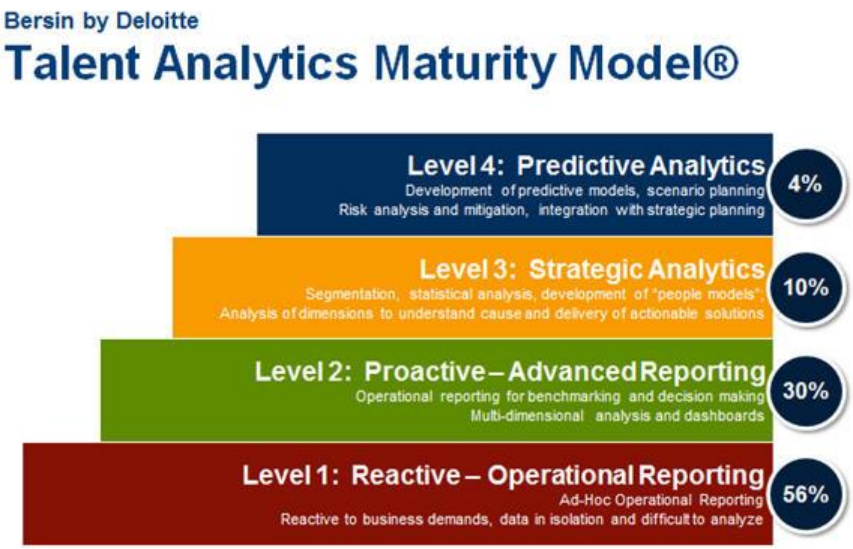
TRANSPORATION

Traffic Prediction
Personalized Insurance
Connected Car



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.



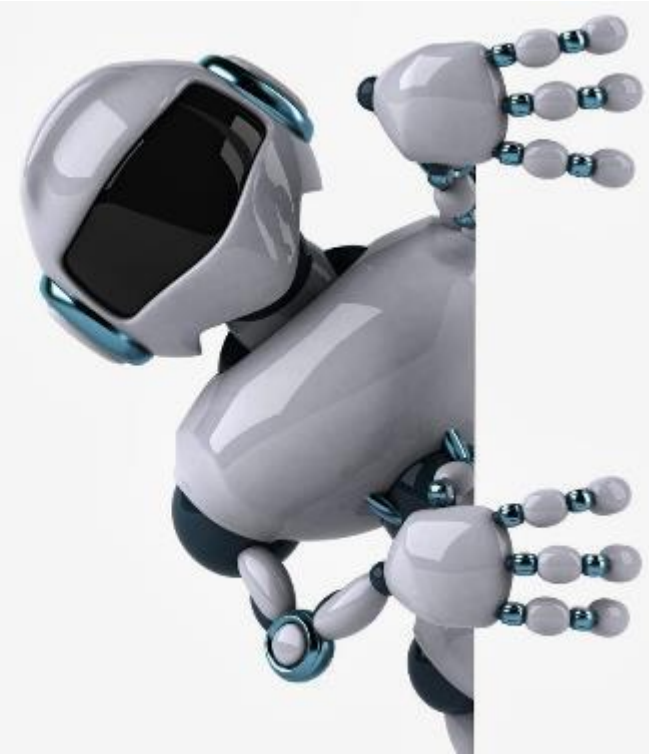
Ref: Forbe

ANSWERING BUSINESS QUESTIONS

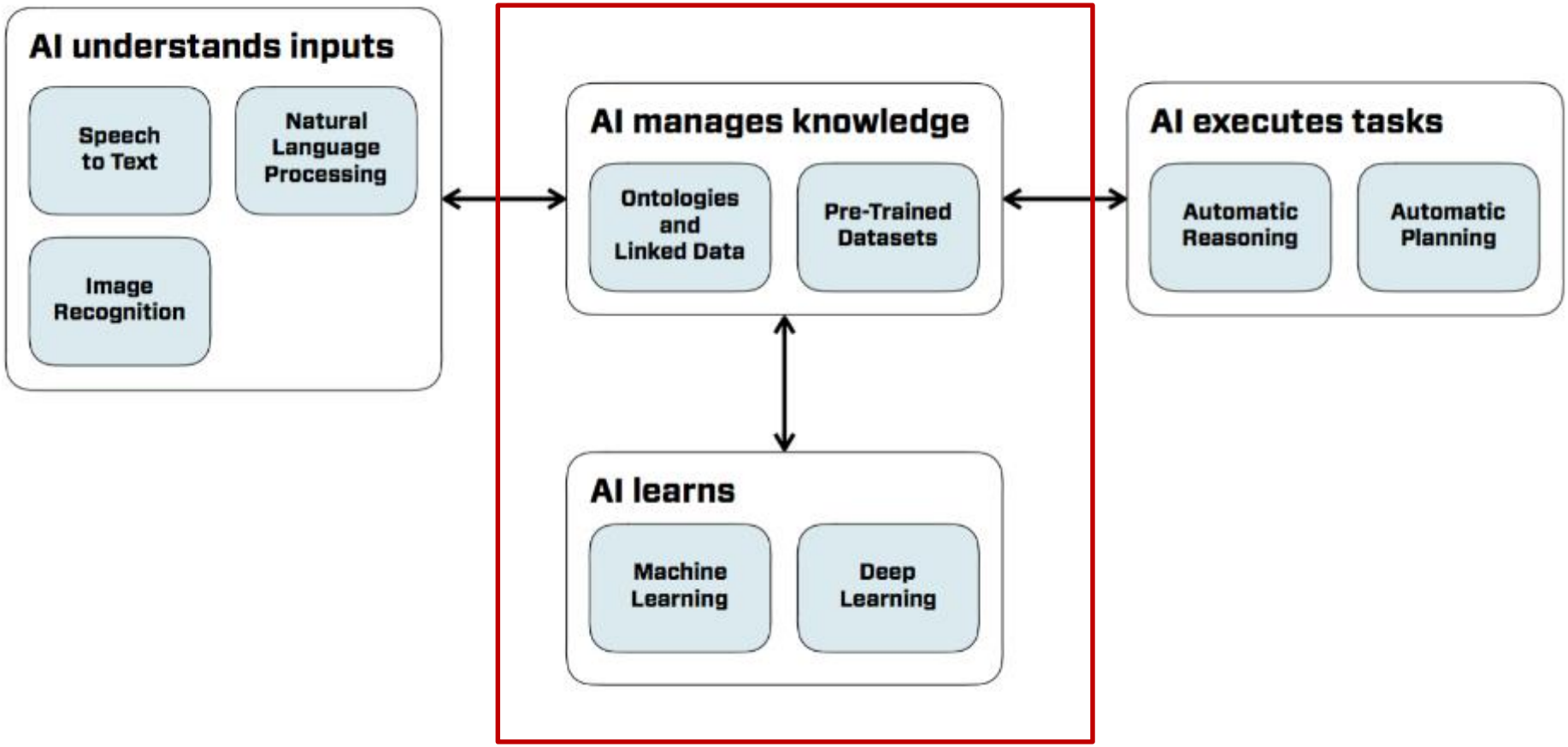
- Who are the most profitable customers?
  - A straightforward database query, if “profitable” can be defined clearly.
- Is there really a difference between the profitable customers and the average customer?
  - Statistical Hypothesis testing
- But who really are these customers? Can I characterize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

MACHINE LEARNING and AI

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



PUTTING AIs TO WORK



Credit: nicolamattina.it

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 Genera Int gence

Work in Progress

specialized in all areas.

Artificial Super Int gence

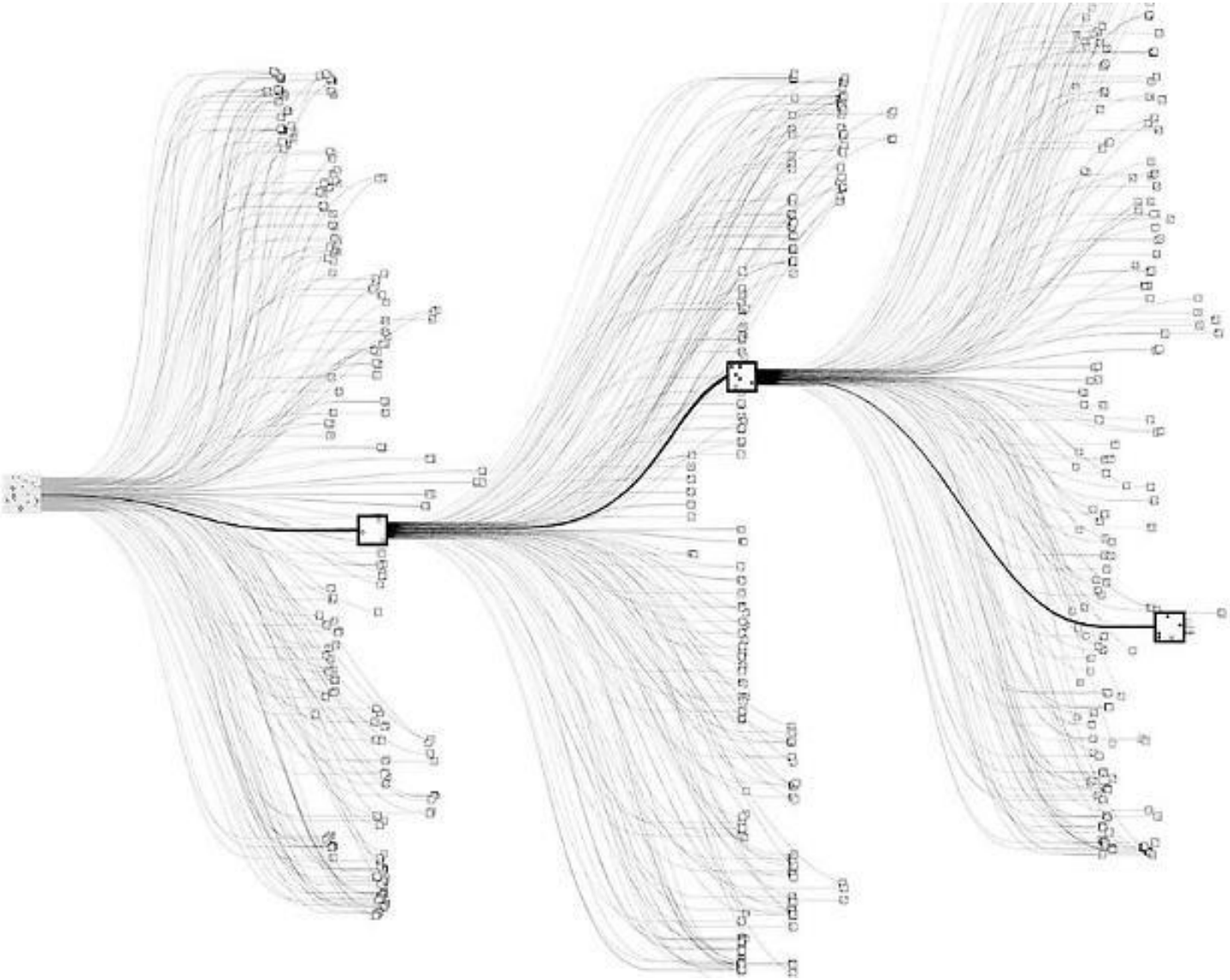
Singularity

harter than human in every way.

ALPHA GO

KBA
Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
ALPHAGO
00:00:54
AlphaGo
Lee Sedol

ALPHA GO



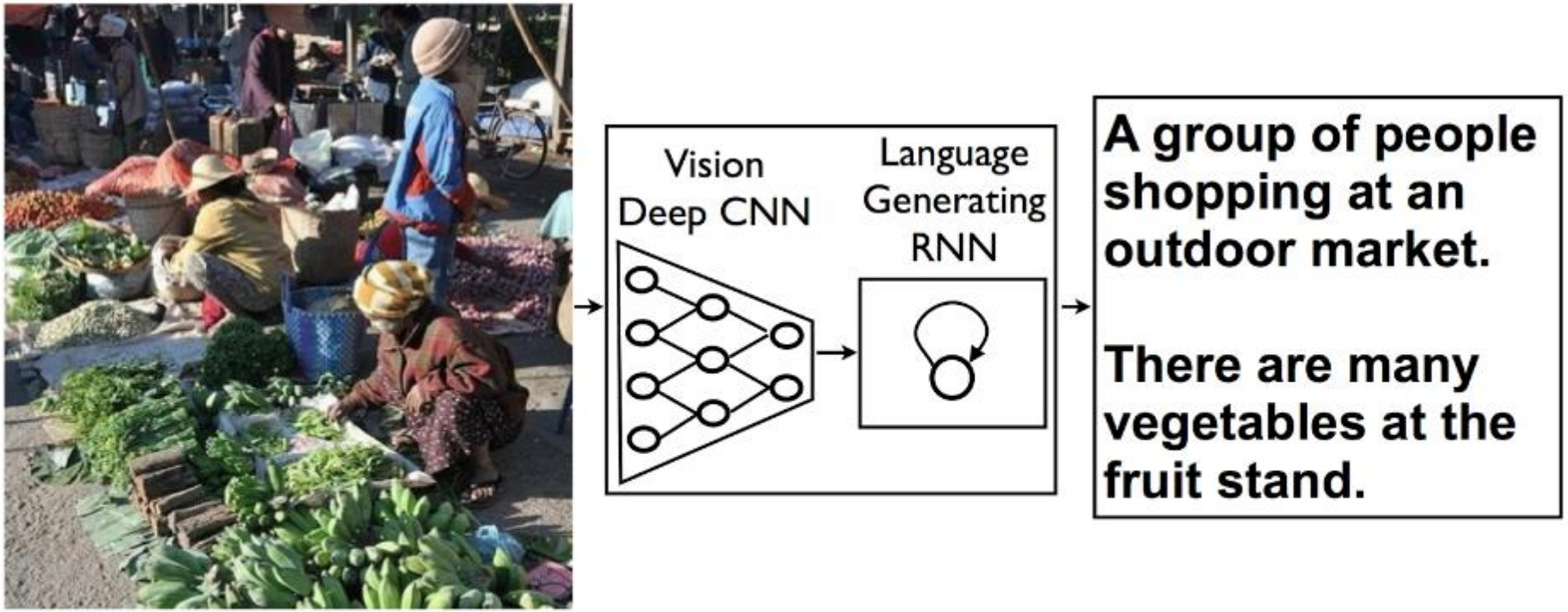
ALPHA GO ZERO

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



Multi-modal Processing

credit: gigaom.com

State of the Art Deep Learning

Generate high-resolution images based on text.

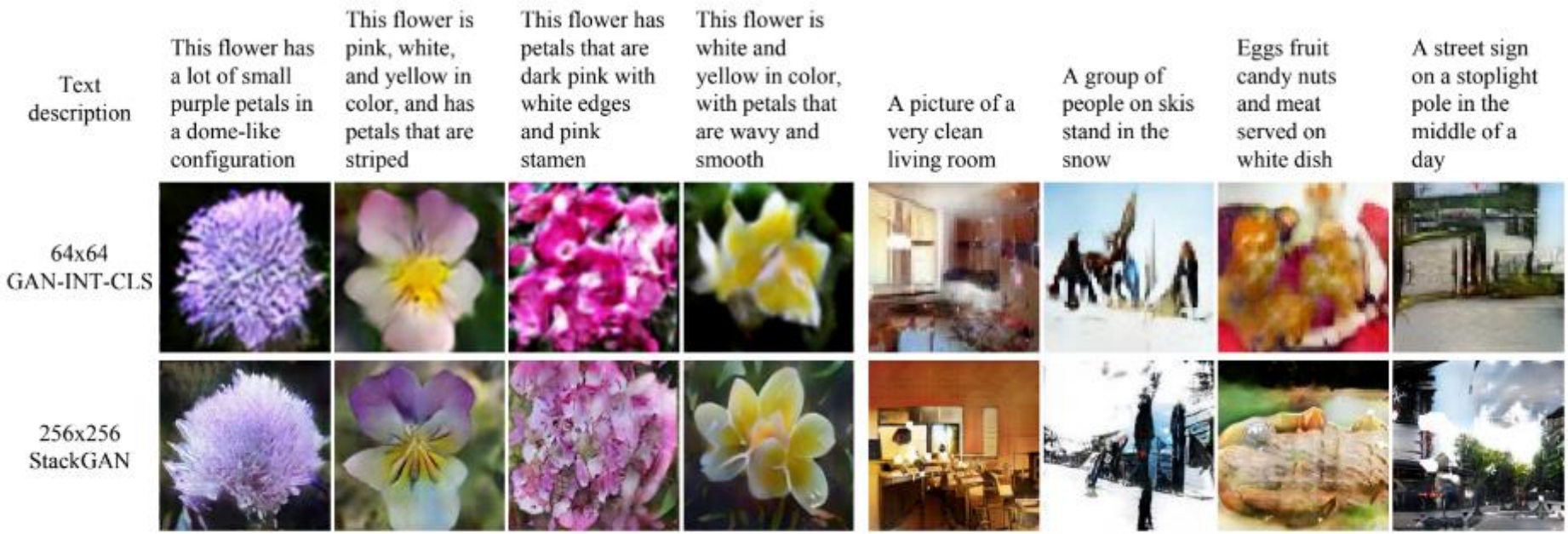
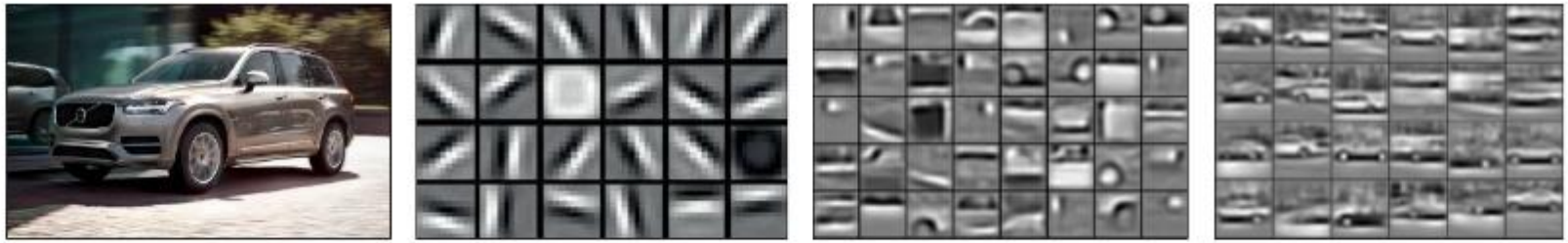


Figure 4. Example results by our StackGAN and GAN-INT-CLS [26] conditioned on text descriptions from Oxford-102 test set (leftmost four columns) and COCO validation set (rightmost four columns).

How does AI Learn?



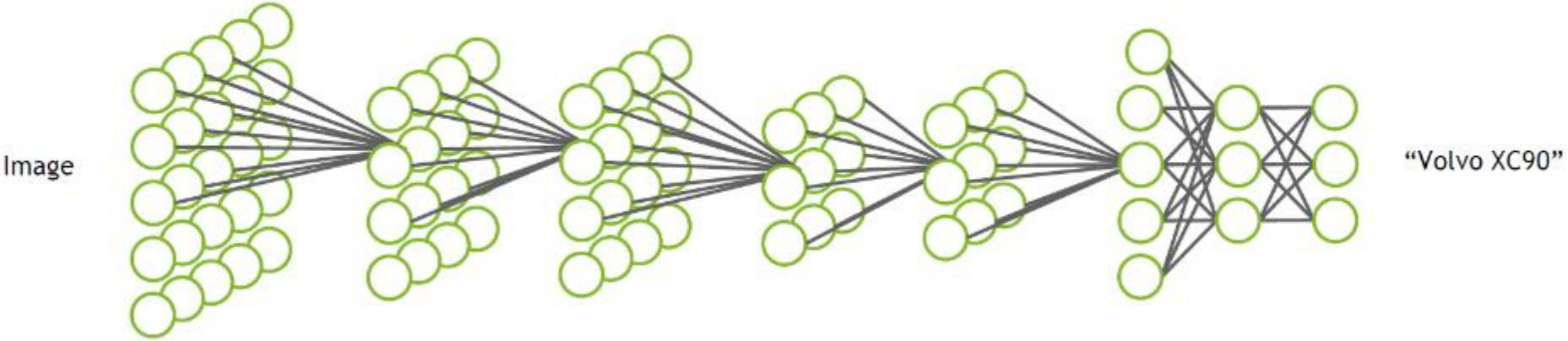
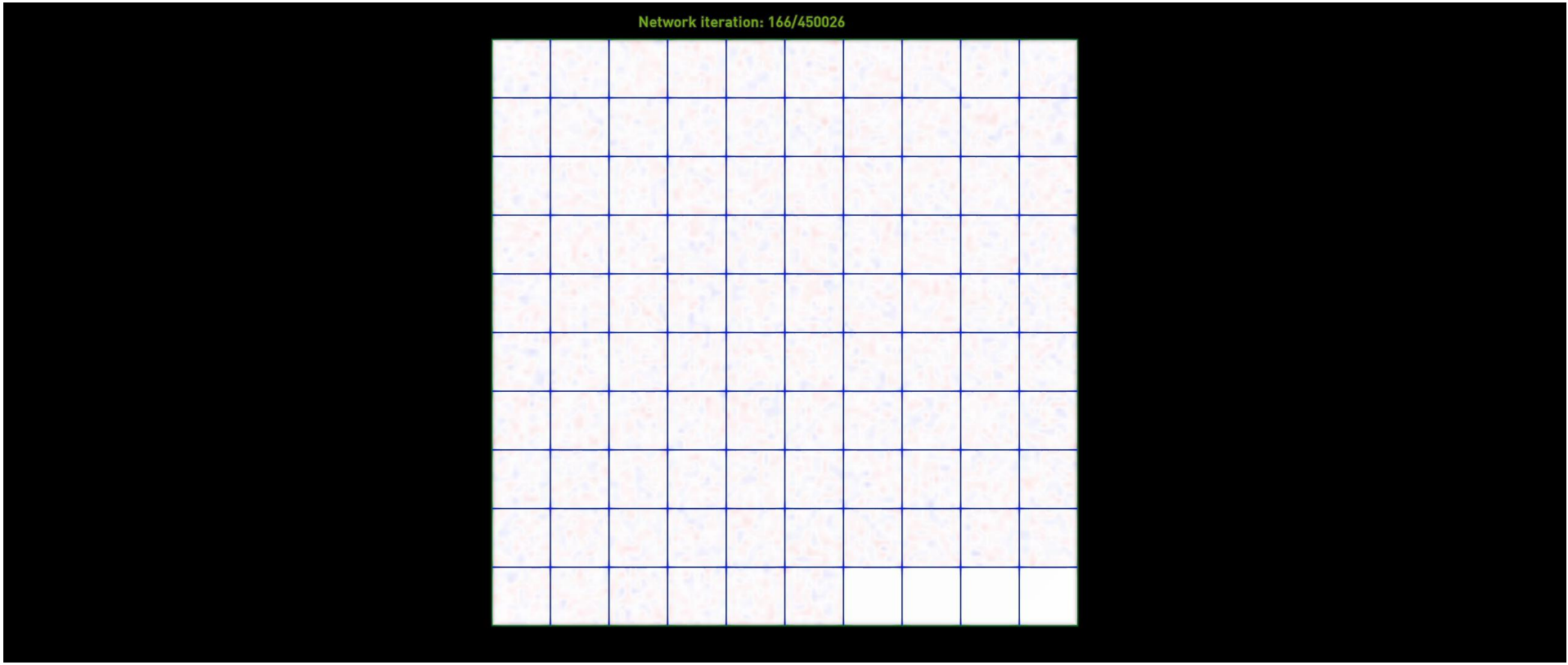


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011.
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

NVIDIA

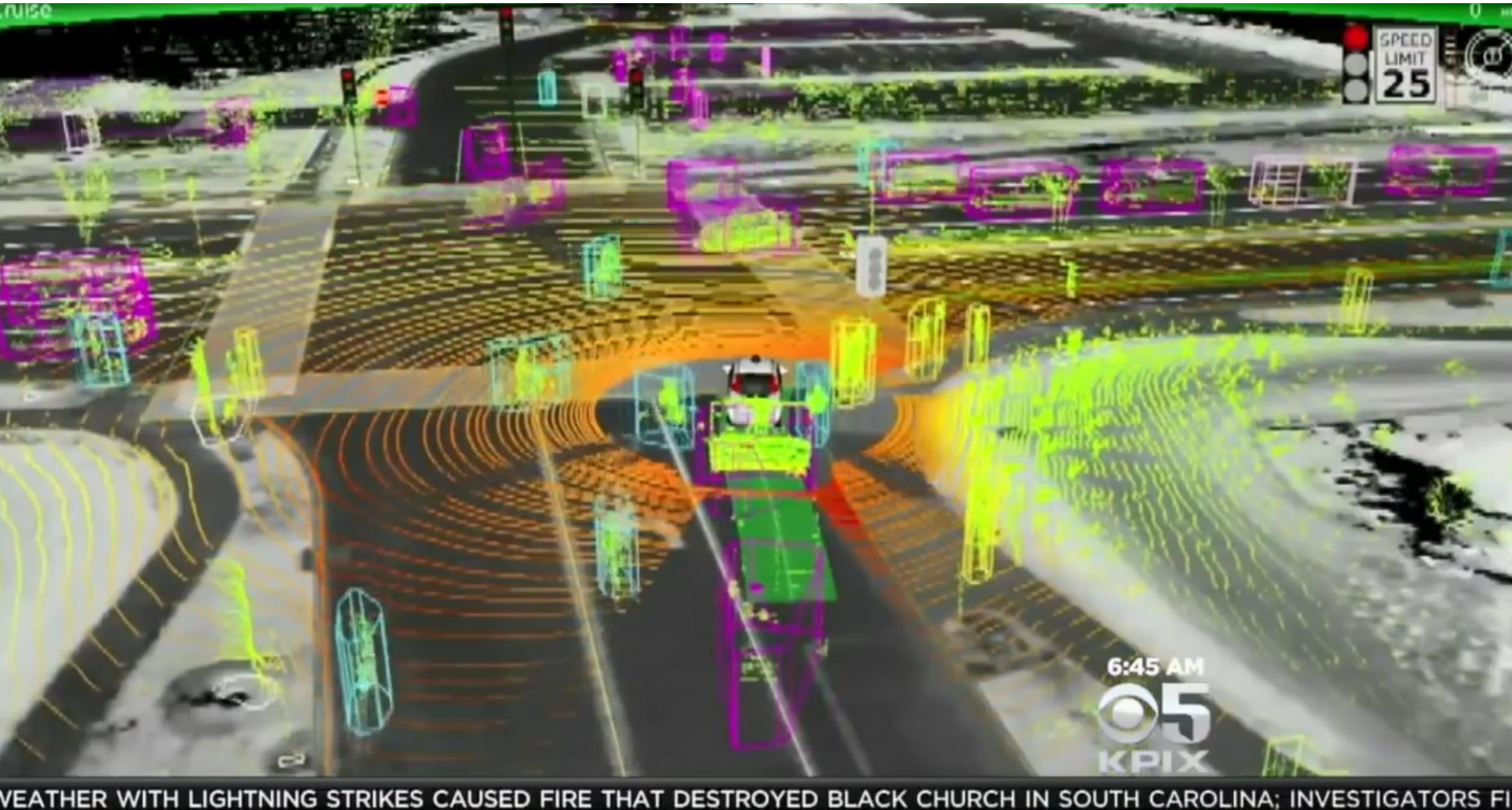
Deep Learning



Real-time Face Recognition

Customer Profiling

Autonomous Car and Drone



A Day in the Life of a Kiva Robot

THE CHEF BOT

Predicting ingredient-pairing and creating recipes

- IBM Chef Watson goes through tens of thousands of recipes, regional and cultural knowledge and statistical, molecular and food pairing theories. Chef Watson generates new recipes, often with unusual ingredients.

- McCormick developed FlayorPrint by researching about food tastes, textures, aromas, and preparation techniques. Data then are encoded into a set of algorithms that helped select recipes.

- Kitchology uses a collective wisdom of chefs. The algorithms reflect on what people who make food designs learn in culinary schools.

MACHINE LEARNING IN FOOD

- Advertising selection: Machine suggests food ads based on your digital pattern on food interest
- Prepopulating shopping lists: You write salads a lot on your shopping list, machine prepopulate your shopping list with vegetable, salad dressing, taking into account the size of your family and dietary restrictions.
- Which product will sell: Machine can assess which attributes of food are relevant to acceptance by your customer segments. It helps suppliers to the right product decisions from design to packaging to branding (Companies are experiencing 12 product failures for every 1 success).

https://blog.kitchology.com/

Pre-Crime Intelligence System

Palantir

Beware by intrado®

DAS by Microsoft NYPD

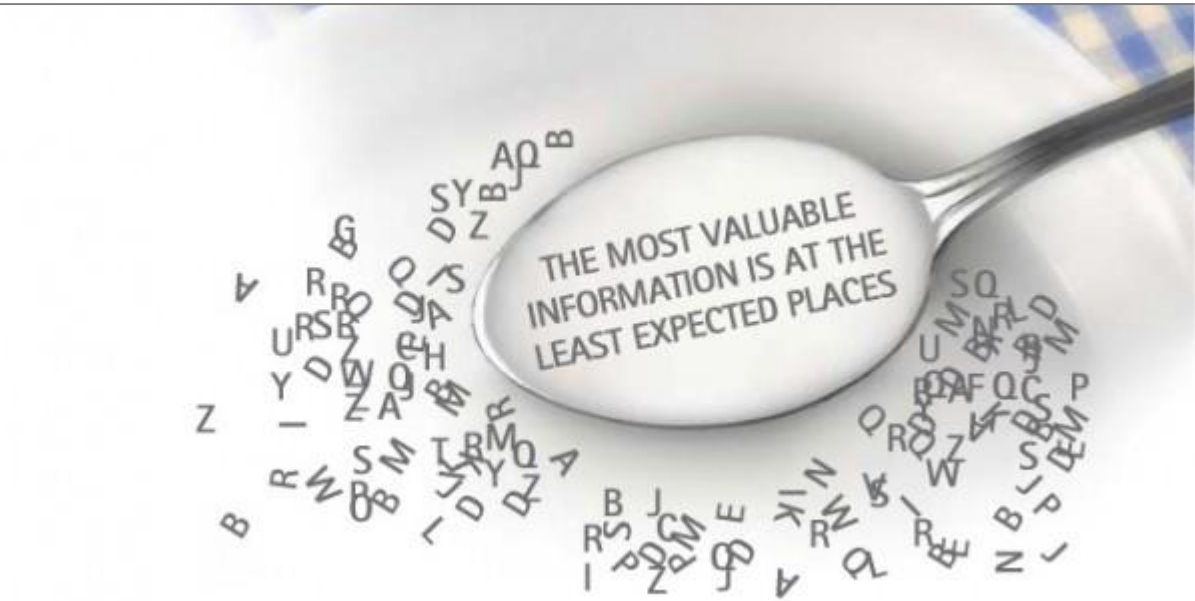
COPLINK. TODAY.
HIGH PERFORMANCE LAW ENFORCEMENT

internet crawling, searching, data aggregation, data analysis, data visualization, data extraction and image and VDO analysis.

PERSONALIZED MEDICINE

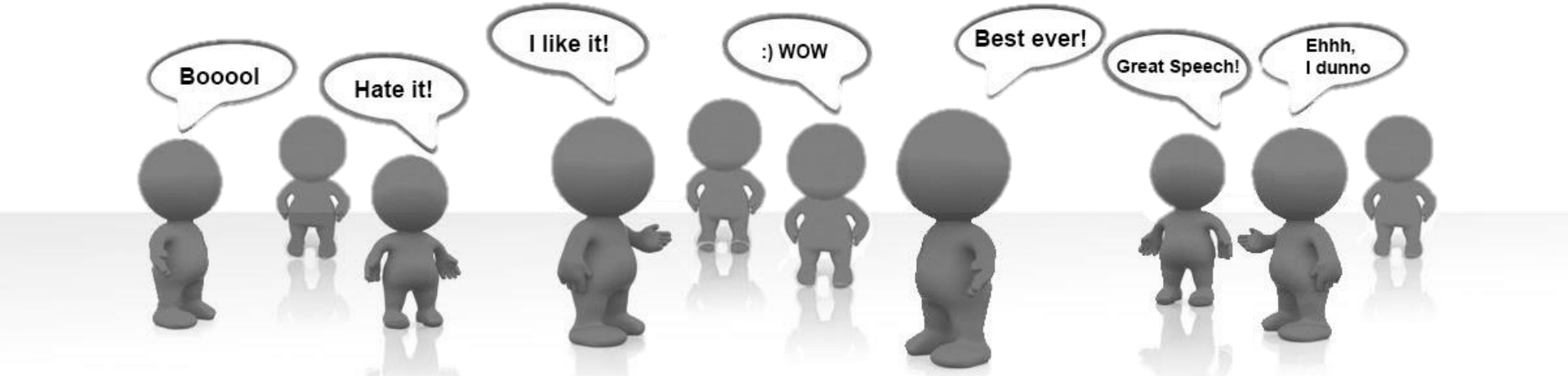
TEXT MINING AND NLP

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



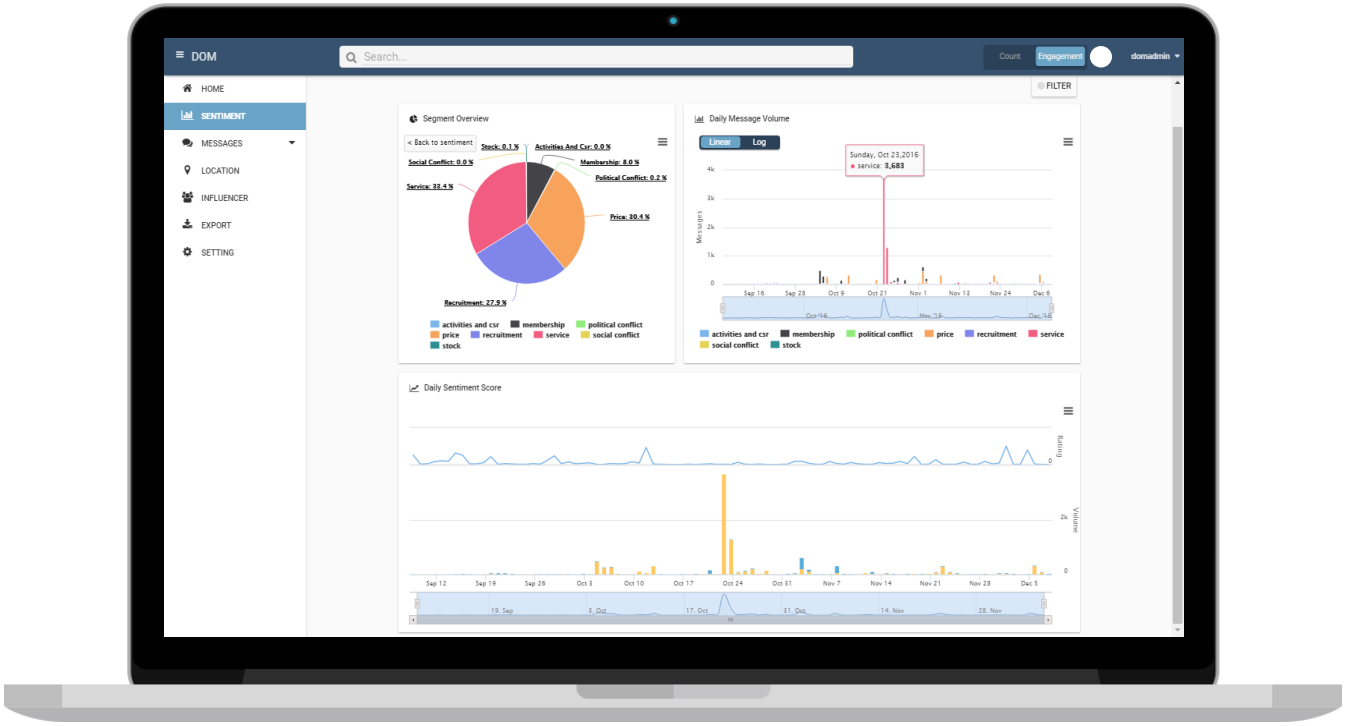
SENTIMENT ANALYSIS

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



VOICE OF CUSTOMERS

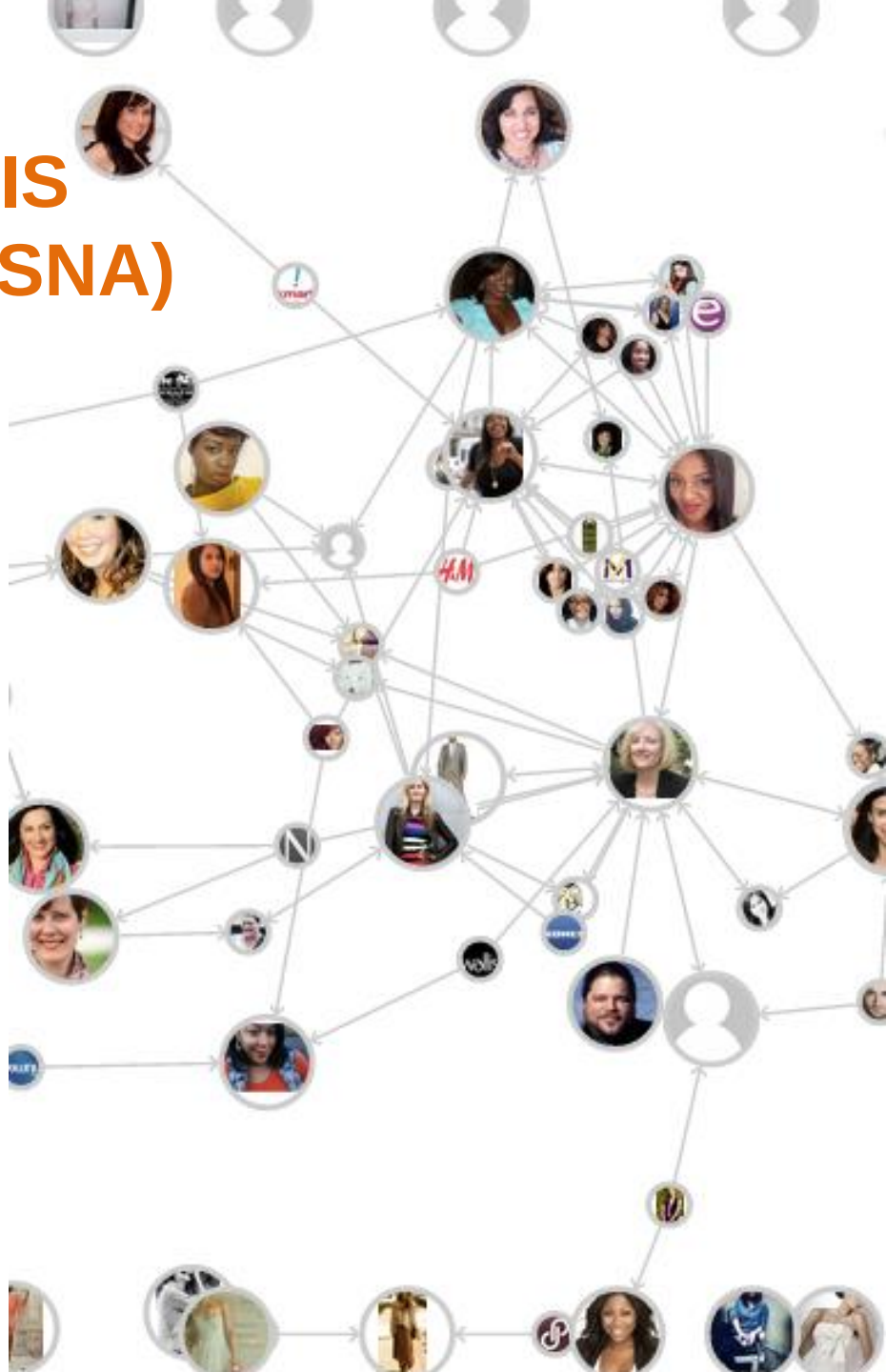
Call Center, Social Network, Front Offices



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

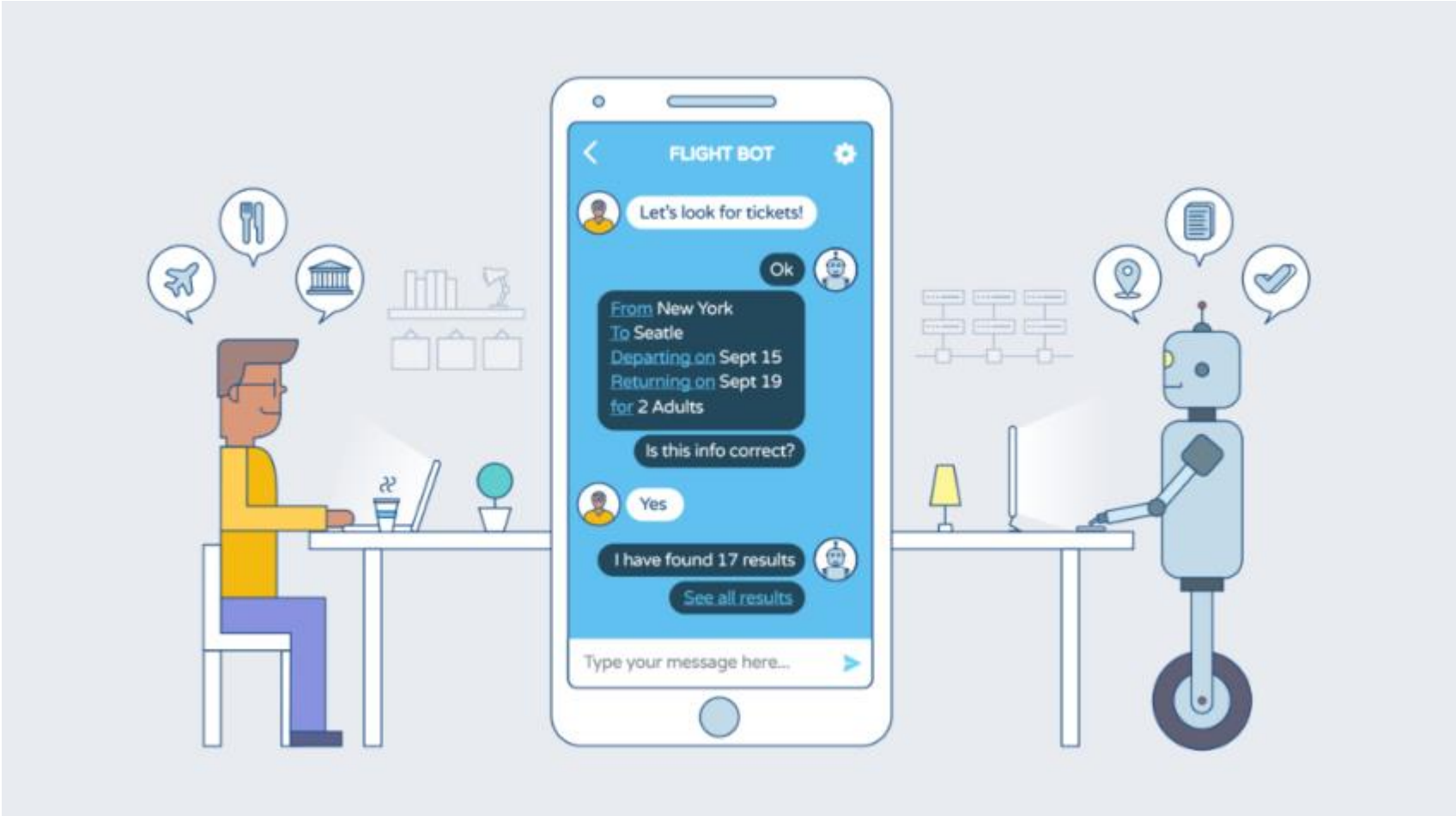
INFLUENCER ANALYSIS
Social Network Analysis (SNA)

- 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 referred to the most shaping a discussion about a topic.



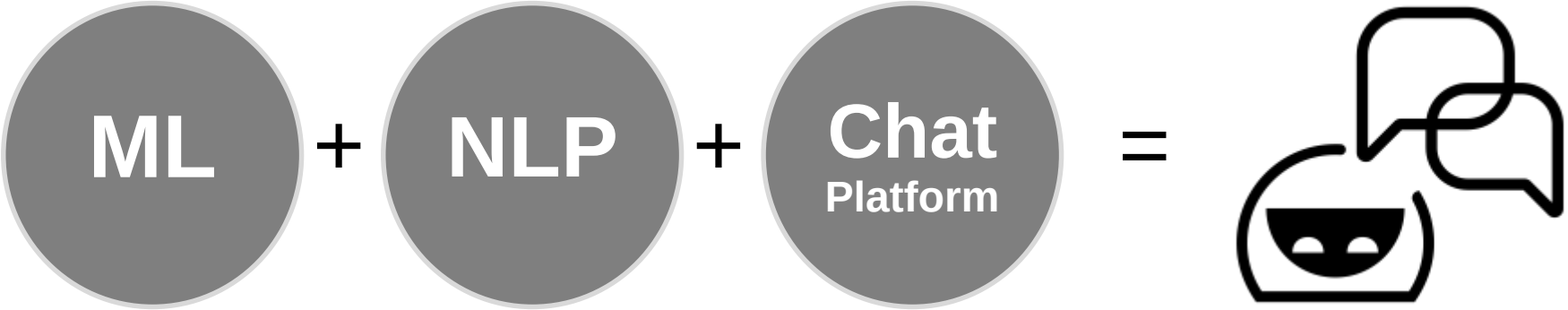
CHATBOT

Customer Services

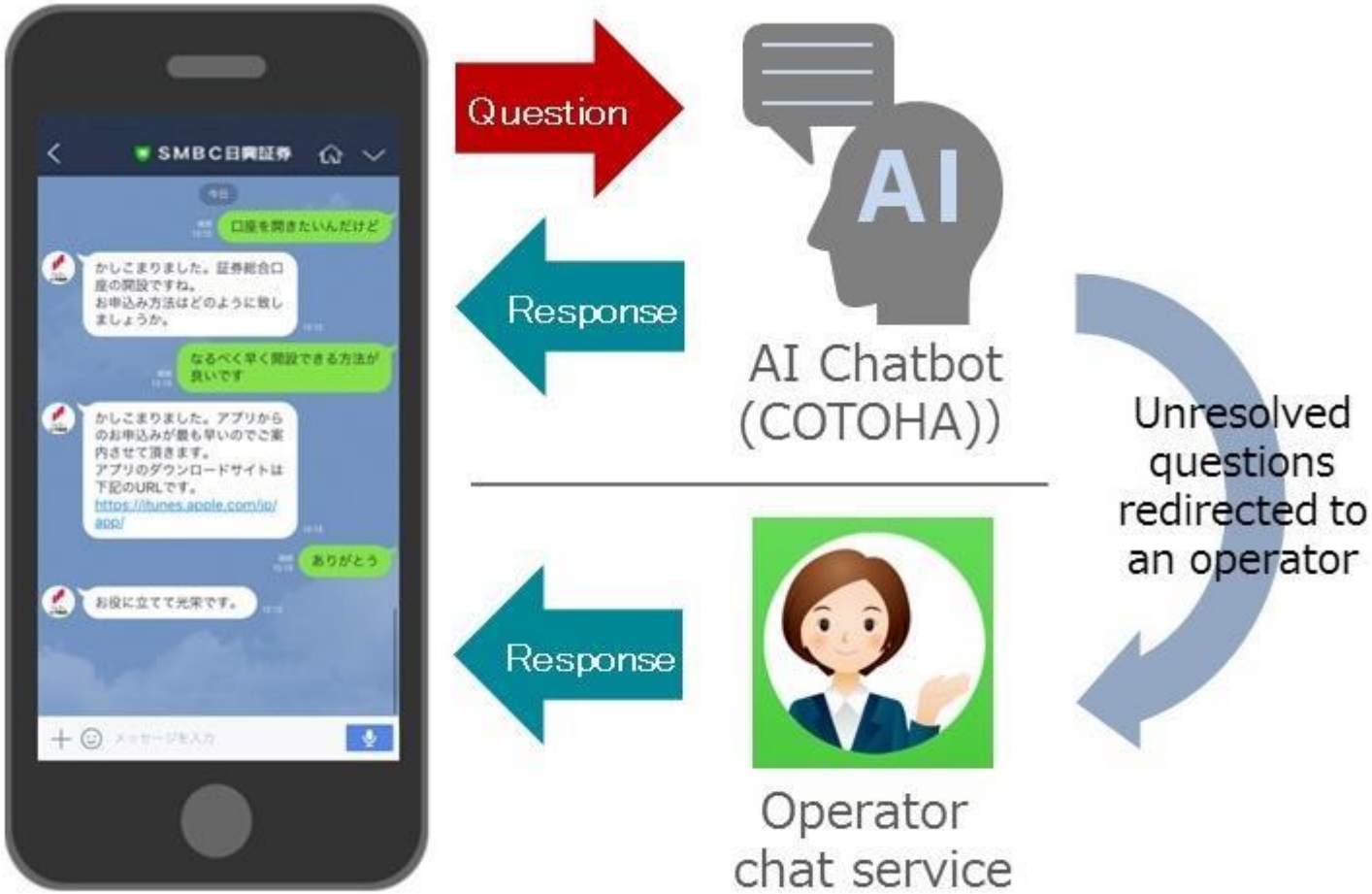


Chatbots have revolutionized the customer service space

Chatbot = Conversational interface powered by AI

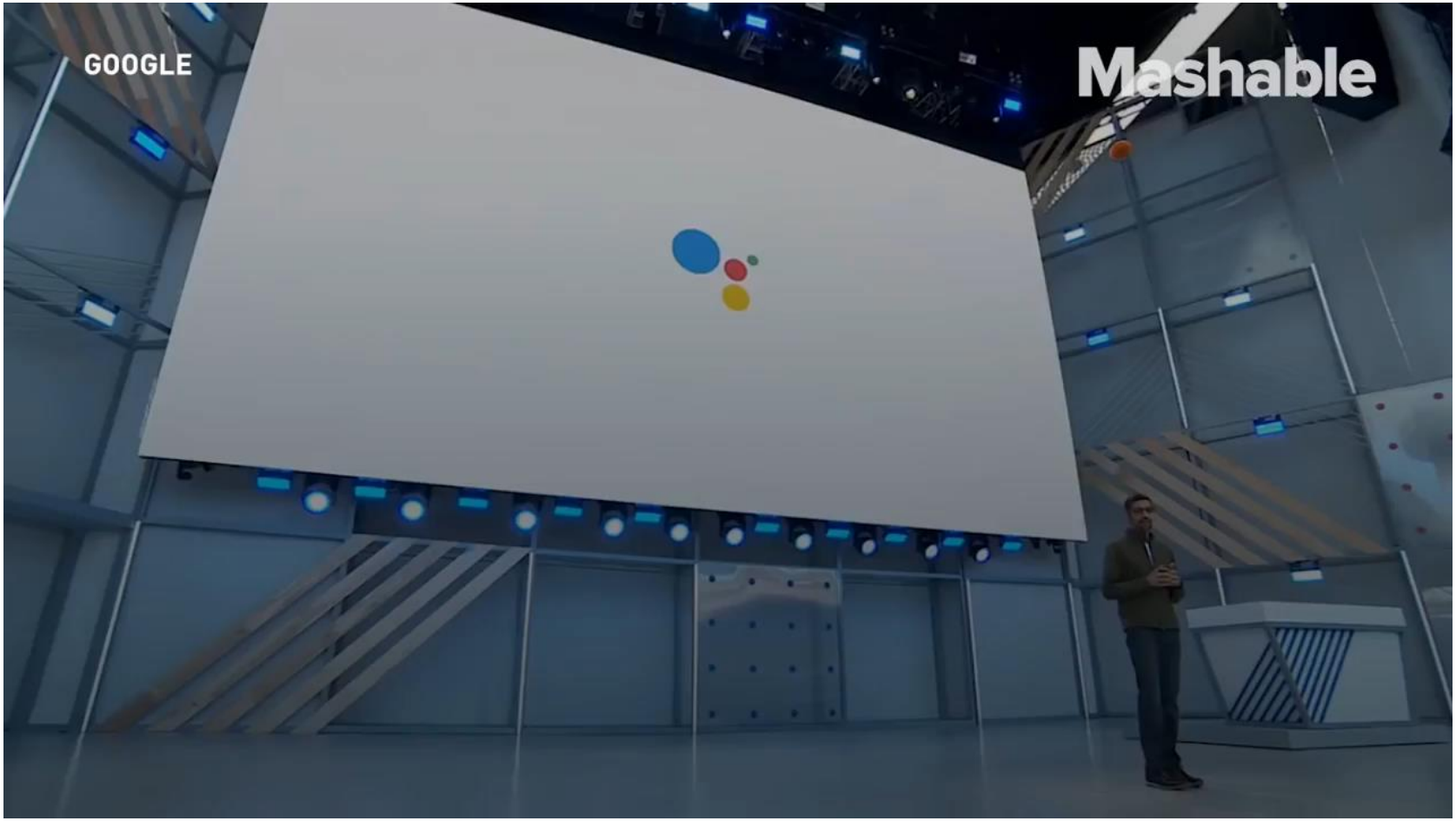


HUMAN HANDOVER



https://disruptive.asia/smbc-ntt-com-ai-chatbot/

Google Assistant !

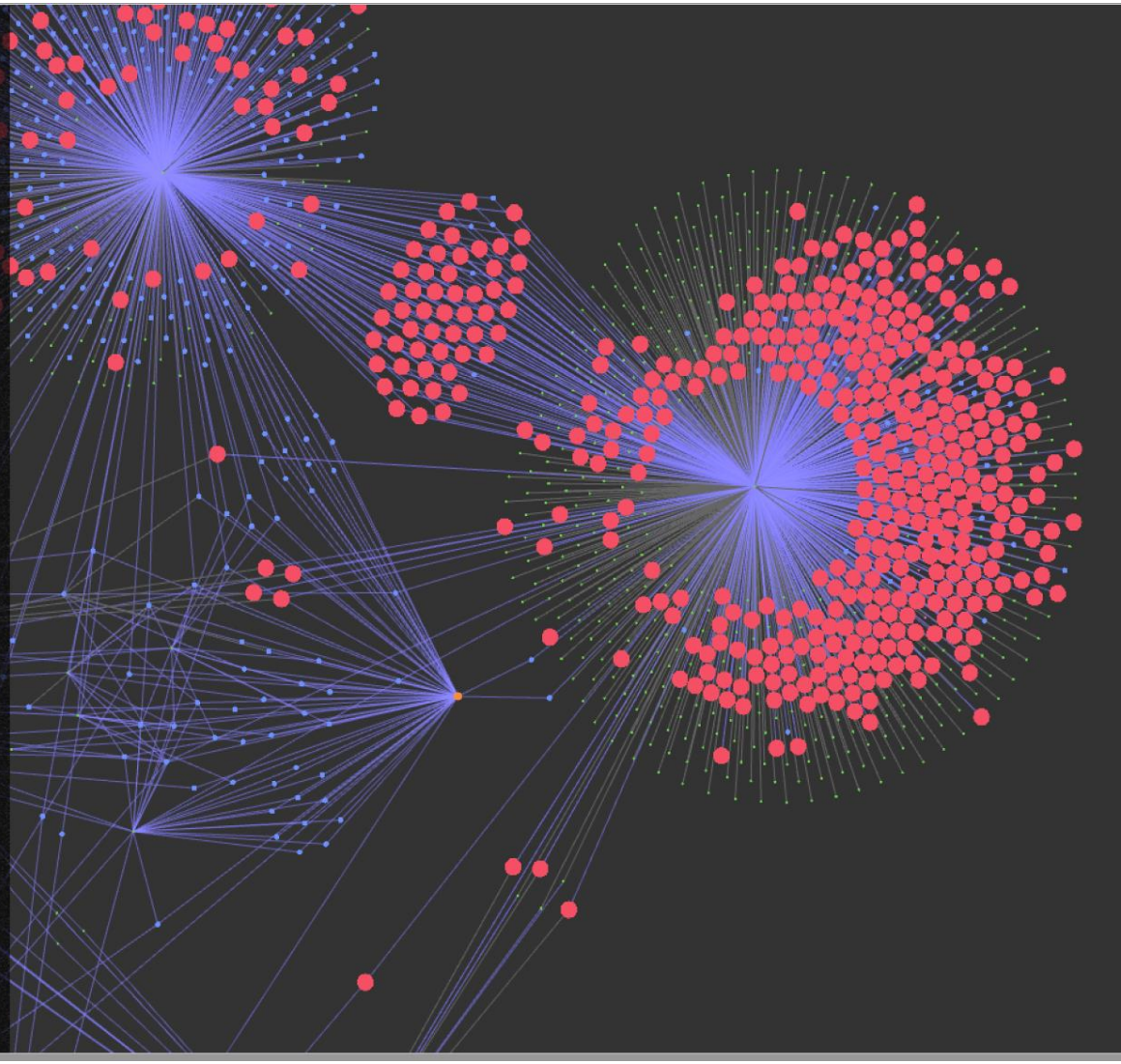


DIAGNOSTIC WITH AI

ACCELERATING DISCOVERIES WITH AI

New drugs typically take 12-14 years and $2.6 billion to bring to market. BenevolentAI is using GPU deep learning for NLP to bring new therapies to market quickly and more affordably. They've automated the process of identifying patterns within large amounts of research literature, enabling scientists to form hypotheses and draw conclusions quicker than any human researcher could. And using the NVIDIA DGX-1 AI supercomputer, they identified two potential drug targets for Alzheimer's in less than one month.

benevolent.ai



Ref: NVIDIA

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

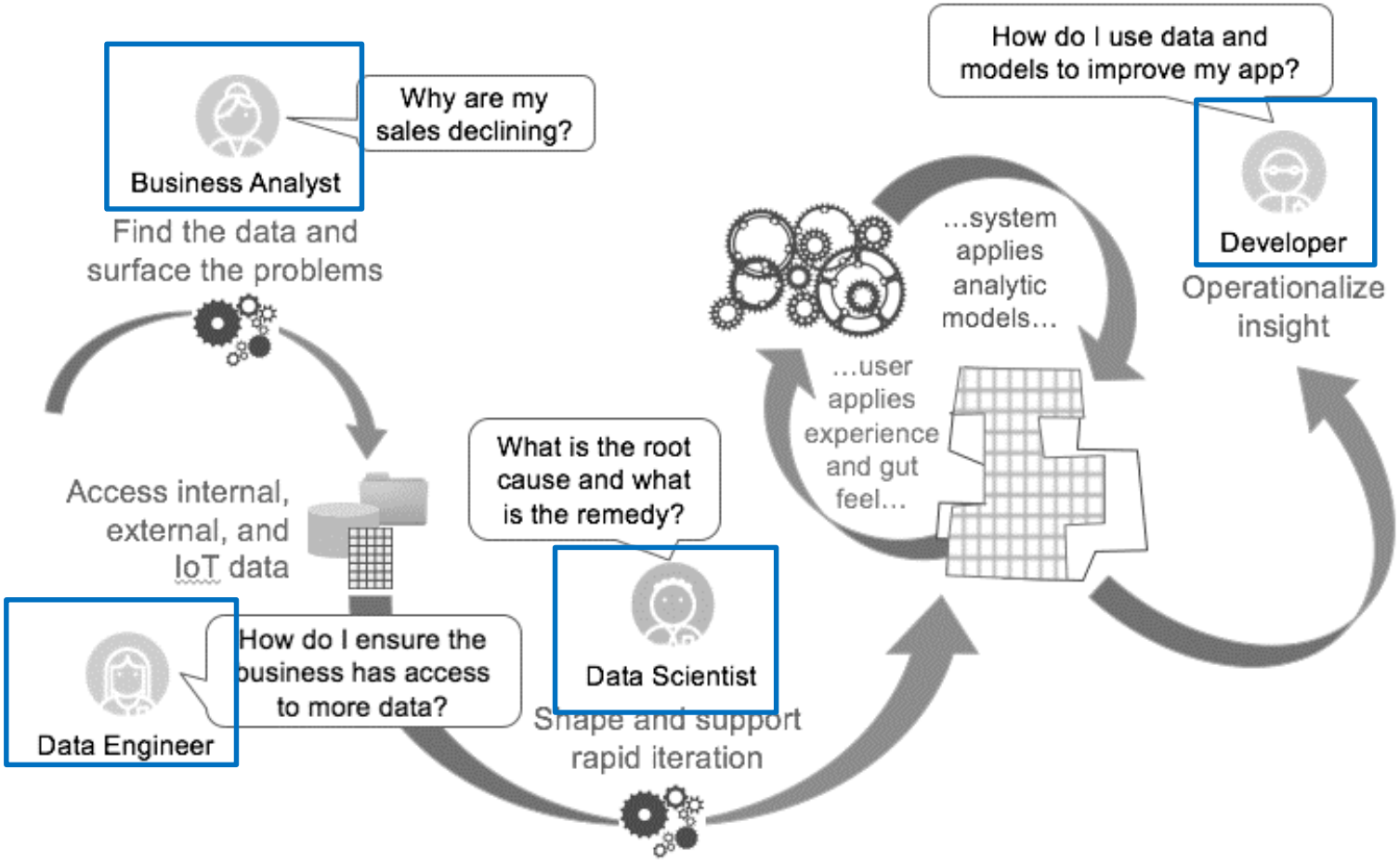
project what will happen next; provide indications of possible outcomes if trends continue.

Prescriptive Analytics

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

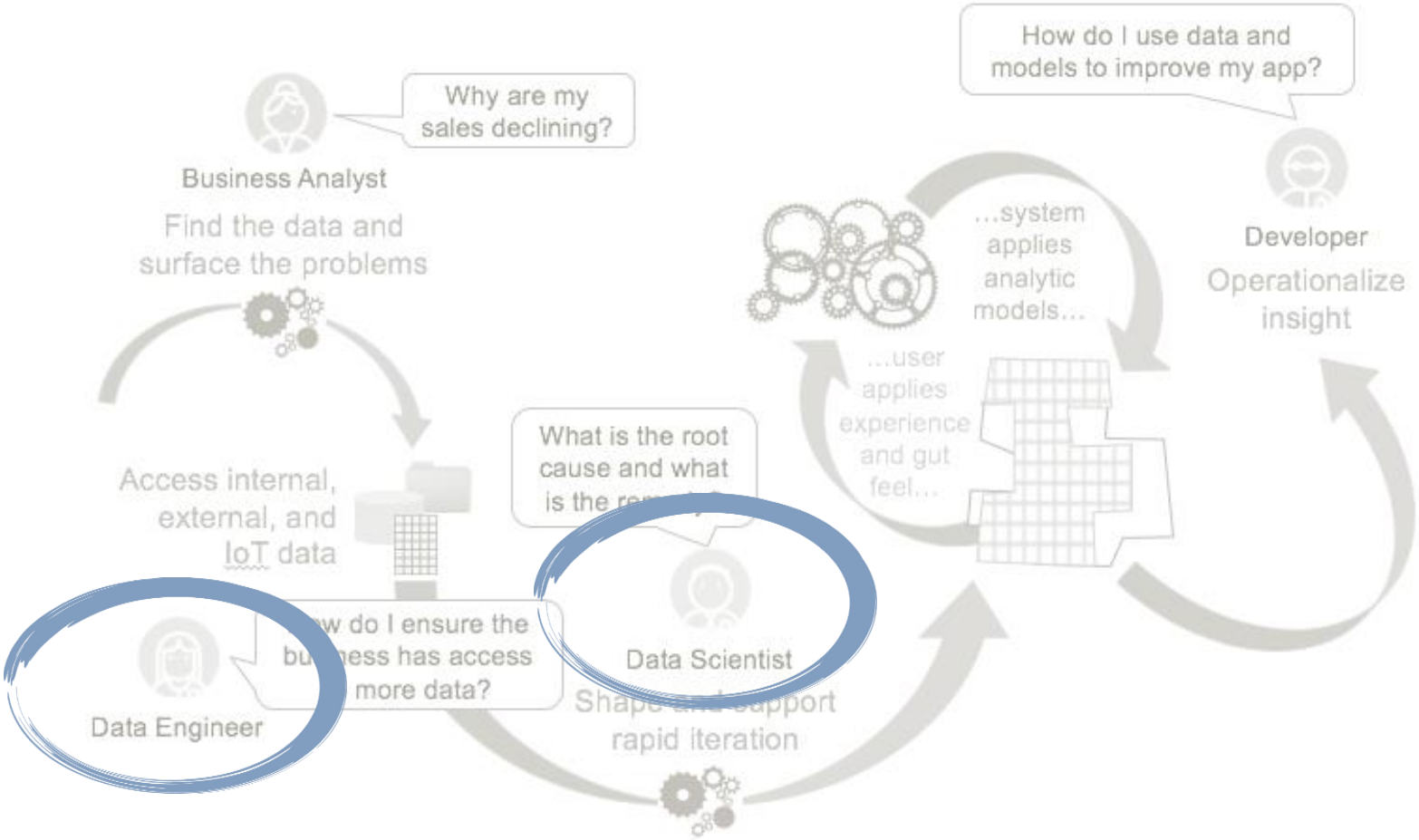
1 2 3

The 4 Personas for Data Analytics

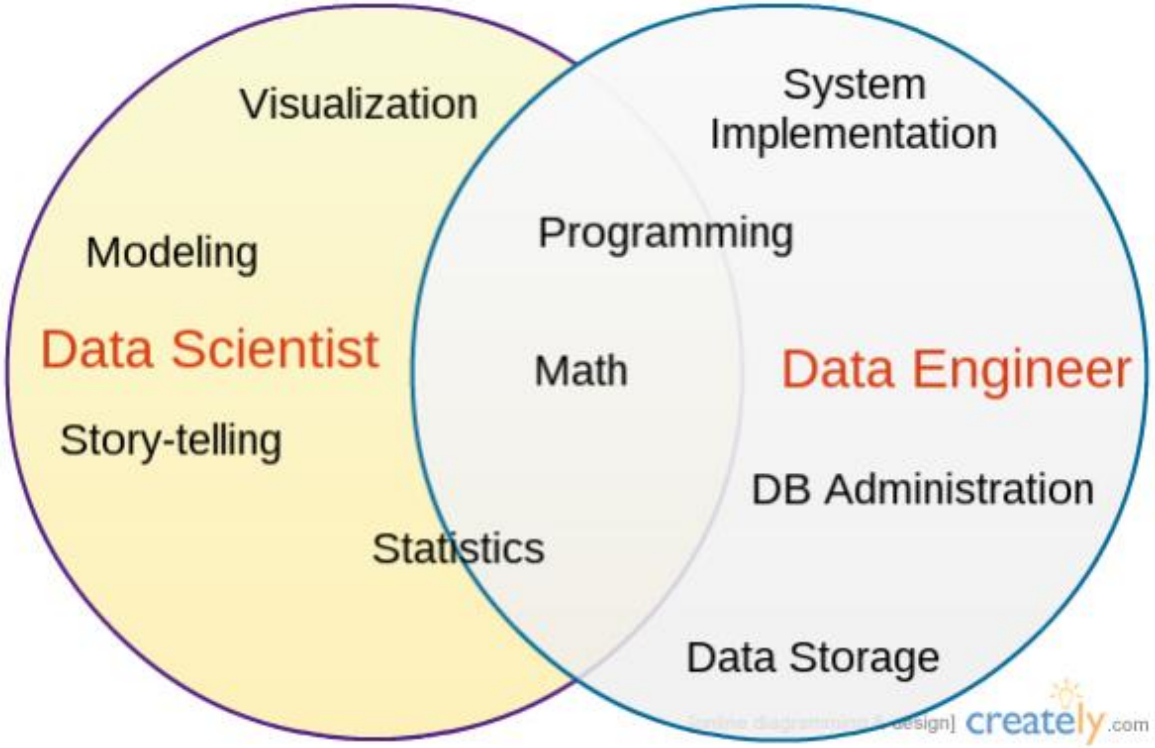


https://namitkabra.wordpress.com/2016/12/05/the-4-personas-for-data-analytics/

The 4 Personas for Data Analytics



https://namitkabra.wordpress.com/2016/12/05/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

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.)

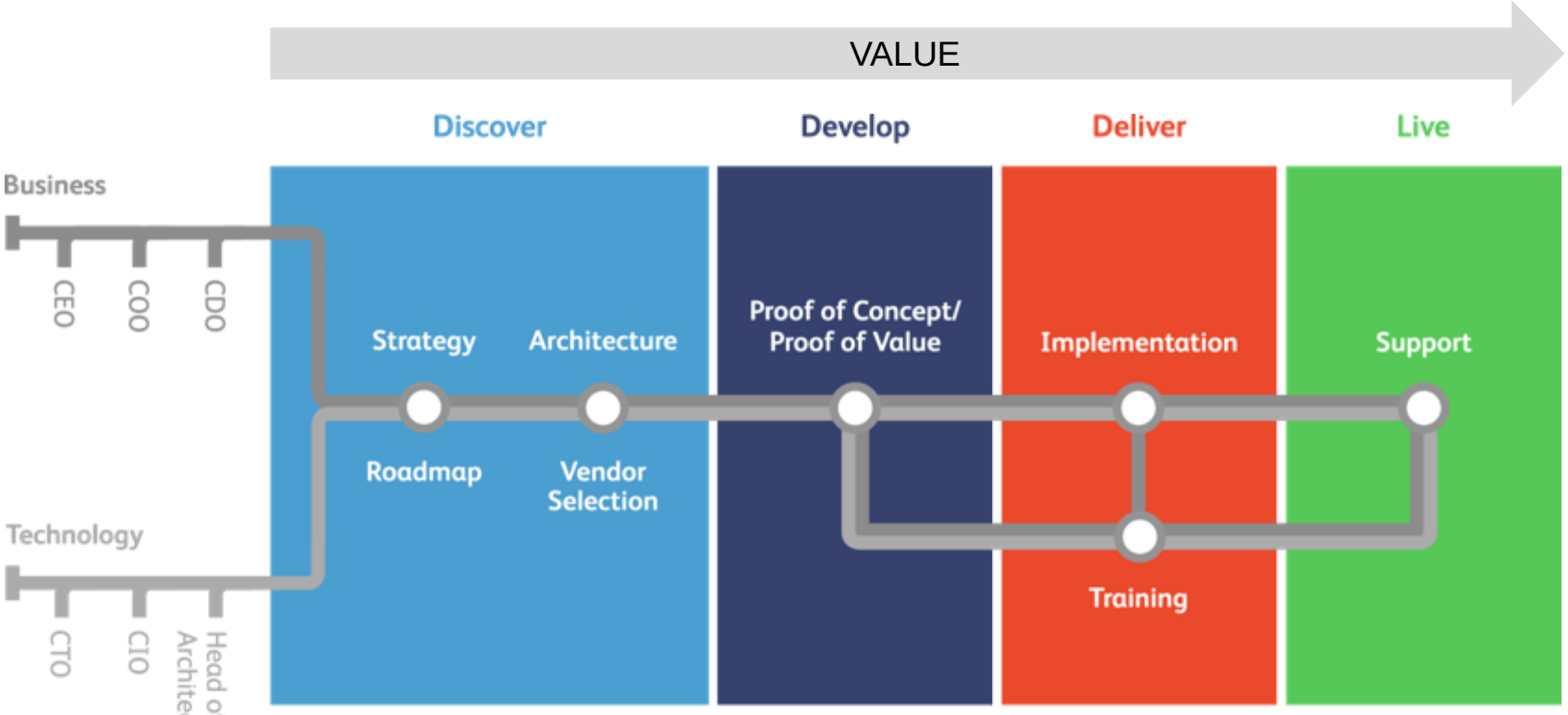
BIG DATA SANDBOX

- 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



- Understand big data
- Ensure management buy-in
- Create a multi-disciplinary team
- Define a list of use cases
- Explore Data Landscape
- Solution Architecture Design (Vendor)
- Initial Investment on Infrastructure (rent)

- Select proof of concepts
- Collect data
- Create use case prototypes

- Share the results
- Define the ROI
- Invest in infrastructure
- Expand the team and start implementation
- Organization wide training
- Improve overall organization