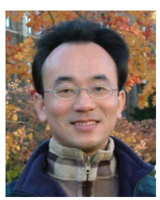
Tamkang University

Big Data Mining

ABC:

AI, Big Data, Cloud Computing

1071BDM02
TLVXM1A (M2244) (8619) (Fall 2018)
(MBA, DBETKU) (3 Credits, Required) [Full English Course]
(Master's Program in Digital Business and Economics)
Mon, 9, 10, 11, (16:10-19:00) (B206)



Min-Yuh Day, Ph.D.
Assistant Professor
Department of Information Management
Tamkang University

http://mail.tku.edu.tw/myday

2018-09-17

Course Schedule (1/2)

Tamkang University

Week	Date	Subject/Topics
1	2018/09/10	Course Orientation for Big Data Mining
2	2018/09/17	ABC: AI, Big Data, Cloud Computing
3	2018/09/24	Mid-Autumn Festival (Day off)
4	2018/10/01	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data
5	2018/10/08	Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem
6	2018/10/15	Foundations of Big Data Mining in Python
7	2018/10/22	Supervised Learning: Classification and Prediction
8	2018/10/29	Unsupervised Learning: Cluster Analysis
9	2018/11/05	Unsupervised Learning: Association Analysis

Course Schedule (2/2)

Tamkang University

Week	Date	Subject/Topics
10	2018/11/12	Midterm Project Report
11	2018/11/19	Machine Learning with Scikit-Learn in Python
12	2018/11/26	Deep Learning for Finance Big Data with TensorFlow
13	2018/12/03	Convolutional Neural Networks (CNN)
14	2018/12/10	Recurrent Neural Networks (RNN)
15	2018/12/17	Reinforcement Learning (RL)
16	2018/12/24	Social Network Analysis (SNA)
17	2018/12/31	Bridge Holiday (Extra Day Off)
18	2019/01/07	Final Project Presentation

ABC:

AI,

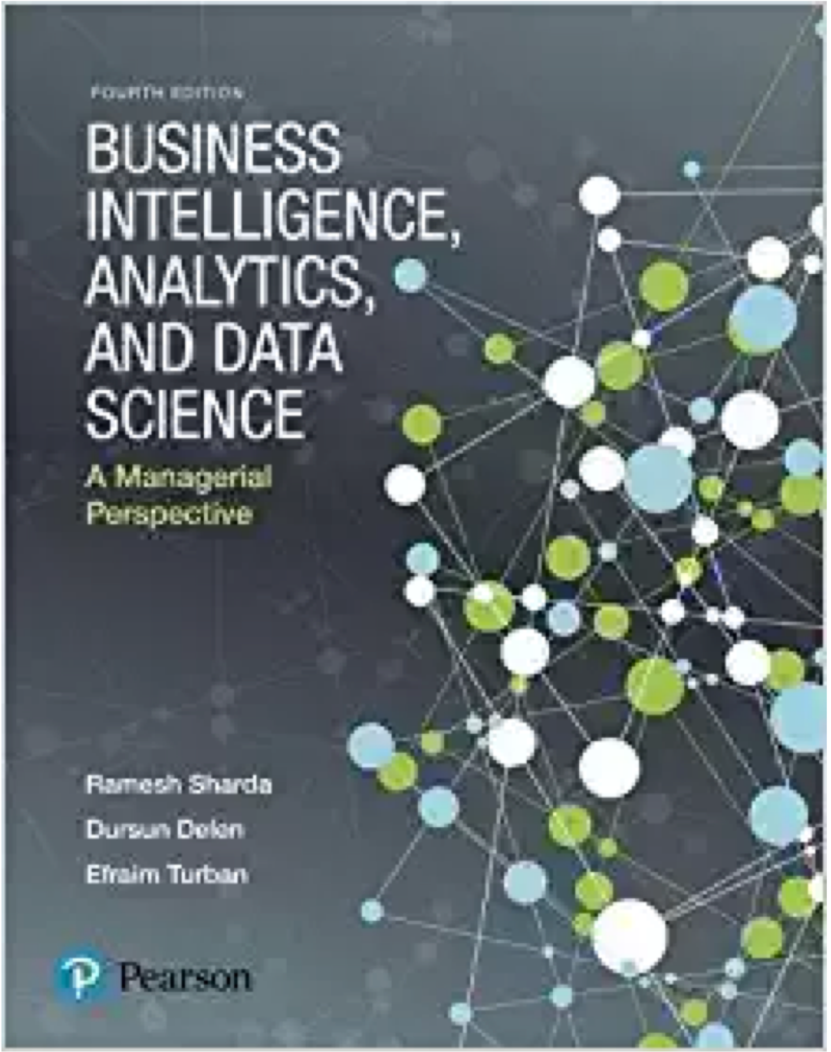
Big Data,

Cloud Computing

Outline

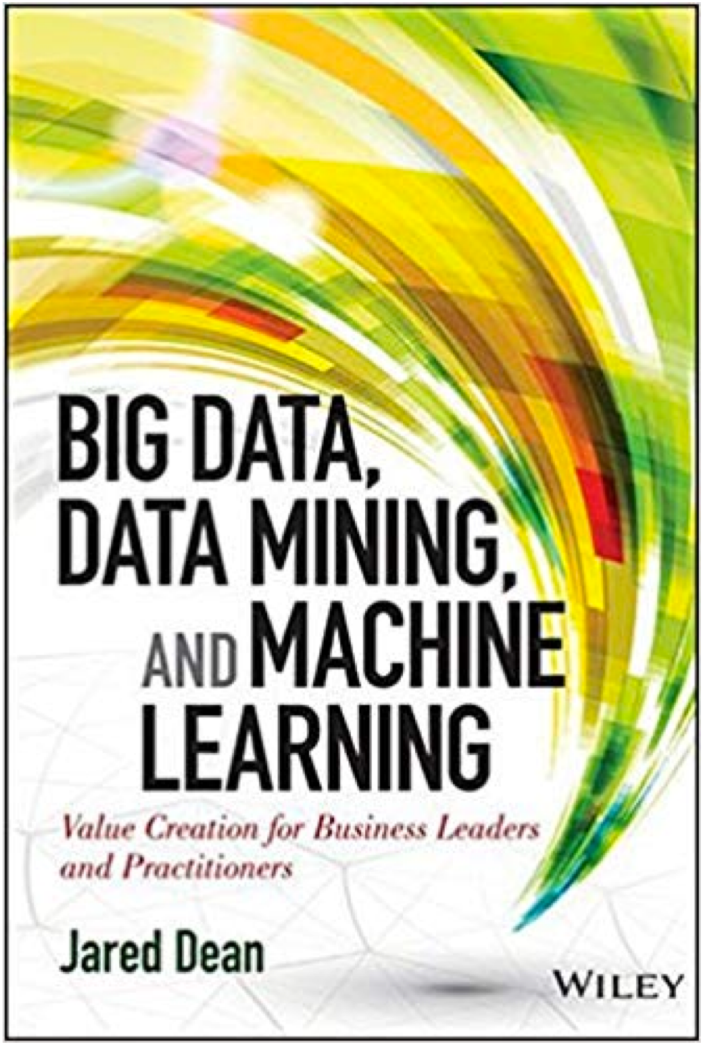
- AI
- Big Data
- Cloud Computing

Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Ramesh Sharda, Dursun Delen, and Efraim Turban, Pearson, 2017.



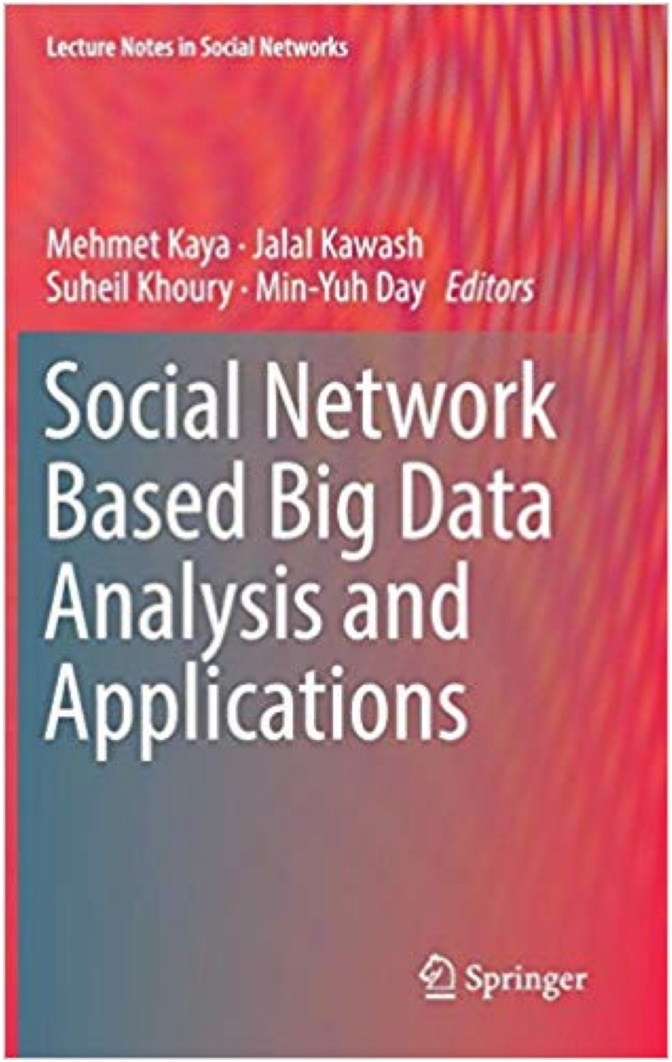
Source: https://www.amazon.com/Business-Intelligence-Analytics-Data-Science/dp/0134633288

Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, Jared Dean, Wiley, 2014.



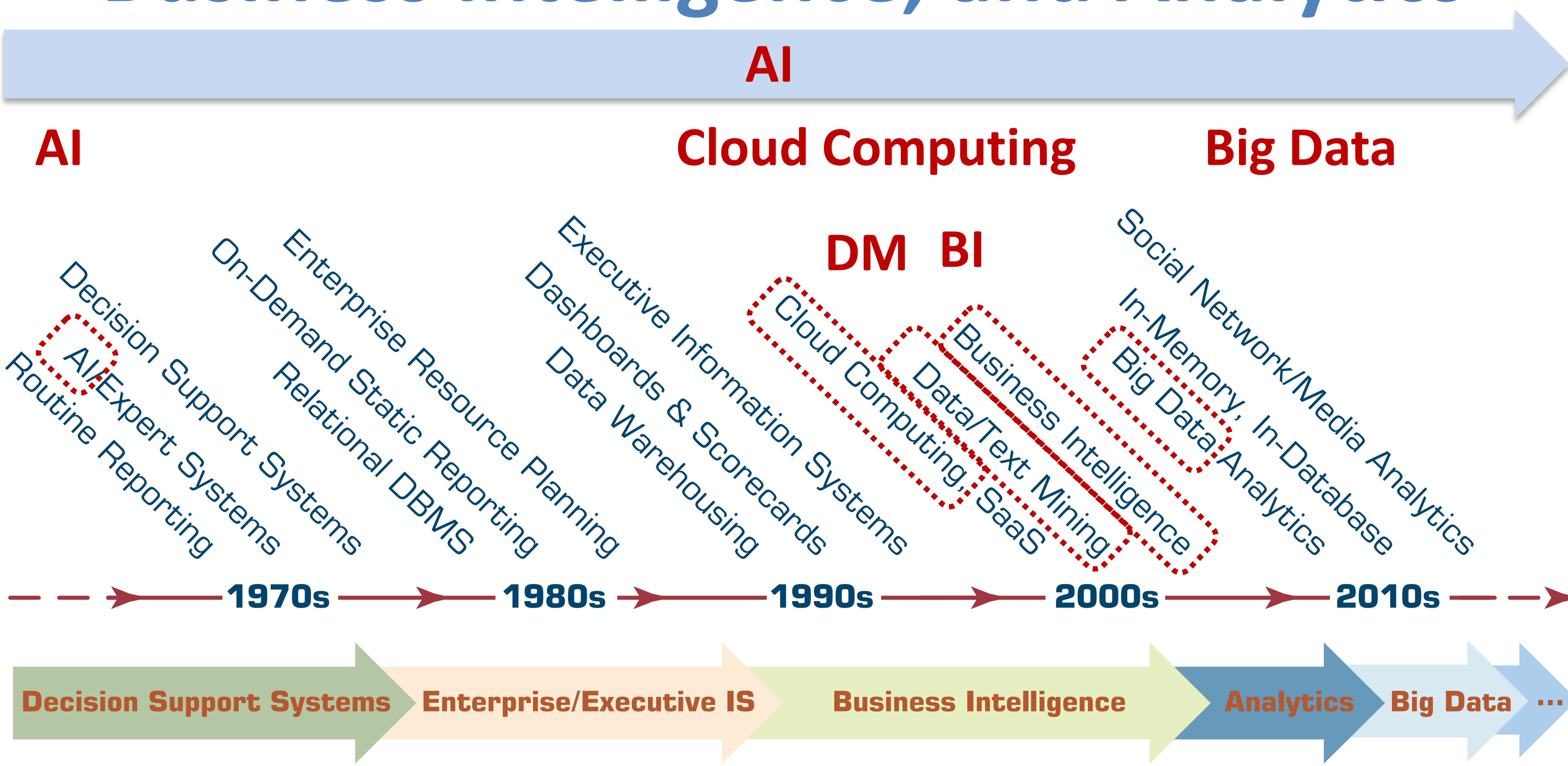
Source: https://www.amazon.com/Data-Mining-Machine-Learning-Practitioners/dp/1118618041

Social Network Based Big Data Analysis and Applications, Lecture Notes in Social Networks, Mehmet Kaya, Jalal Kawash, Suheil Khoury, Min-Yuh Day, Springer International Publishing, 2018.



Source: https://www.amazon.com/Network-Analysis-Applications-Lecture-Networks/dp/3319781952

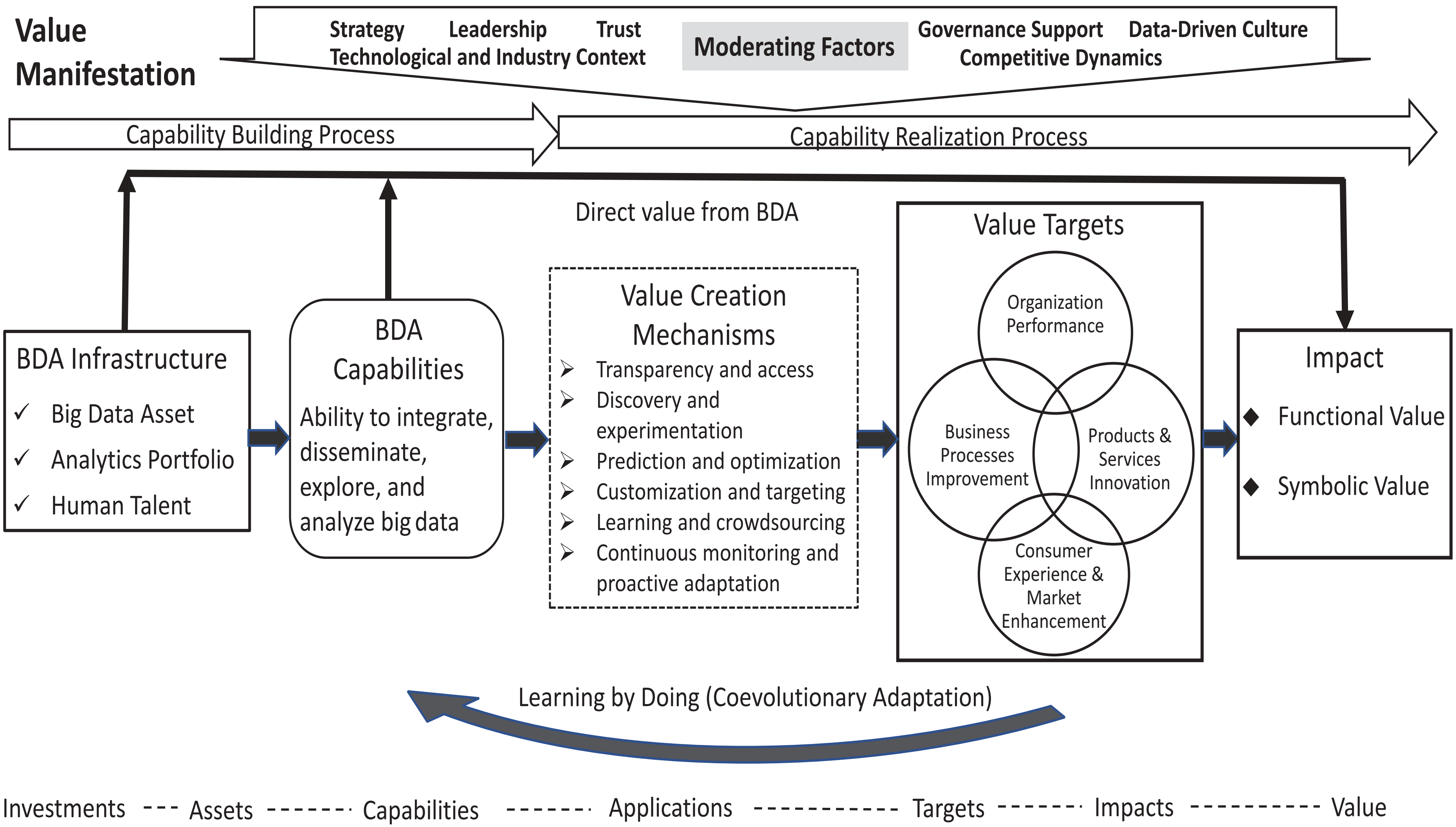
AI, Big Data, Cloud Computing
Evolution of Decision Support, Business Intelligence, and Analytics



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Value Creation by Big Data Analytics

(Grover et al., 2018)



Source: Varun Grover, Roger HL Chiang, Ting-Peng Liang, and Dongsong Zhang (2018), "Creating Strategic Business Value from Big Data Analytics: A Research Framework", Journal of Management Information Systems, 35, no. 2, pp. 388-423.

Research Landscape of Business Intelligence and Big Data Analytics: A bibliometrics study

- A bibliometric analysis on Big Data and Business Intelligence from 1990 to 2016.
- Big Data papers grow much faster than Business Intelligence papers
- Computer Science and information systems are two core disciplines.
- Most influential papers are identified and a research framework is proposed.

Source: Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, Volume 111, 30, 2018, pp. 2-10

Evolution of top keywords in "BD & BI" publications

2014
- Management
- Text Mining
- Data Mining
- Data Science

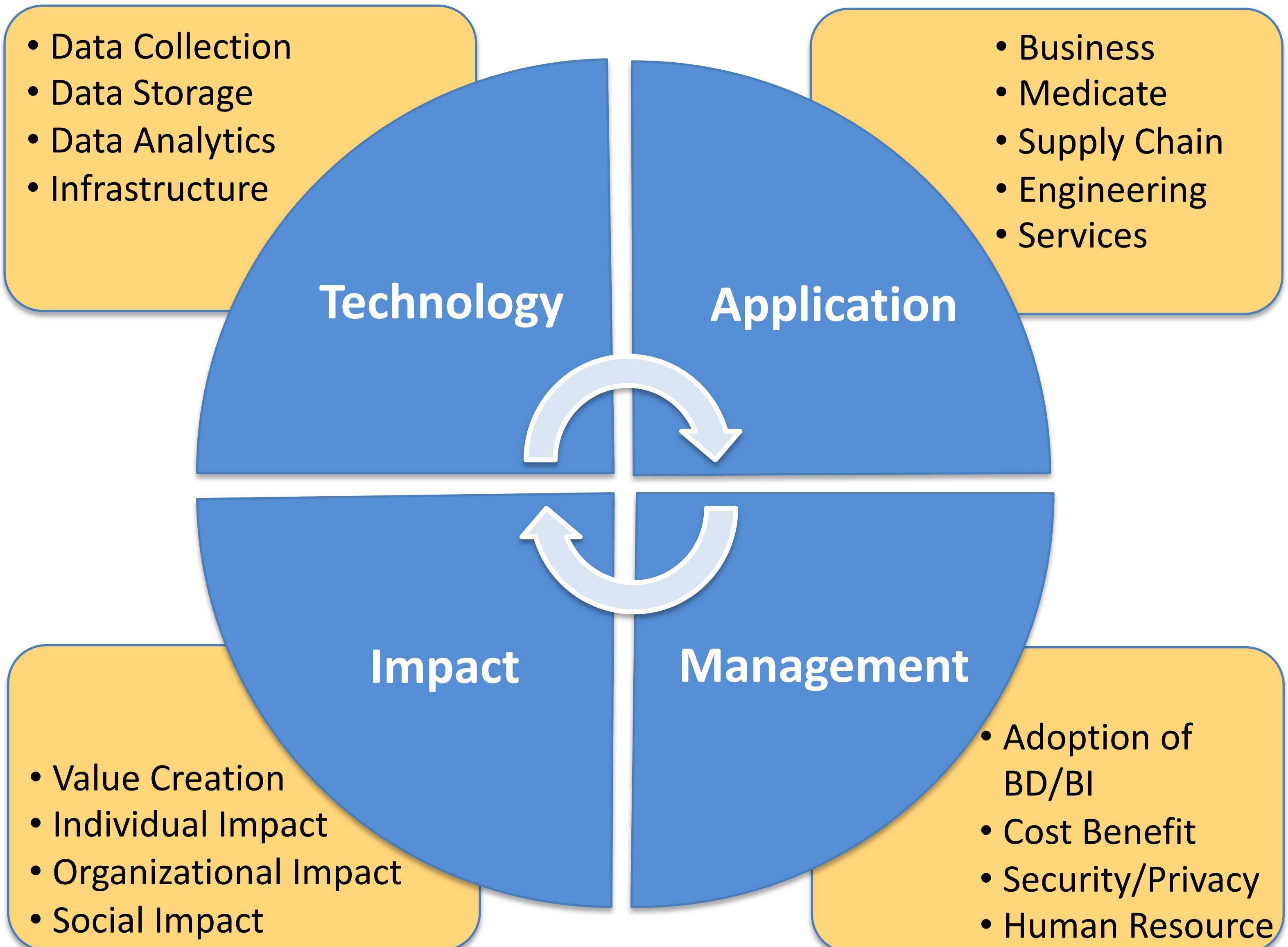
2015
- Big Data Analytics
- Social Media
- Business Analytics
- Information System

2016
- Cloud Computing
- Data Warehouse

2017
- Knowledge Management

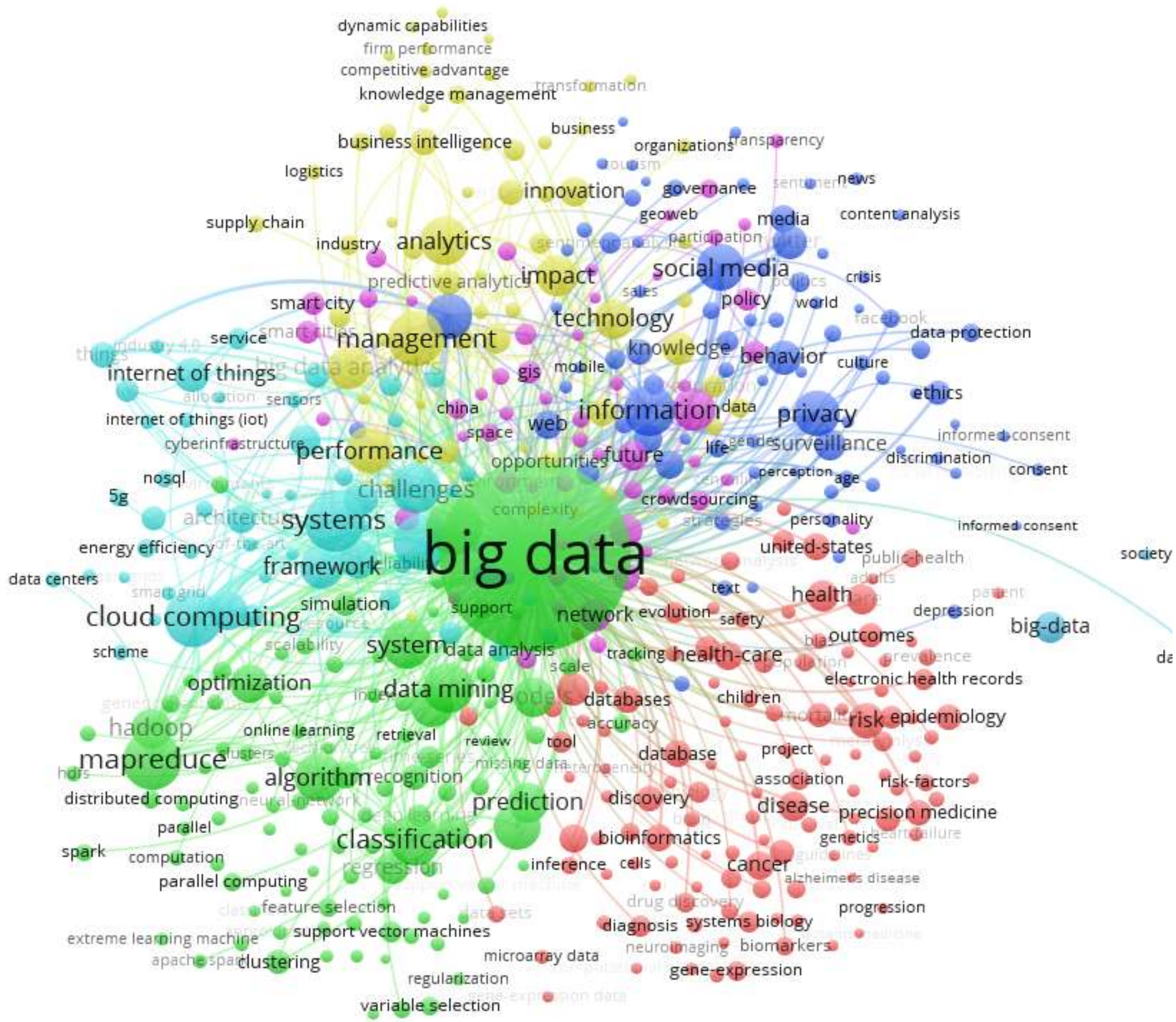
Source: Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, Volume 111, 30, 2018, pp. 2-10

Framework for BD and BI Research



Source: Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, Volume 111, 30, 2018, pp. 2-10

Business Intelligence and Big Data analytics



Source: Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, Volume 111, 30, 2018, pp. 2-10

AI

Definition
of
Artificial Intelligence
(A.I.)

Artificial Intelligence

“… the science and

engineering

of

making

intelligent machines”

(John McCarthy, 1955)

Source: https://digitalintelligencetoday.com/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/

Artificial Intelligence

"… technology that thinks and acts like humans"

Source: https://digitalintelligencetoday.com/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/

Artificial Intelligence

“… intelligence exhibited by machines or software”

Source: https://digitalintelligencetoday.com/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/

4 Approaches of AI

Thinking Humanly	**Thinking Rationally**
Acting Humanly	**Acting Rationally**

Source: Stuart Russell and Peter Norvig (2016) , Artificial Intelligence: A Modern Approach, 3rd Edition, Pearson International

4 Approaches of AI

2. Thinking Humanly: The Cognitive Modeling Approach	**3. Thinking Rationally: The “Laws of Thought” Approach**
1. Acting Humanly: The Turing Test Approach (1950)	**4. Acting Rationally: The Rational Agent Approach**

Source: Stuart Russell and Peter Norvig (2016) , Artificial Intelligence: A Modern Approach, 3rd Edition, Pearson International

AI Acting Humanly: The Turing Test Approach

(Alan Turing, 1950)

- Natural Language Processing (NLP)
- Knowledge Representation
- Automated Reasoning
- Machine Learning (ML)
- Computer Vision
- Robotics

Source: Stuart Russell and Peter Norvig (2016) , Artificial Intelligence: A Modern Approach, 3rd Edition, Pearson International

Boston Dynamics: Atlas

#13 ON TRENDING

What's new, Atlas?

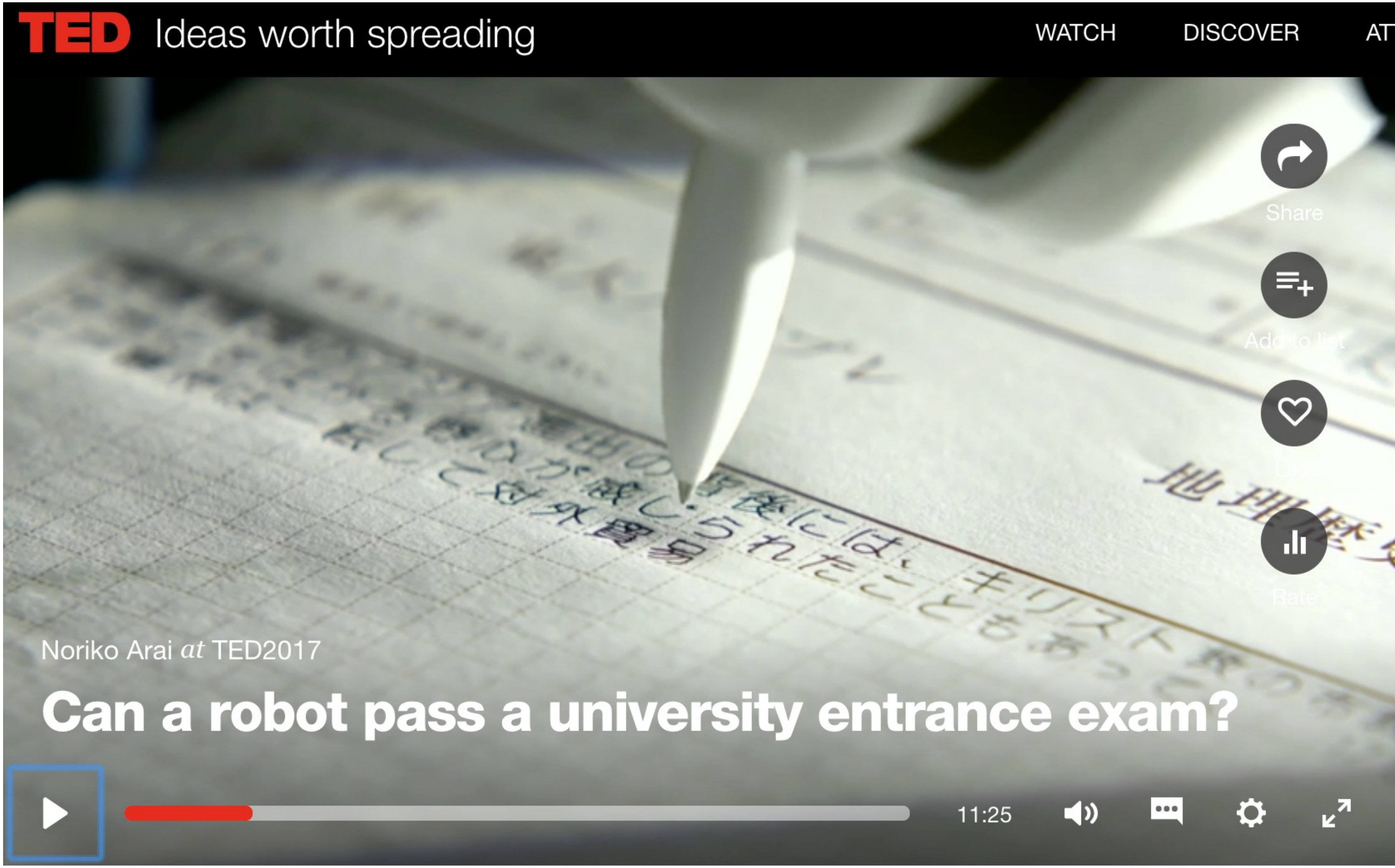
https://www.youtube.com/watch?v=fRj34o4hN4I

Humanoid Robot: Sophia

https://www.youtube.com/watch?v=S5t6K9iwcdw

Can a robot pass a university entrance exam?

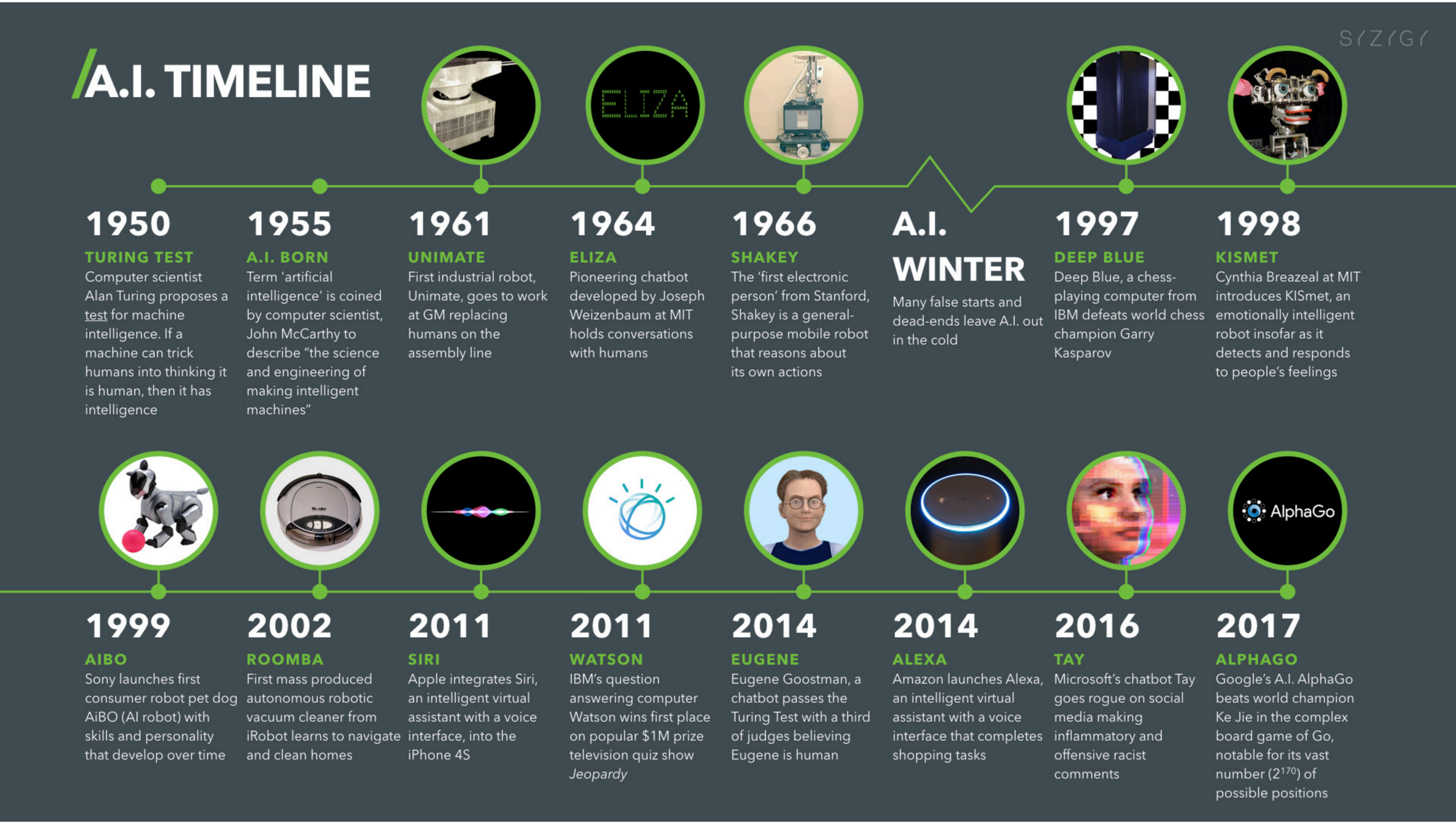
Noriko Arai at TED2017



https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam

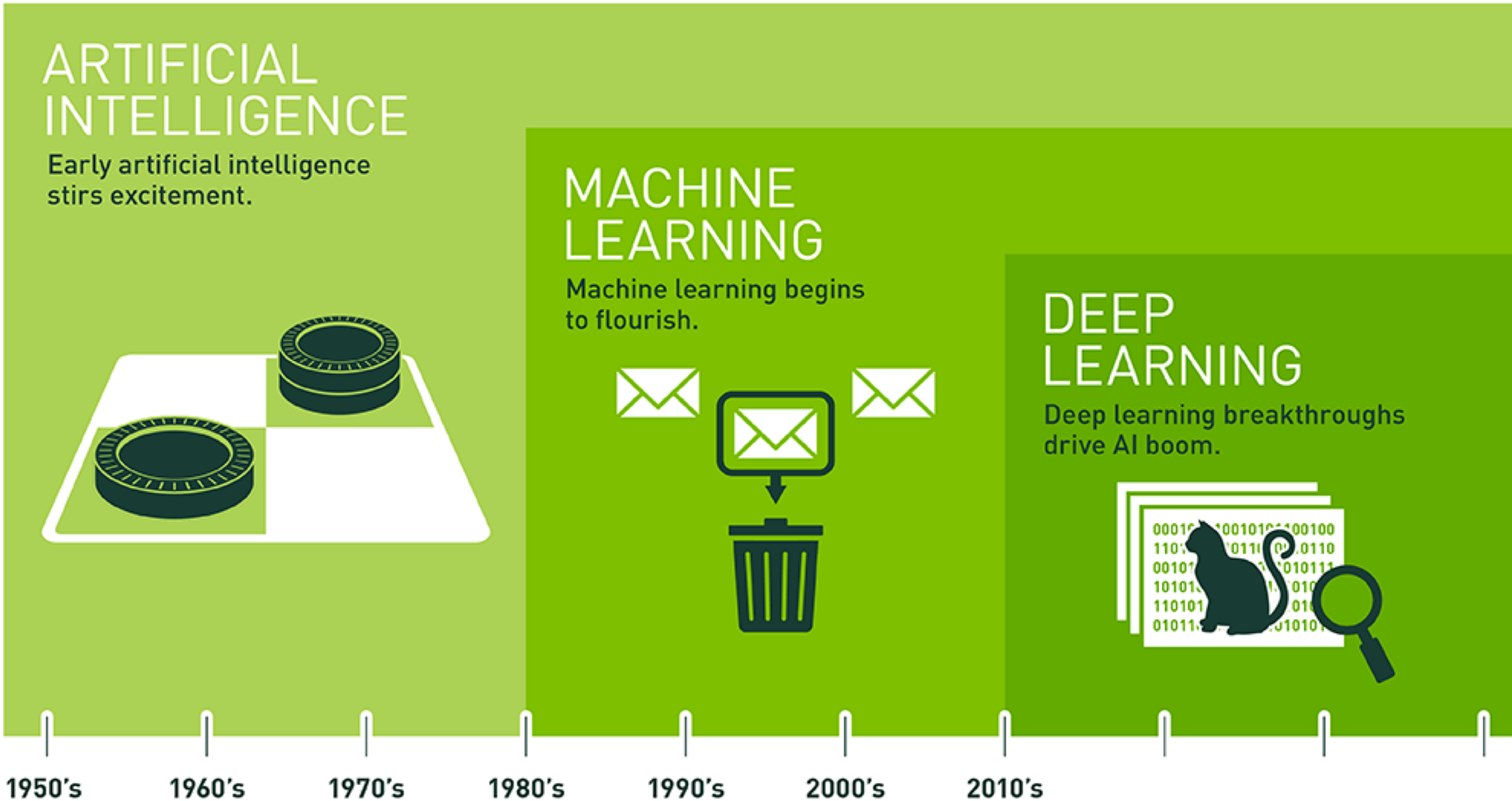
https://www.youtube.com/watch?v=XQZjkPyJ8KU

Artificial Intelligence (A.I.) Timeline



Source: https://digitalintelligencetoday.com/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/

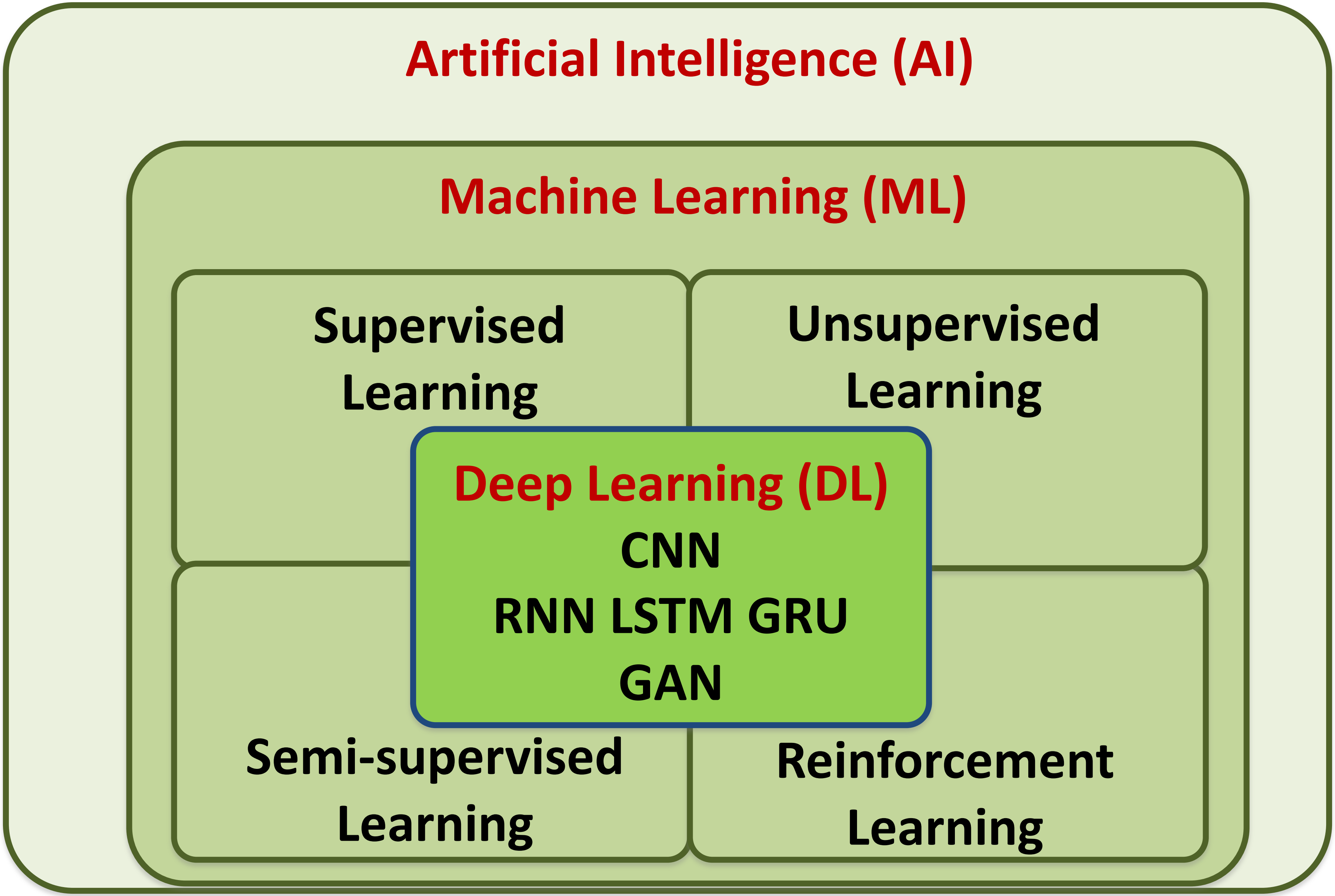
Artificial Intelligence
Machine Learning & Deep Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

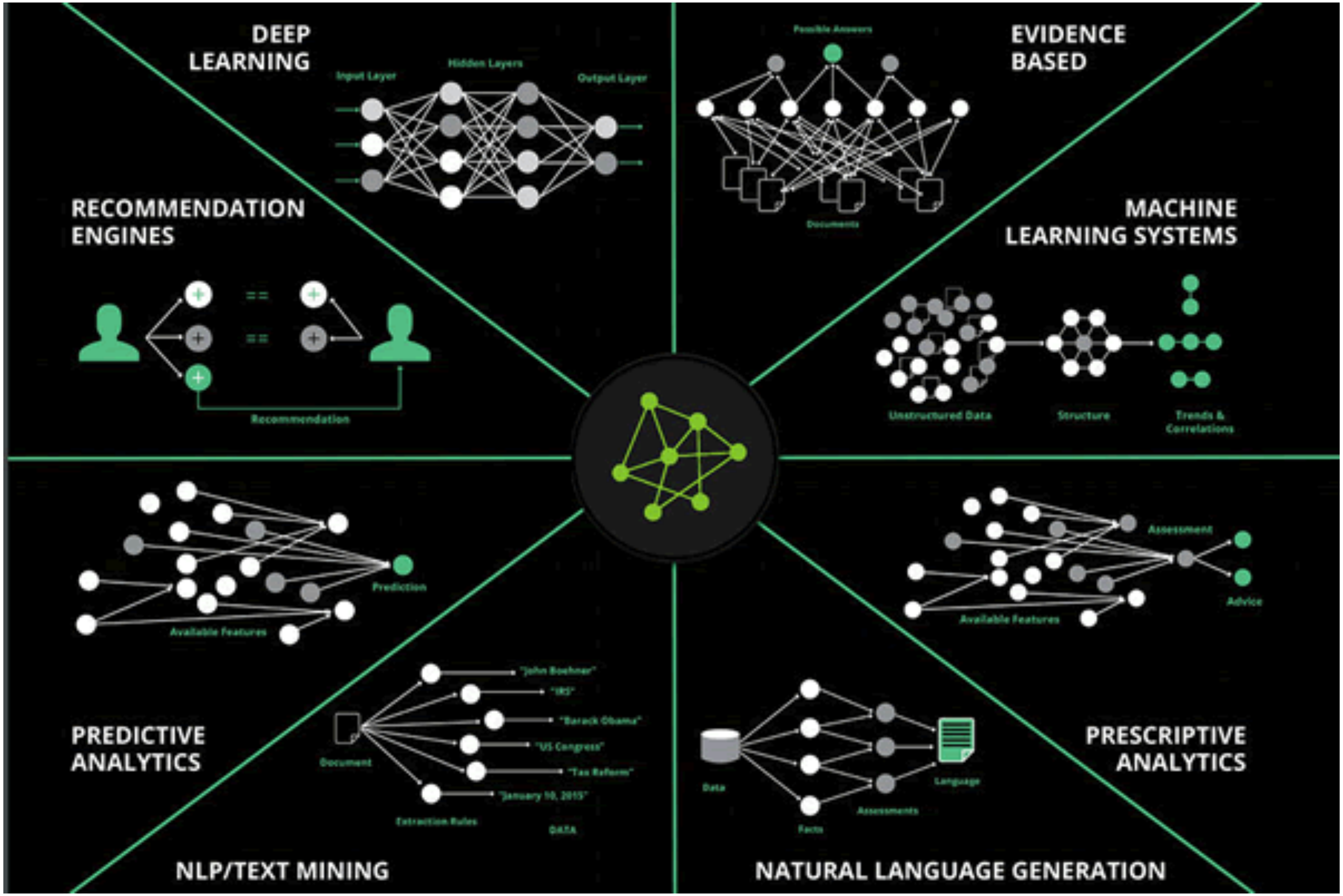
Source: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

AI, ML, DL



Source: https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/deep_learning.html

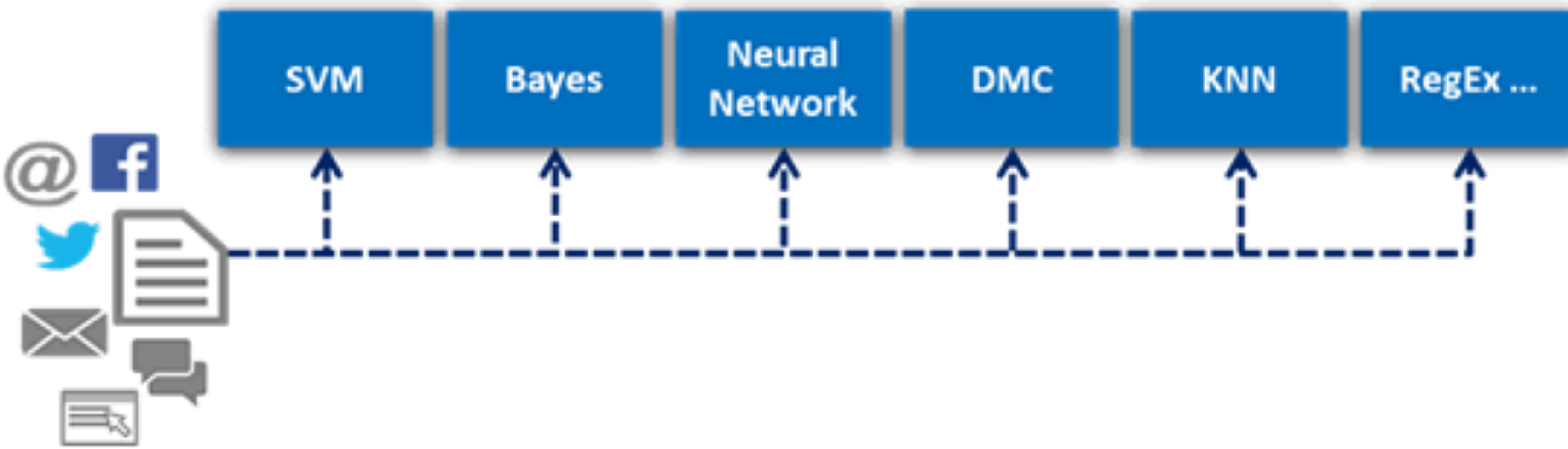
Artificial Intelligence (AI) is many things



Ecosystem of AI

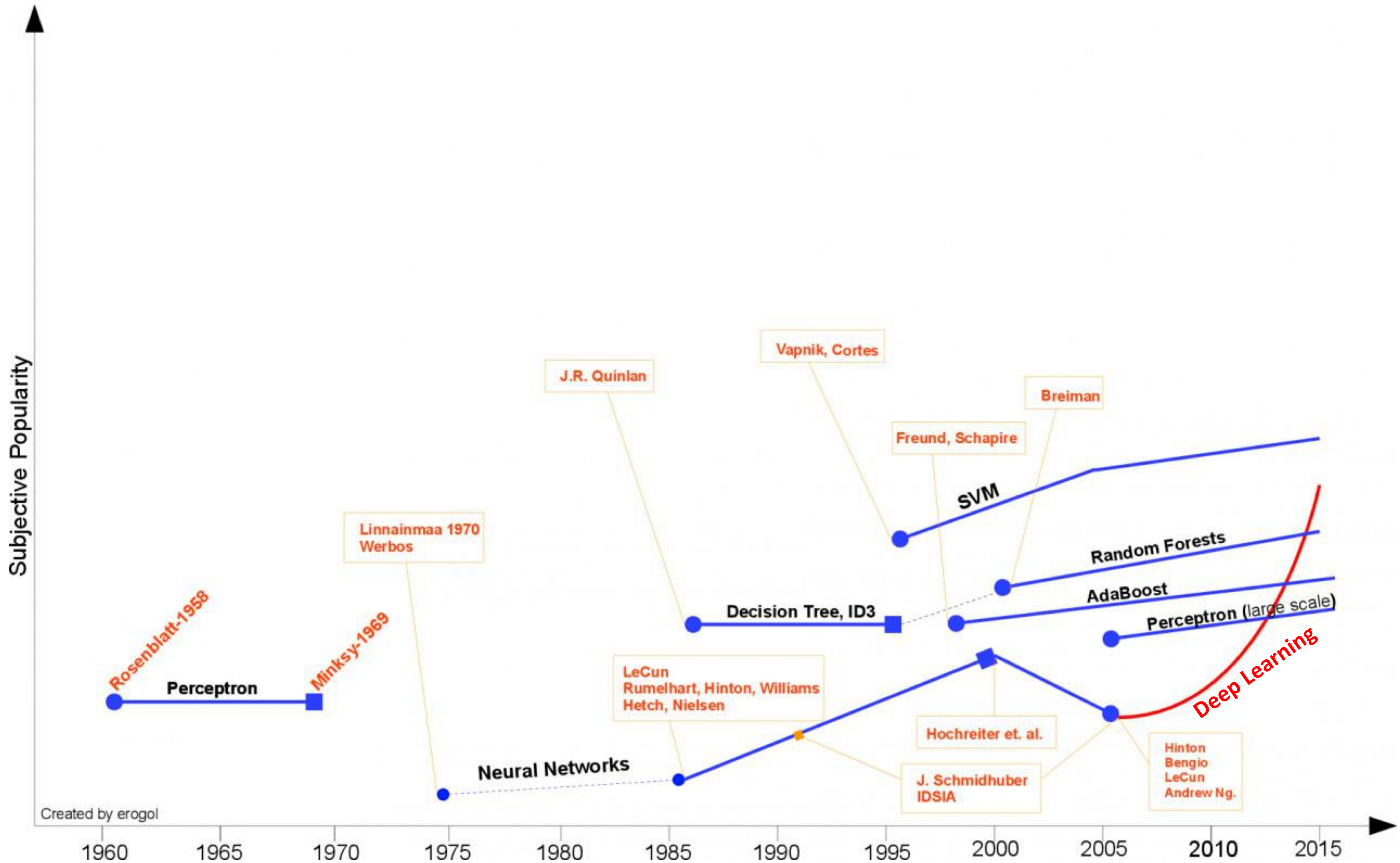
Source: https://www.i-scoop.eu/artificial-intelligence-cognitive-computing/

Artificial Intelligence (AI)
Intelligent Document Recognition algorithms



Source: https://www.i-scoop.eu/artificial-intelligence-cognitive-computing/

Deep Learning Evolution



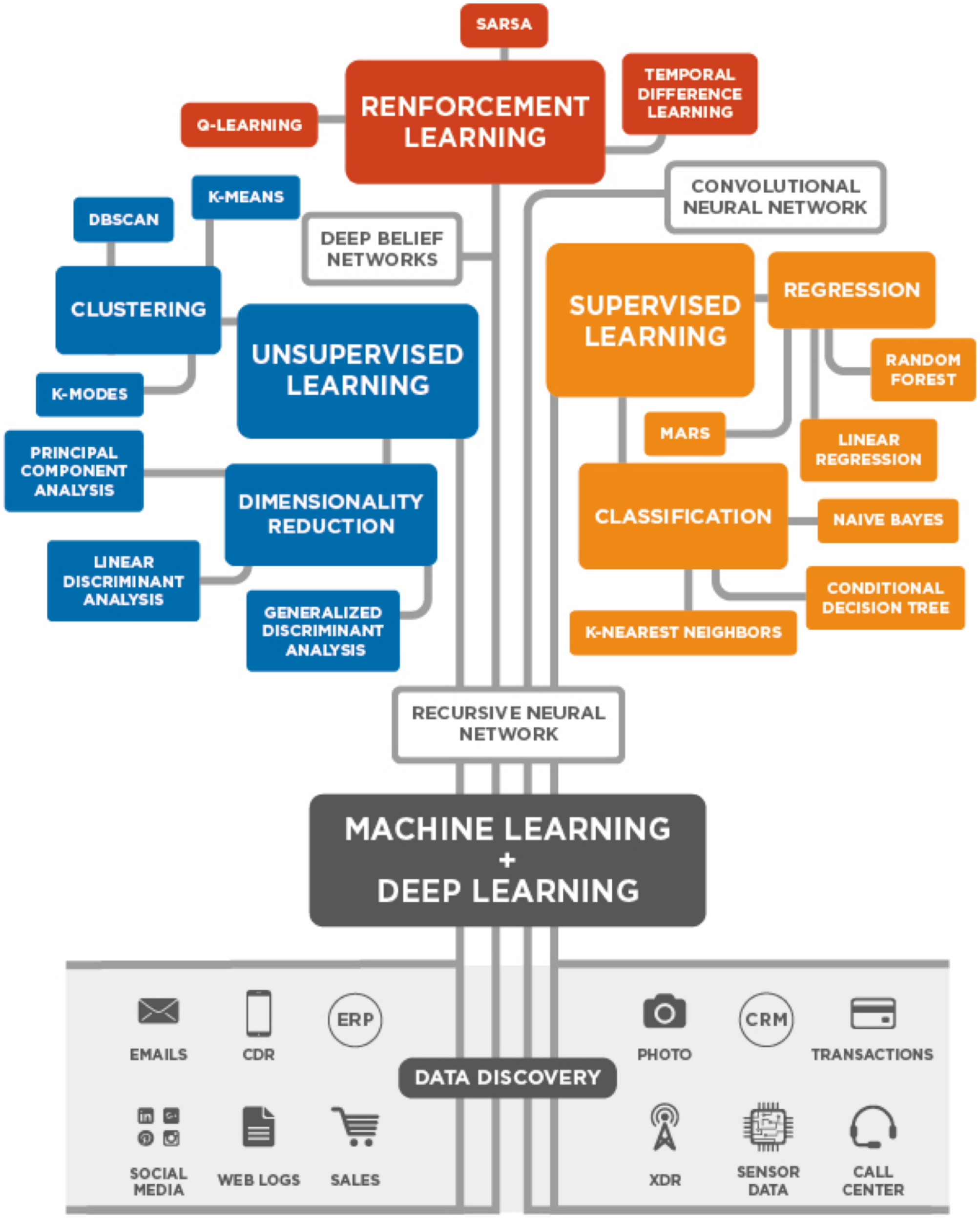
Source: http://www.erogol.com/brief-history-machine-learning/

Machine Learning Models

- Deep Learning
- Kernel
- Association rules
- Ensemble
- Decision tree
- Dimensionality reduction
- Clustering
- Regression Analysis
- Bayesian
- Instance based

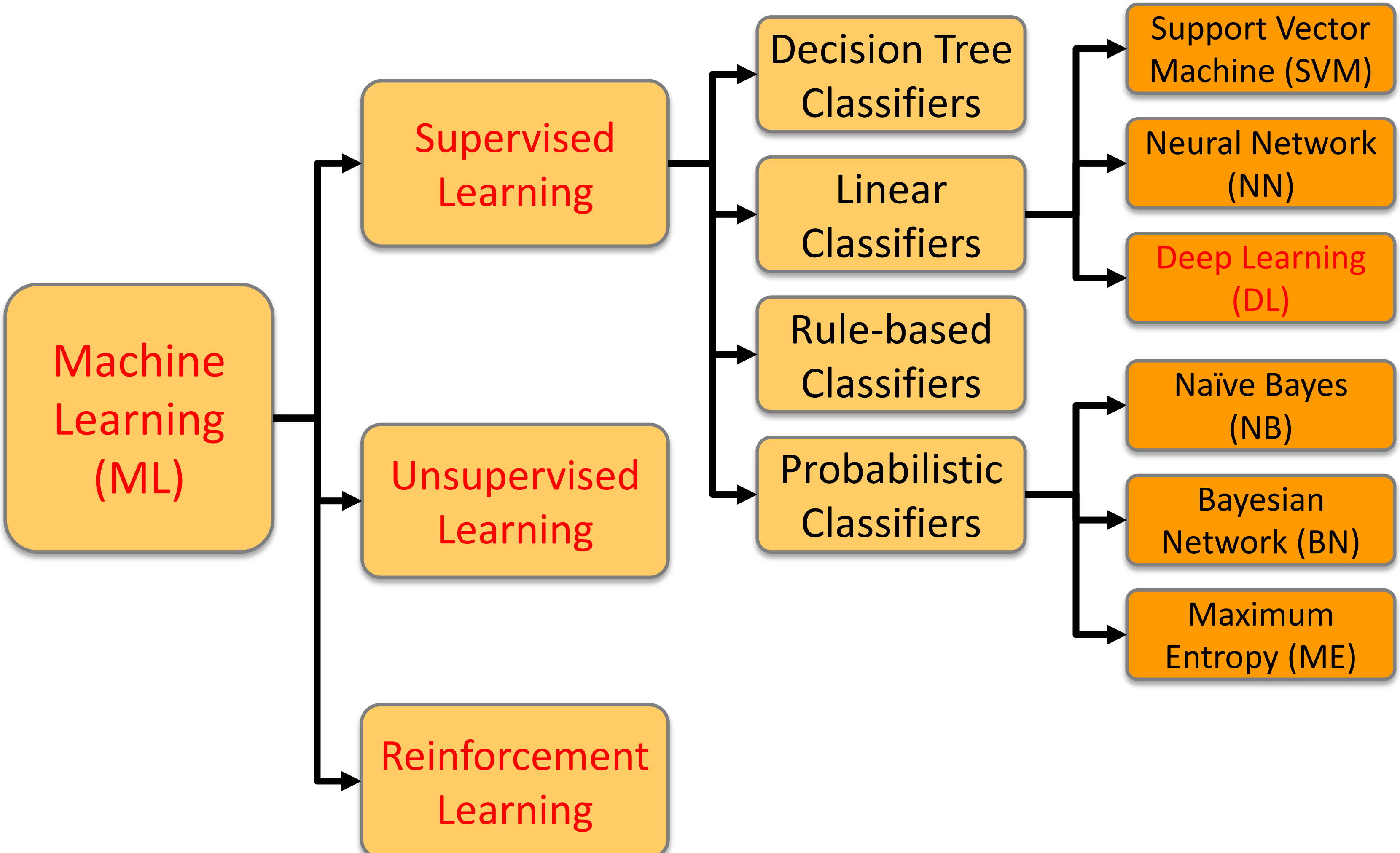
Source: Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing

3 Machine Learning Algorithms



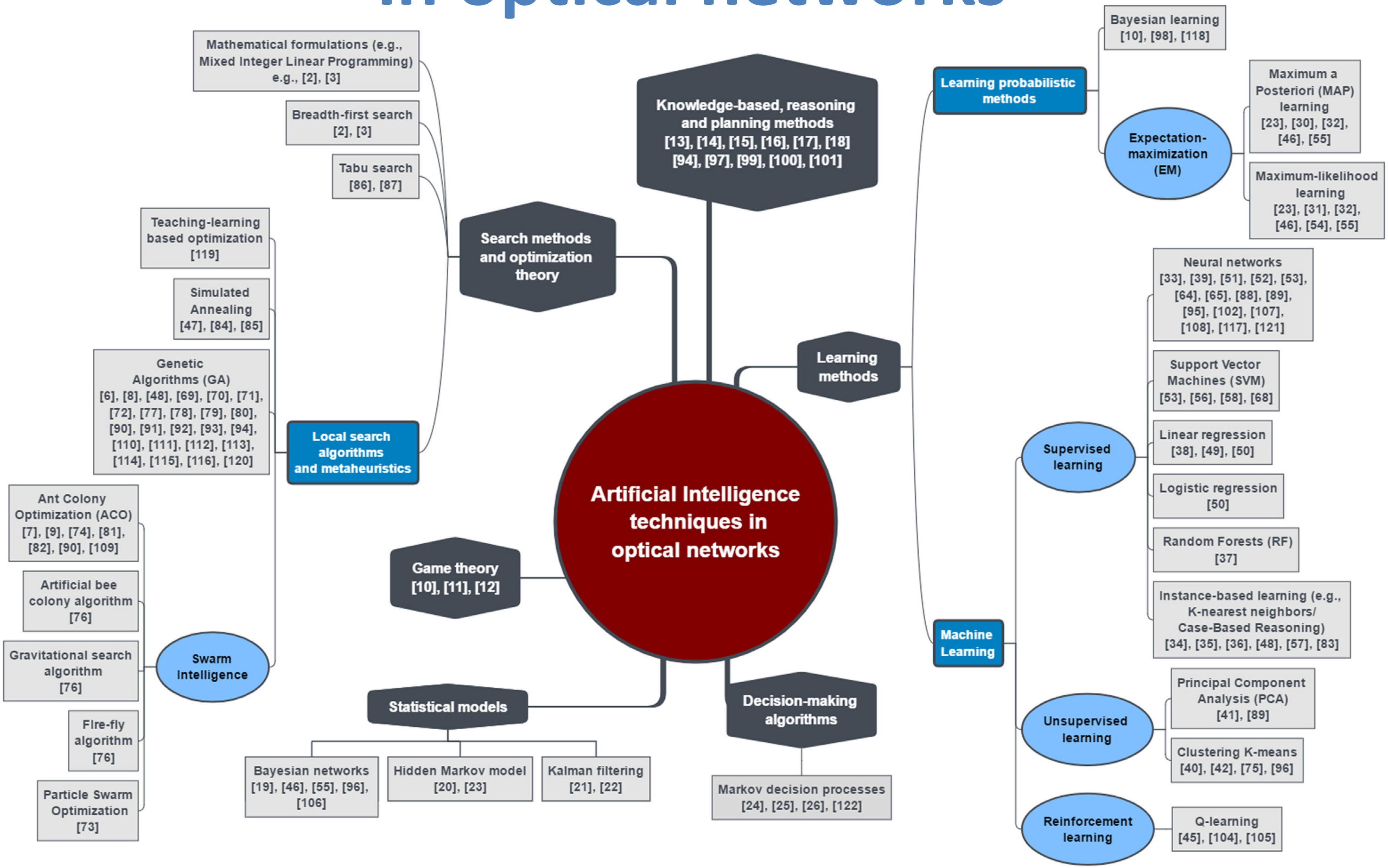
Source: Enrico Galimberti, http://blogs.teradata.com/data-points/tree-machine-learning-algorithms/

Machine Learning (ML) / Deep Learning (DL)



Source: Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.

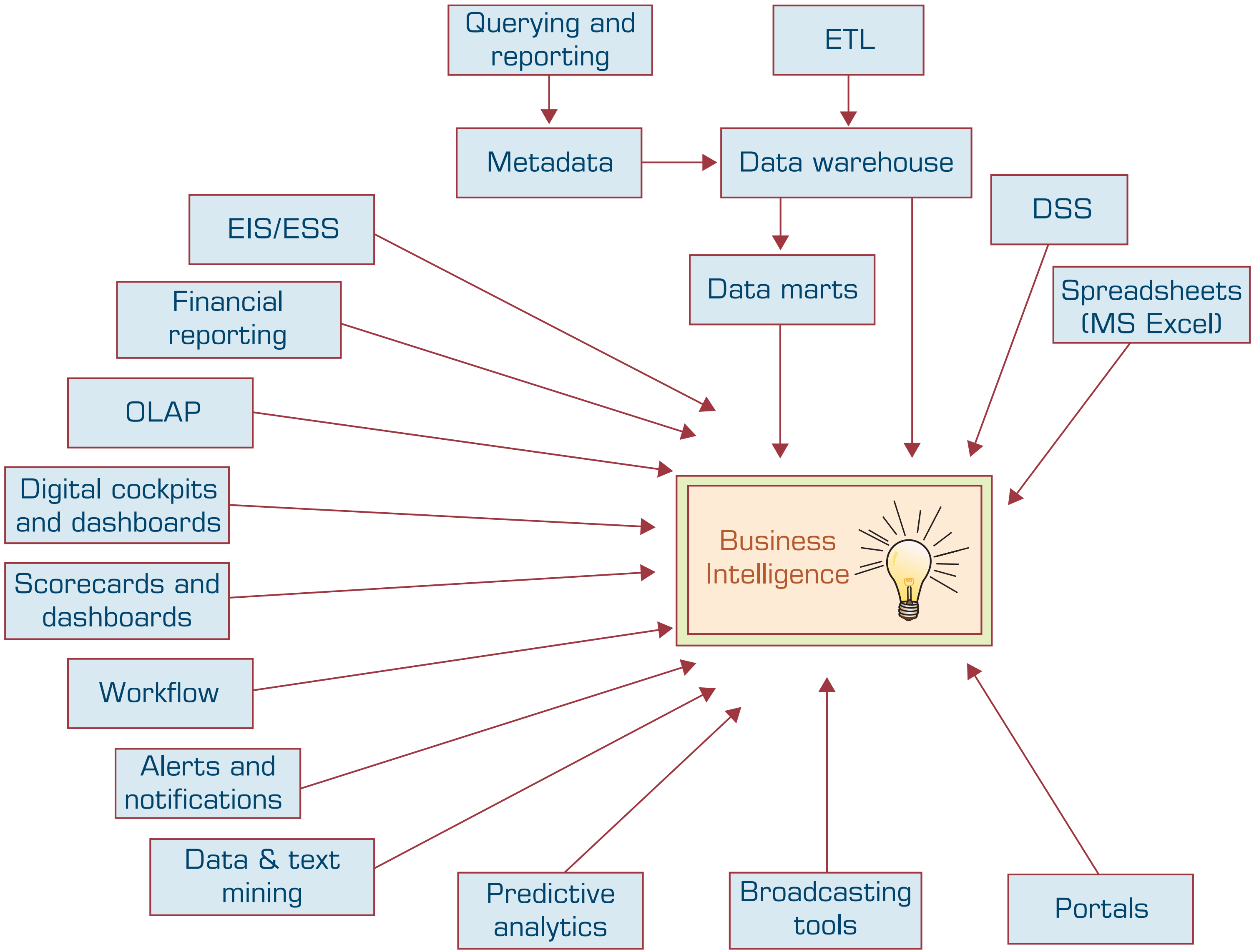
Artificial intelligence (AI) in optical networks



Source: Javier Mata, Ignacio de Miguel, Ramón J. Durán, Noemí Merayo, Sandeep Kumar Singh, Admela Jukan, and Mohit Chamania (2018), "Artificial intelligence (AI) methods in optical networks: A comprehensive survey", Optical Switching and Networking, 28, pp. 43-57

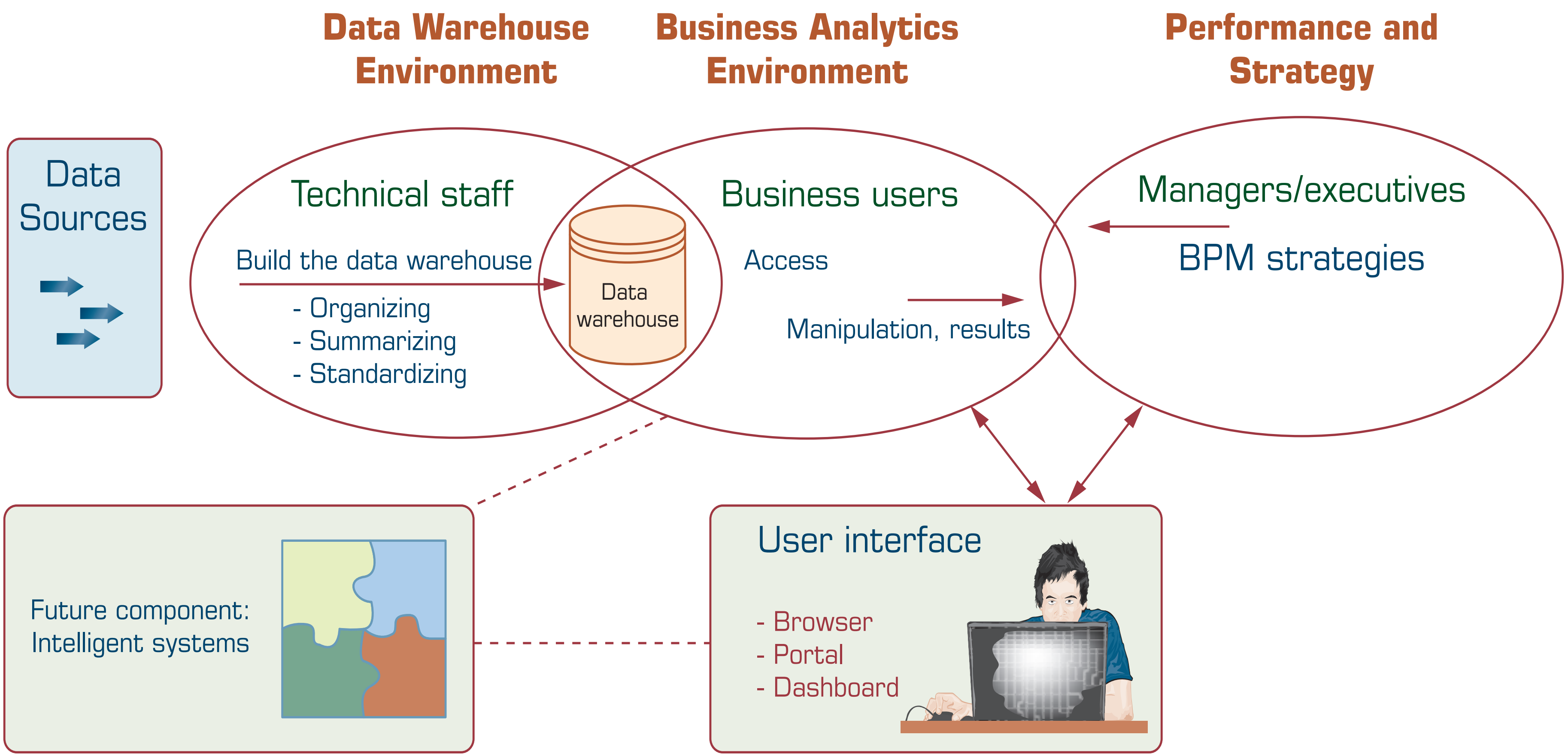
Big Data

Evolution of Business Intelligence (BI)



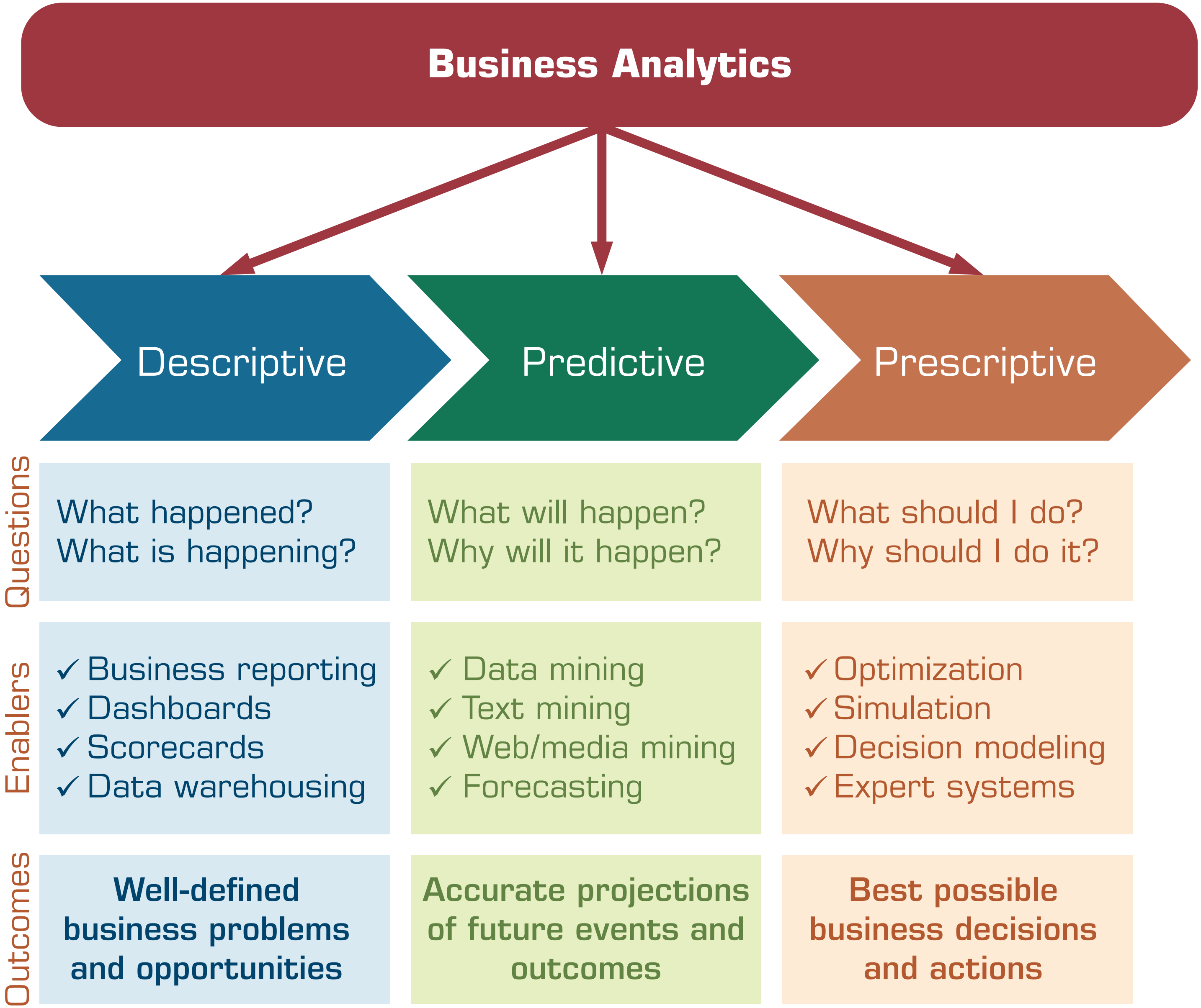
Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017),
Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

A High-Level Architecture of BI



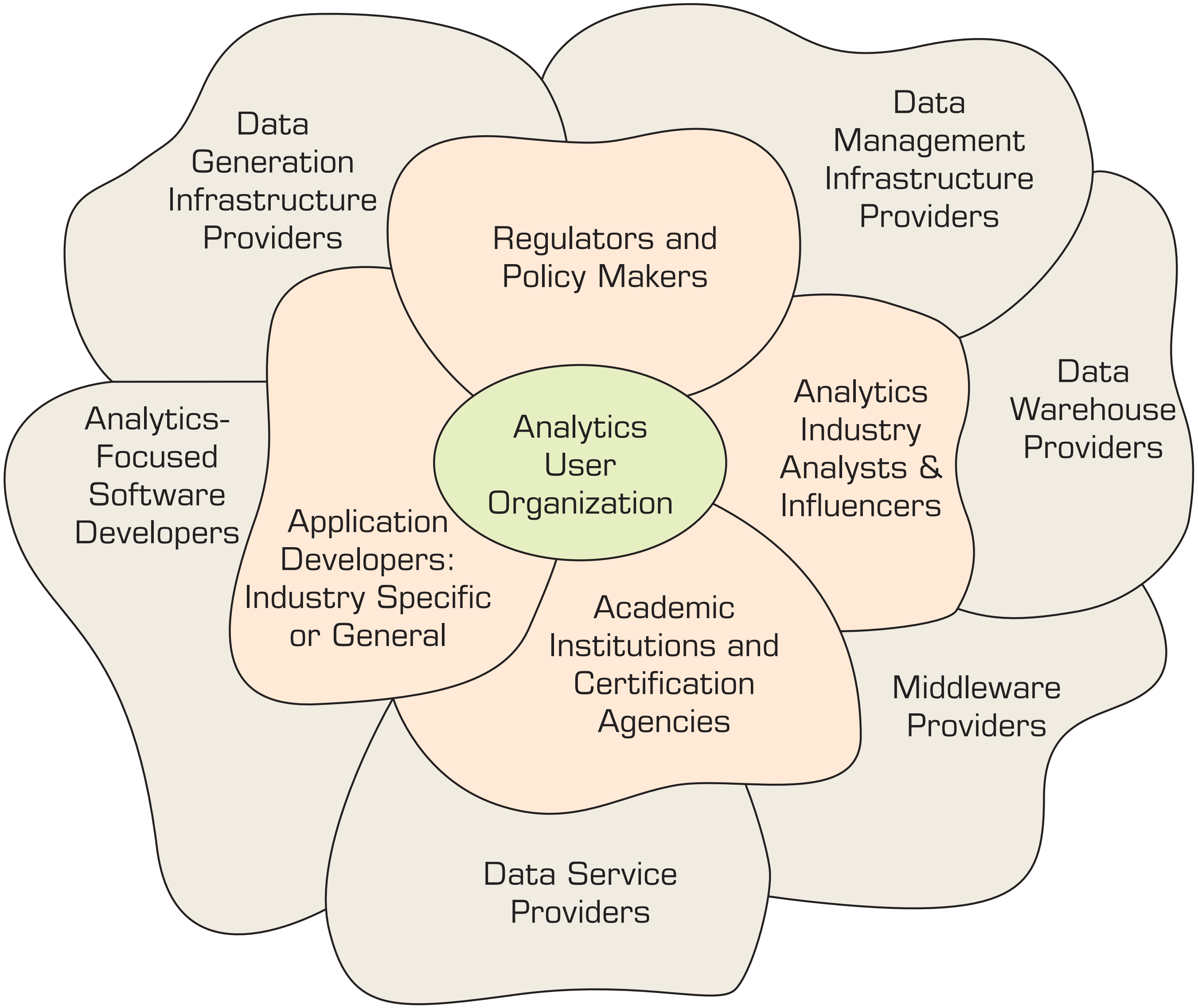
Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Three Types of Analytics



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017),
Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Analytics Ecosystem

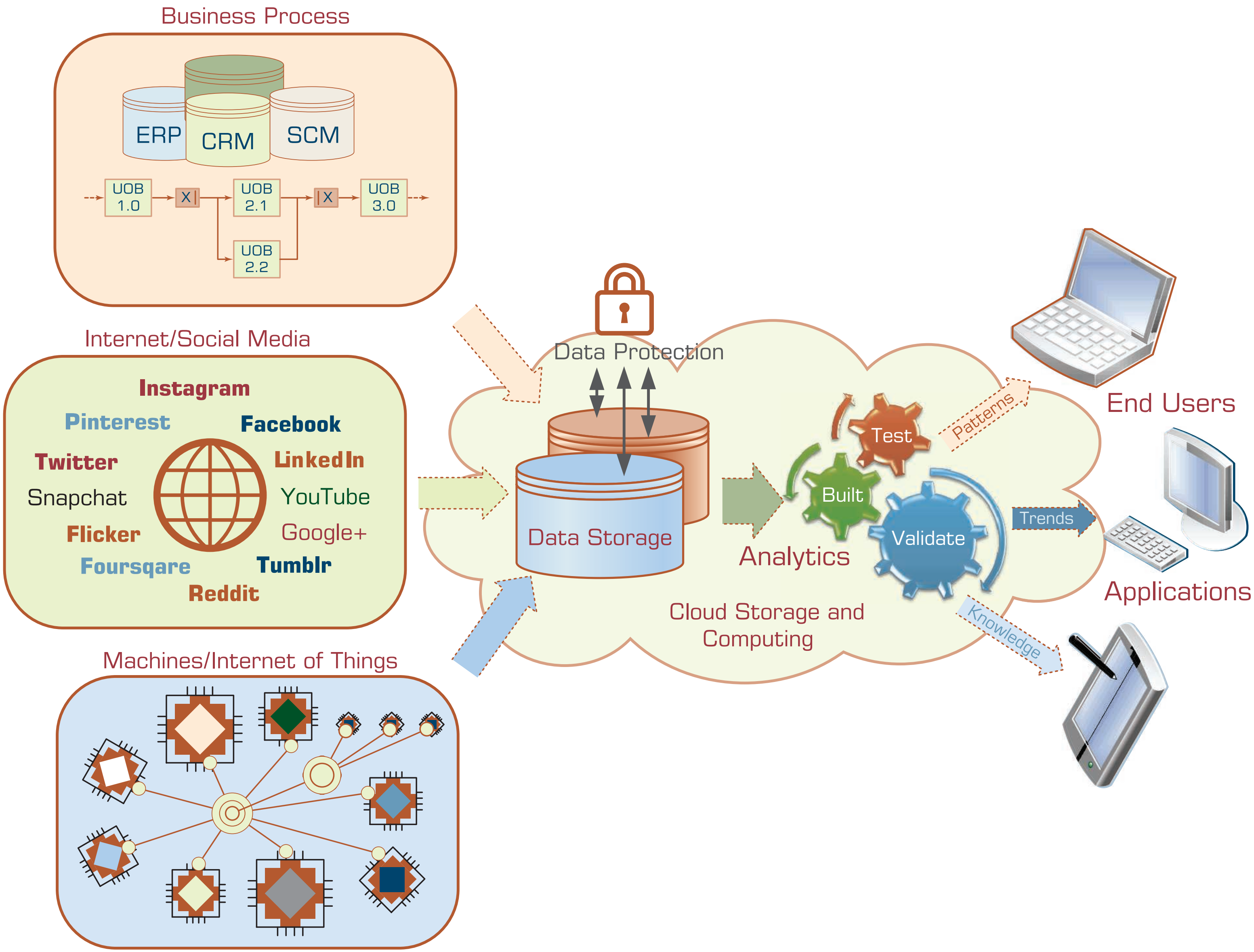


Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017),
Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Job Titles of Analytics

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017),
Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

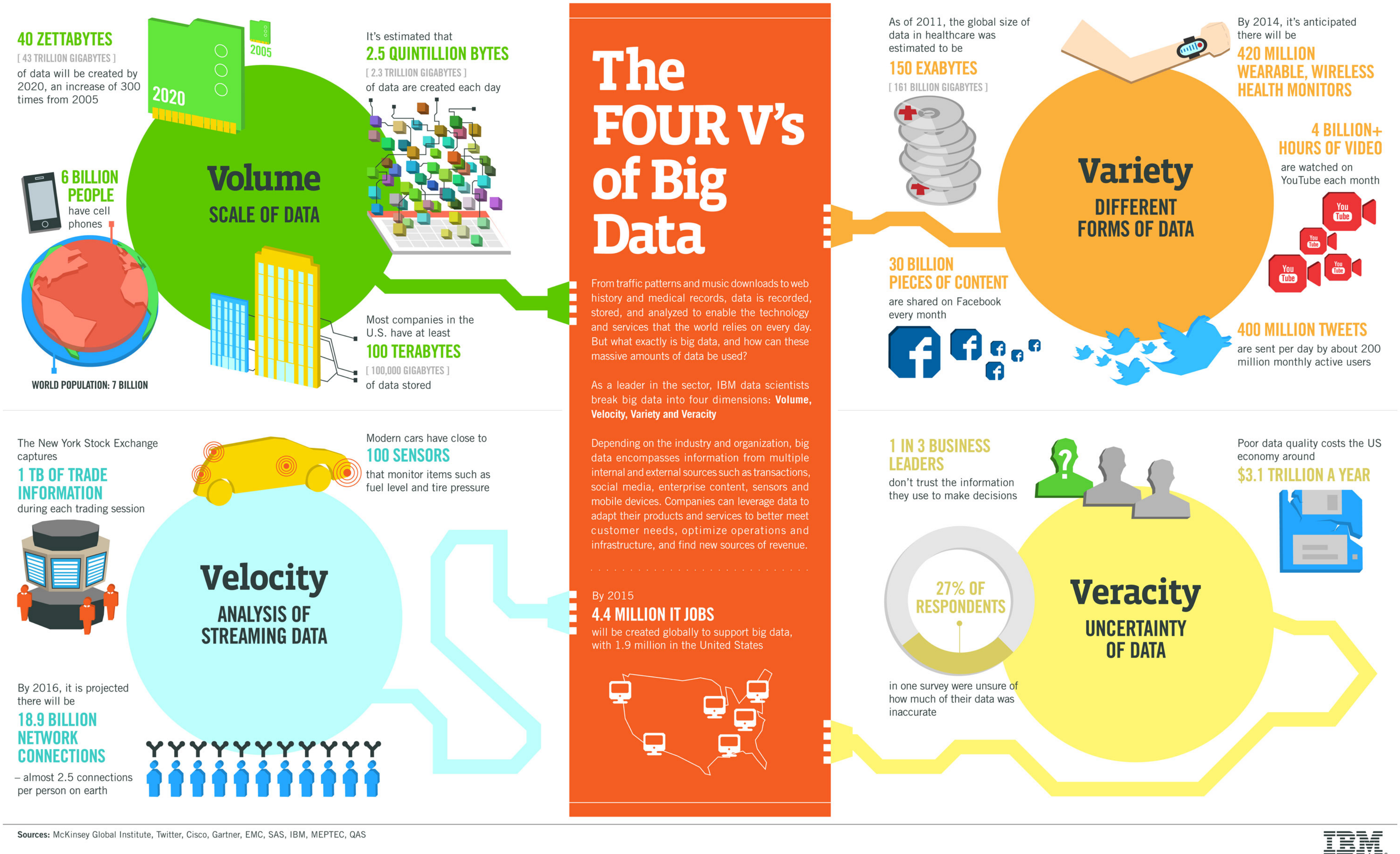
A Data to Knowledge Continuum



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017),
Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Big Data Analytics and Data Mining

Big Data 4 V



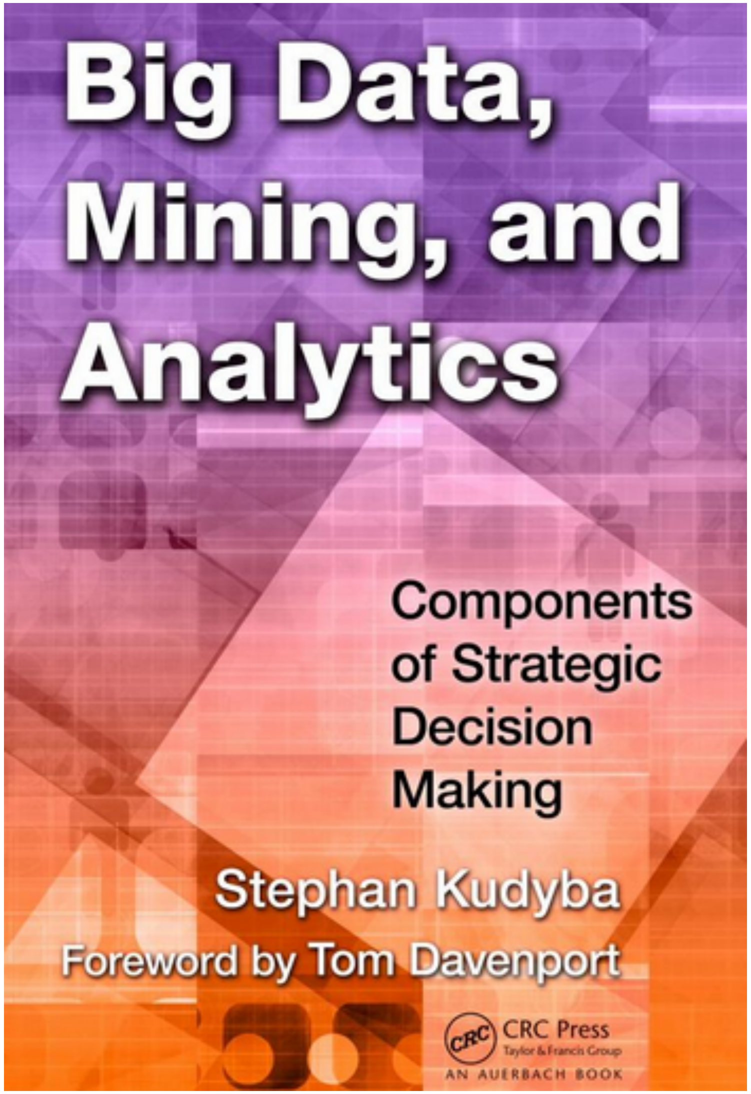
Source: https://www-01.ibm.com/software/data/bigdata/

Value

Stephan Kudyba (2014),

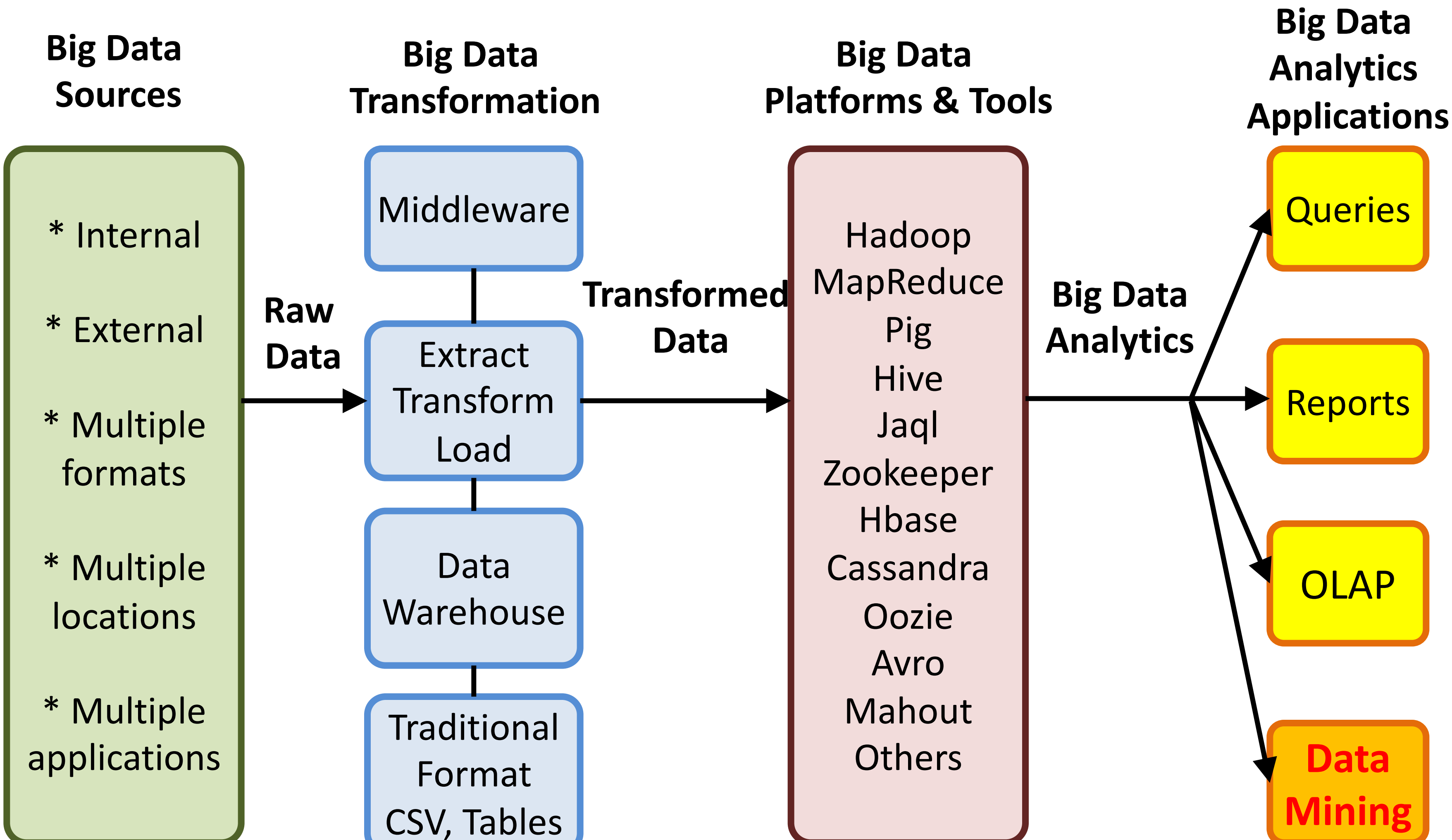
Big Data, Mining, and Analytics:

Components of Strategic Decision Making, Auerbach Publications



Source: http://www.amazon.com/gp/product/1466568704

Architecture of Big Data Analytics



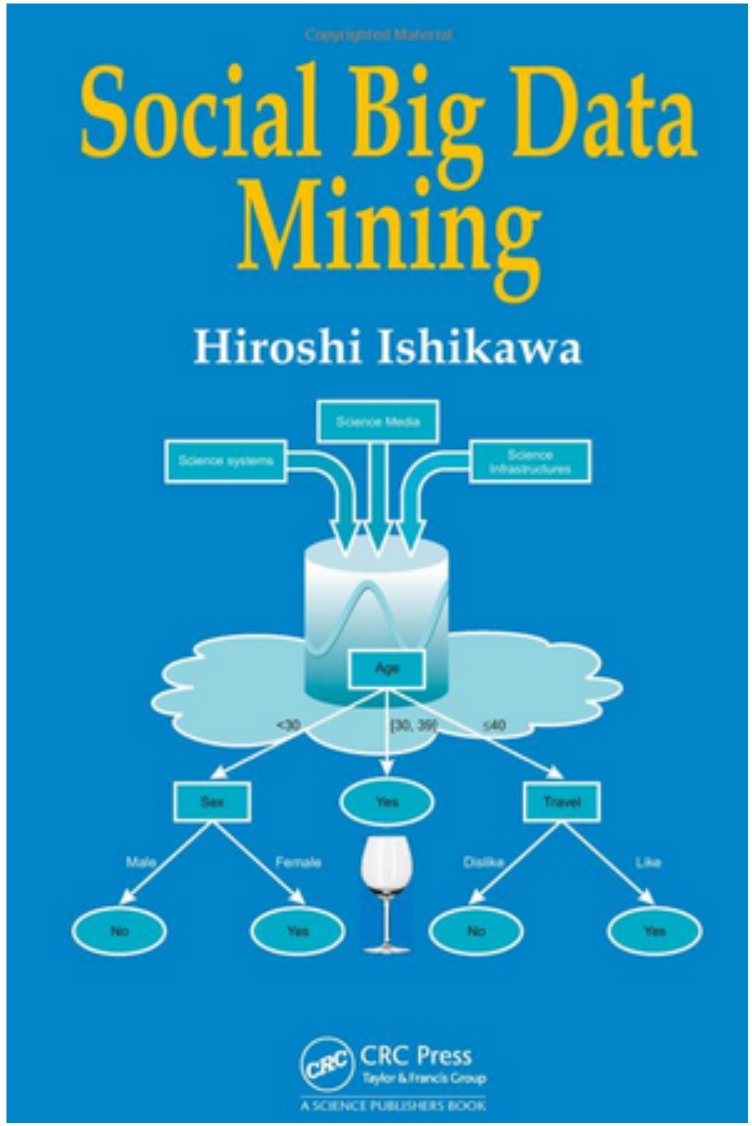
Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Architecture of Big Data Analytics

Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Social Big Data Mining

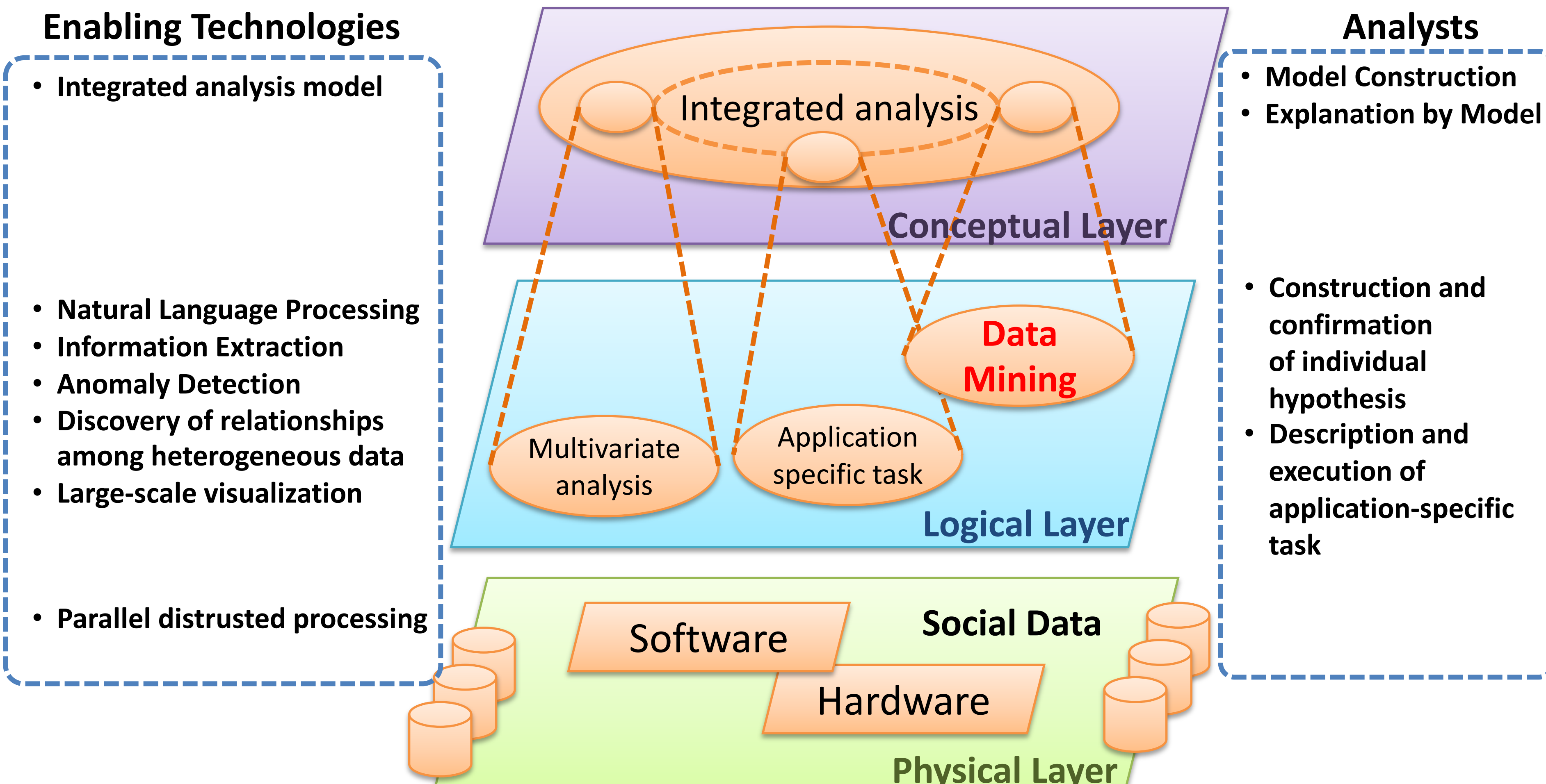
(Hiroshi Ishikawa, 2015)



Source: http://www.amazon.com/Social-Data-Mining-Hiroshi-Ishikawa/dp/149871093X

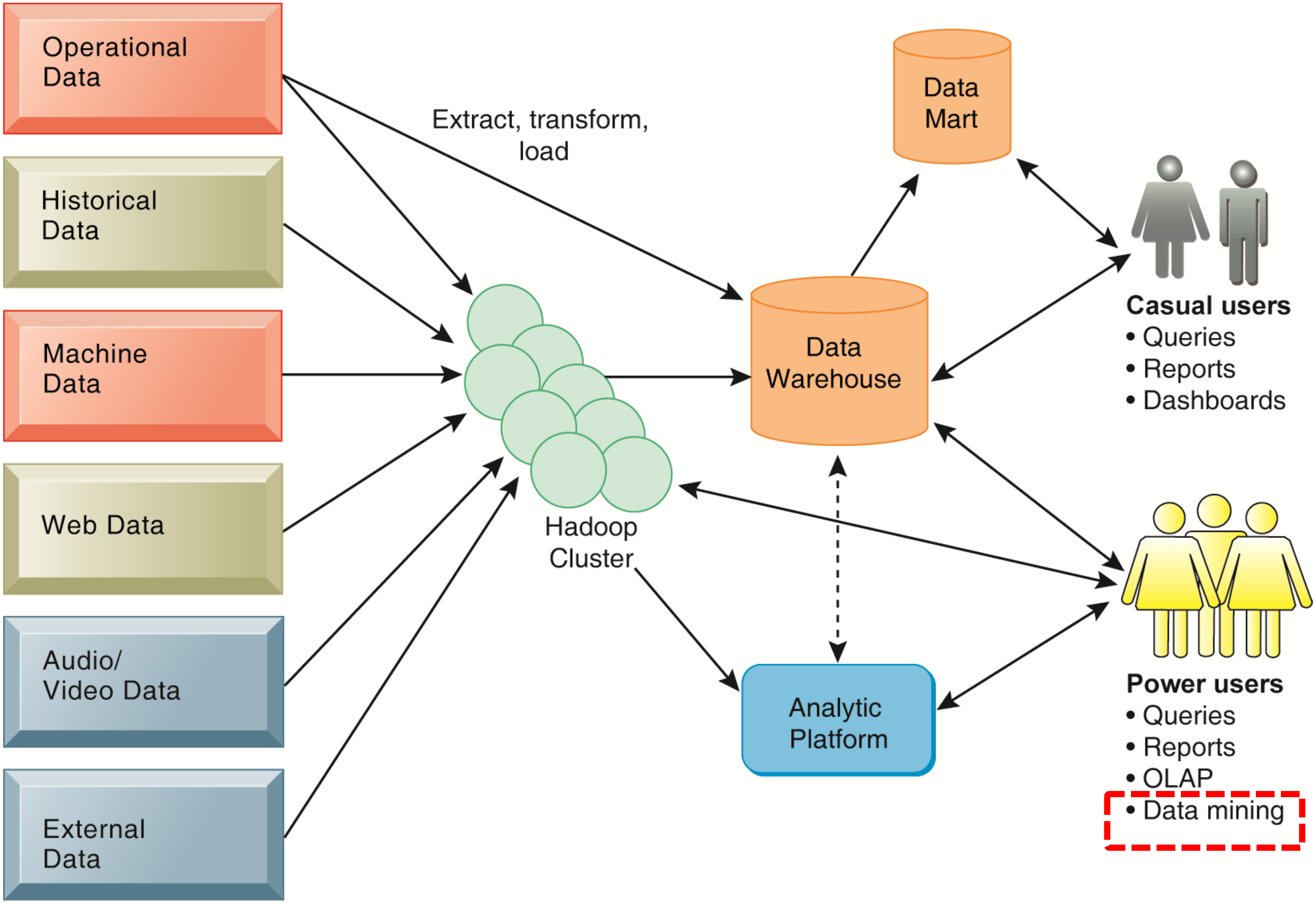
Architecture for Social Big Data Mining

(Hiroshi Ishikawa, 2015)



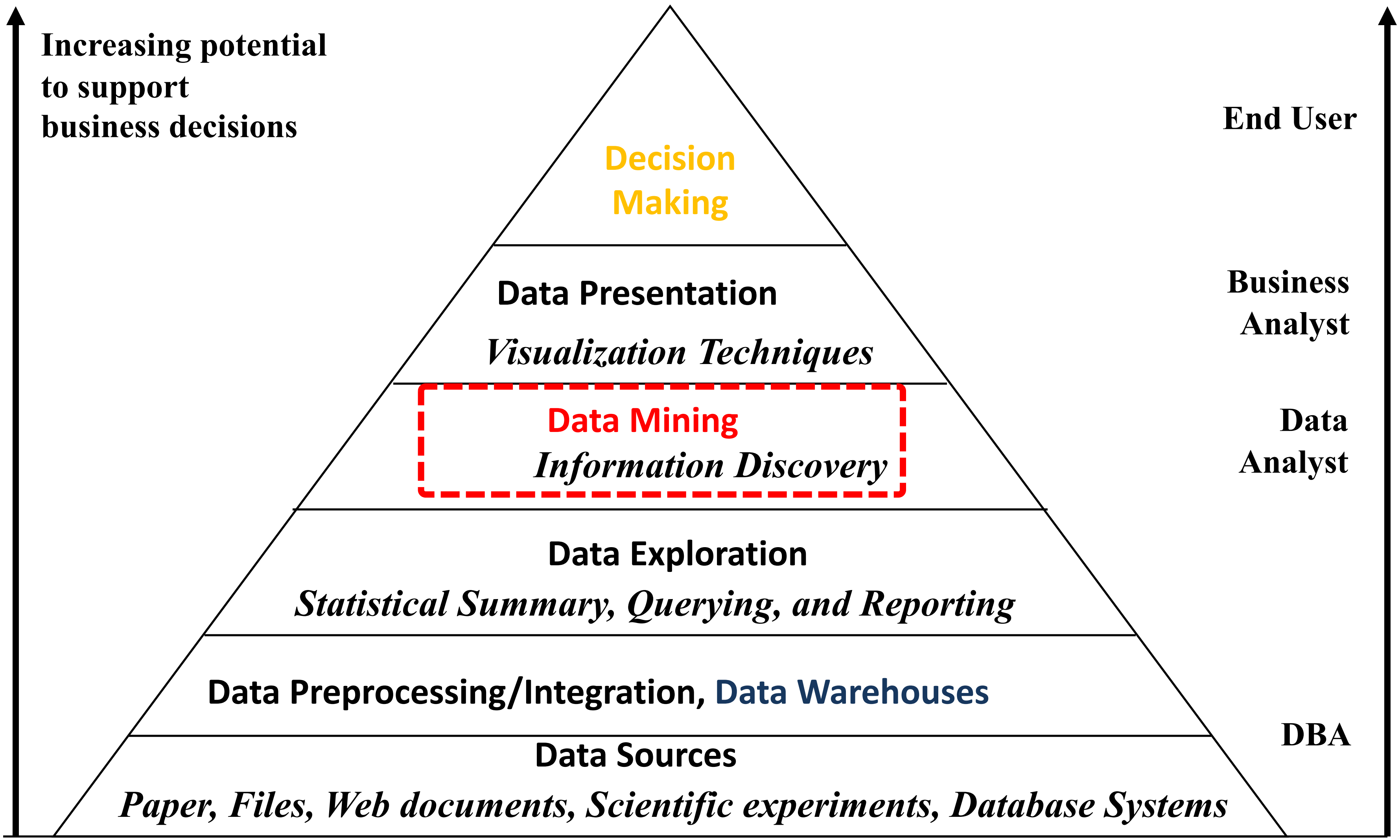
Source: Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press

Business Intelligence (BI) Infrastructure



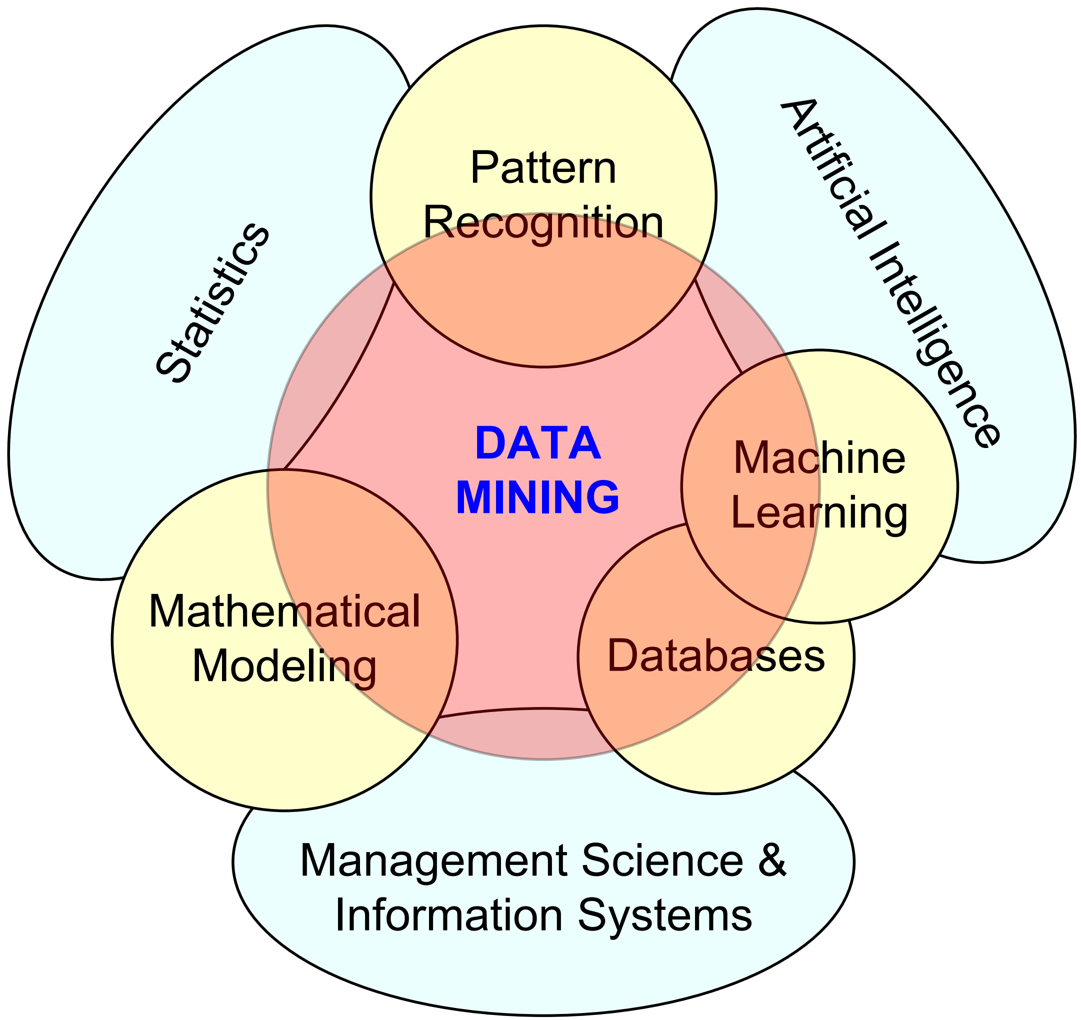
Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.

Business Intelligence and Data Mining



Source: Jiawei Han and Micheline Kamber (2006), Data Mining: Concepts and Techniques, Second Edition, Elsevier

Data Mining at the Intersection of Many Disciplines



Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Data Mining:

Core Analytics Process

The KDD Process for Extracting Useful Knowledge from Volumes of Data

Source: Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, 39(11), 27-34.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996).
The KDD Process for
Extracting Useful Knowledge
from Volumes of Data.
Communications of the ACM, 39(11), 27-34.

Knowledge Discovery in Databases creates the context for developing the tools needed to control the flood of data facing organizations that depend on ever-growing databases of business, manufacturing, scientific, and personal information.

The KDD Process for Extracting Useful Knowledge from Volumes of Data

Usama Fayyad,
Gregory Piatetsky-Shapiro,
and Padhraic Smyth

As we march into the age of digital information, the problem of data overload looms ominously ahead. Our ability to analyze and understand massive datasets lags far behind our ability to gather and store the data. A new generation of computational techniques and tools is required to support the extraction of useful knowledge from the rapidly growing volumes of data. These techniques and tools are the subject of the emerging field of knowledge discovery in databases (KDD) and data mining.

Large databases of digital information are ubiquitous. Data from the neighborhood store's checkout register, your bank's credit card authorization device, records in your doctor's office, patterns in your telephone calls, and many more applications generate streams of digital records archived in huge databases, sometimes in so-called data warehouses.

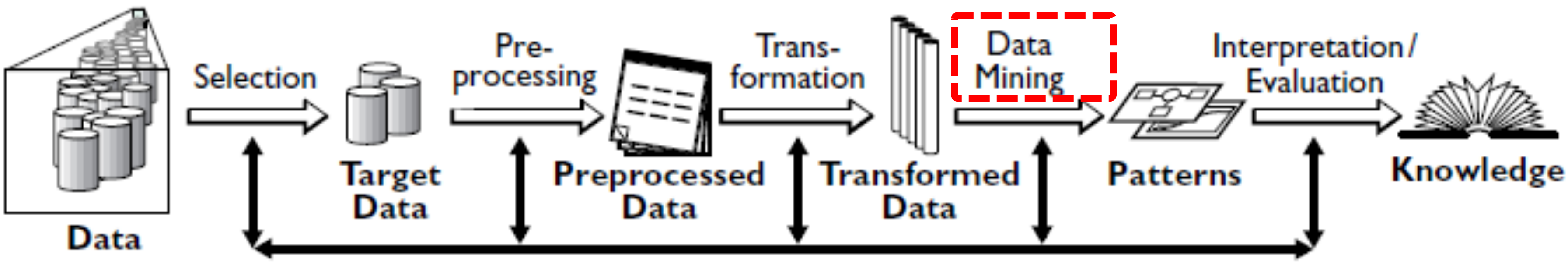
Current hardware and database technology allow efficient and inexpensive reliable data storage and access. However, whether the context is business, medicine, science, or government, the datasets themselves (in raw form) are of little direct value. What is of value is the knowledge that can be inferred from the data and put to use. For example, the marketing database of a consumer

TERRY WIDENER

Data Mining

Knowledge Discovery in Databases (KDD) Process

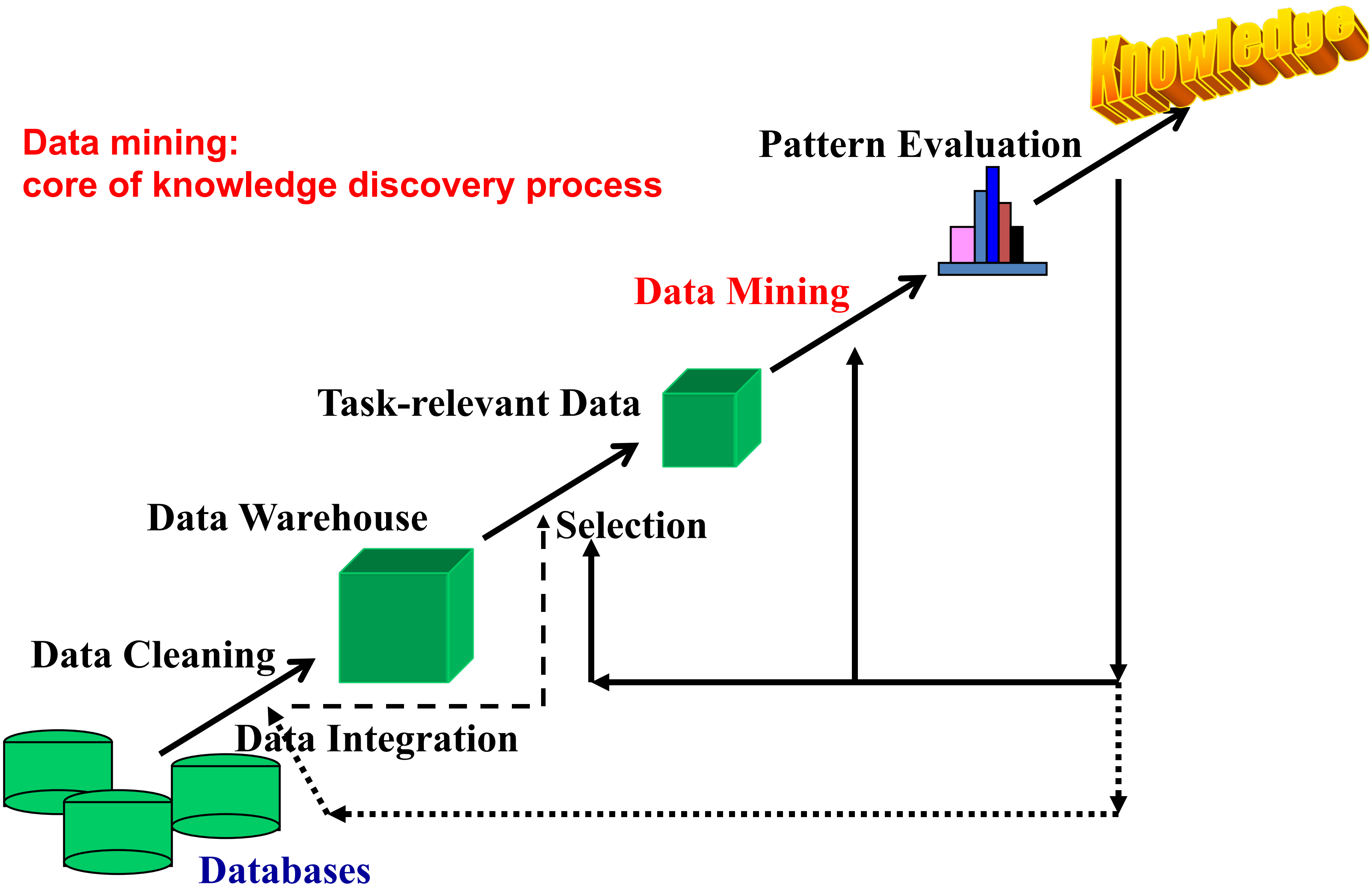
(Fayyad et al., 1996)



Source: Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, 39(11), 27-34.

Knowledge Discovery (KDD) Process

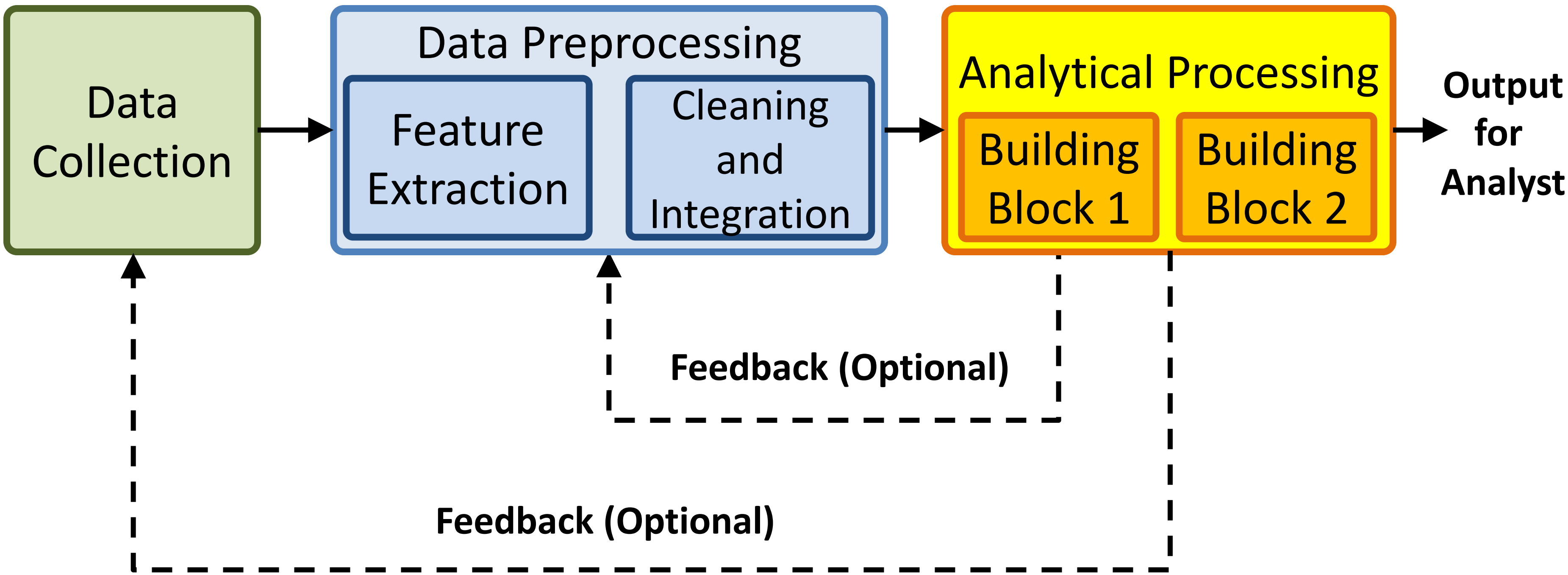
Data mining:
core of knowledge discovery process



Source: Han & Kamber (2006)

Data Mining Processing Pipeline

(Charu Aggarwal, 2015)



Source: Charu Aggarwal (2015), Data Mining: The Textbook Hardcover, Springer

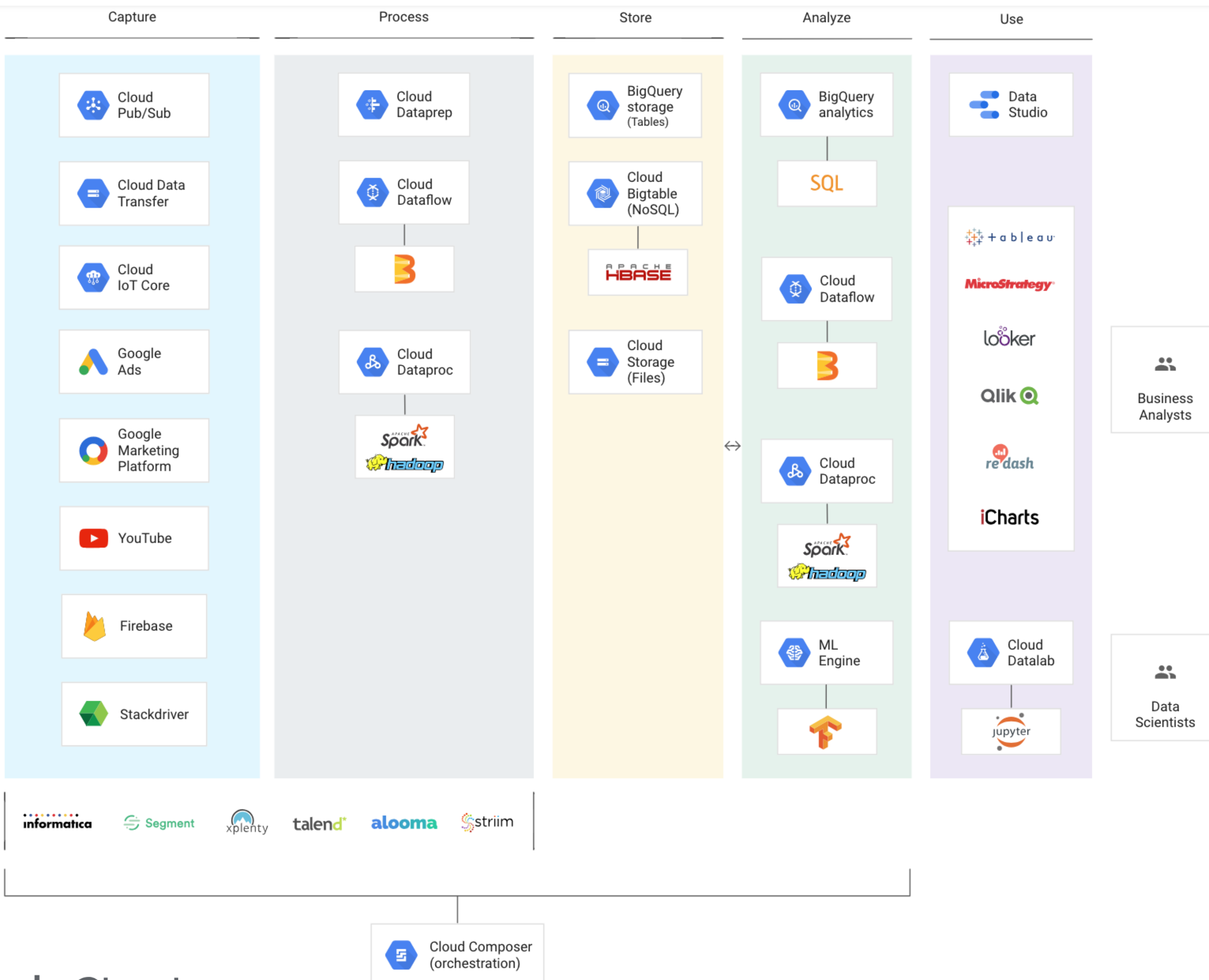
A Taxonomy for Data Mining Tasks

Data Mining	Learning Method	Popular Algorithms
Prediction	Supervised	Classification and Regression Trees, ANN, SVM, Genetic Algorithms
Classification	Supervised	Decision trees, ANN/MLP, SVM, Rough sets, Genetic Algorithms
Regression	Supervised	Linear/Nonlinear Regression, Regression trees, ANN/MLP, SVM
Association	Unsupervised	Apriory, OneR, ZeroR, Eclat
Link analysis	Unsupervised	Expectation Maximization, Apriory Algorithm, Graph-based Matching
Sequence analysis	Unsupervised	Apriory Algorithm, FP-Growth technique
Clustering	Unsupervised	K-means, ANN/SOM
Outlier analysis	Unsupervised	K-means, Expectation Maximization (EM)

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Cloud Computing

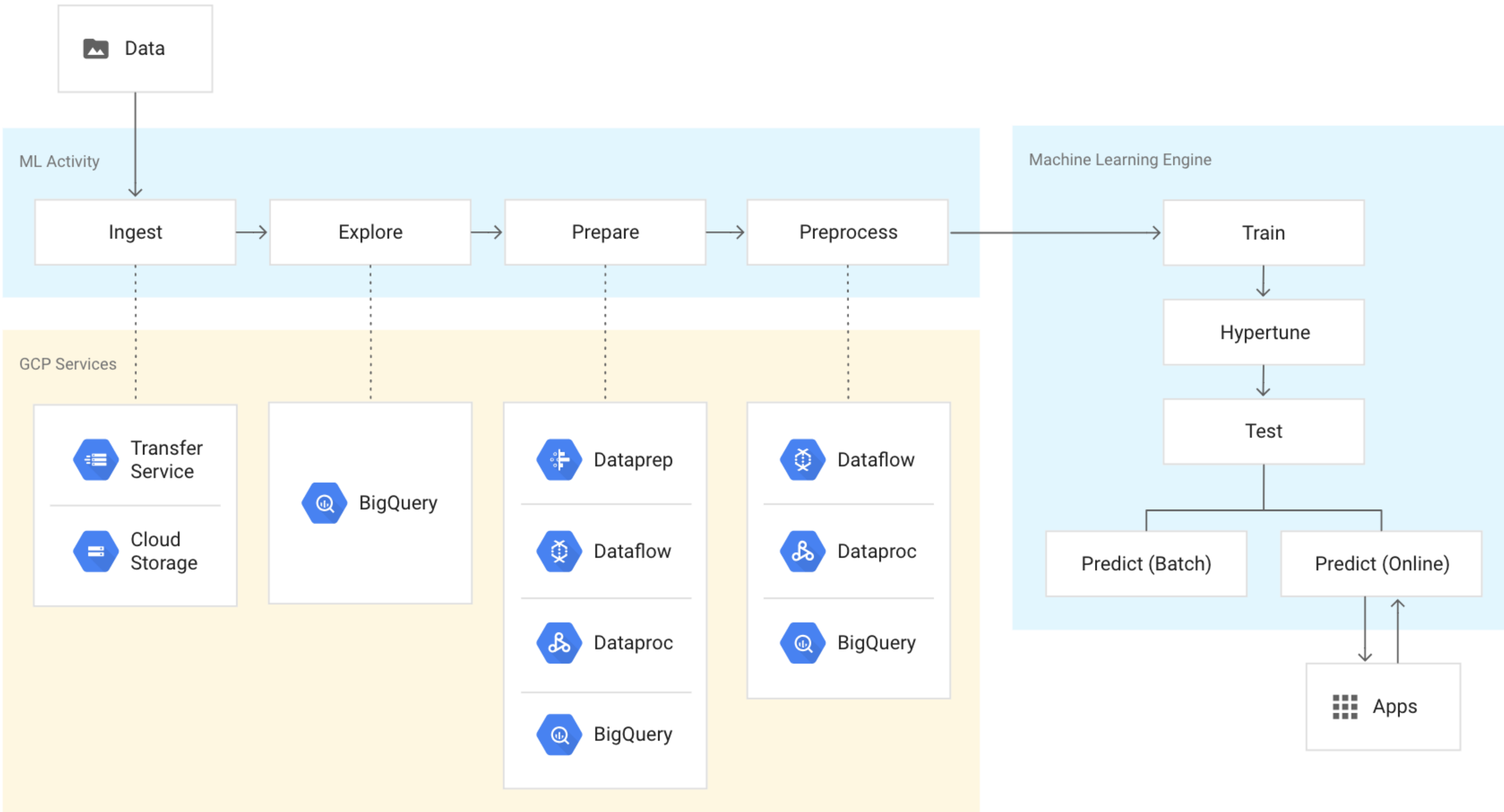
Google Cloud Big Data Analytics



Source: https://cloud.google.com/solutions/big-data/

Google Cloud

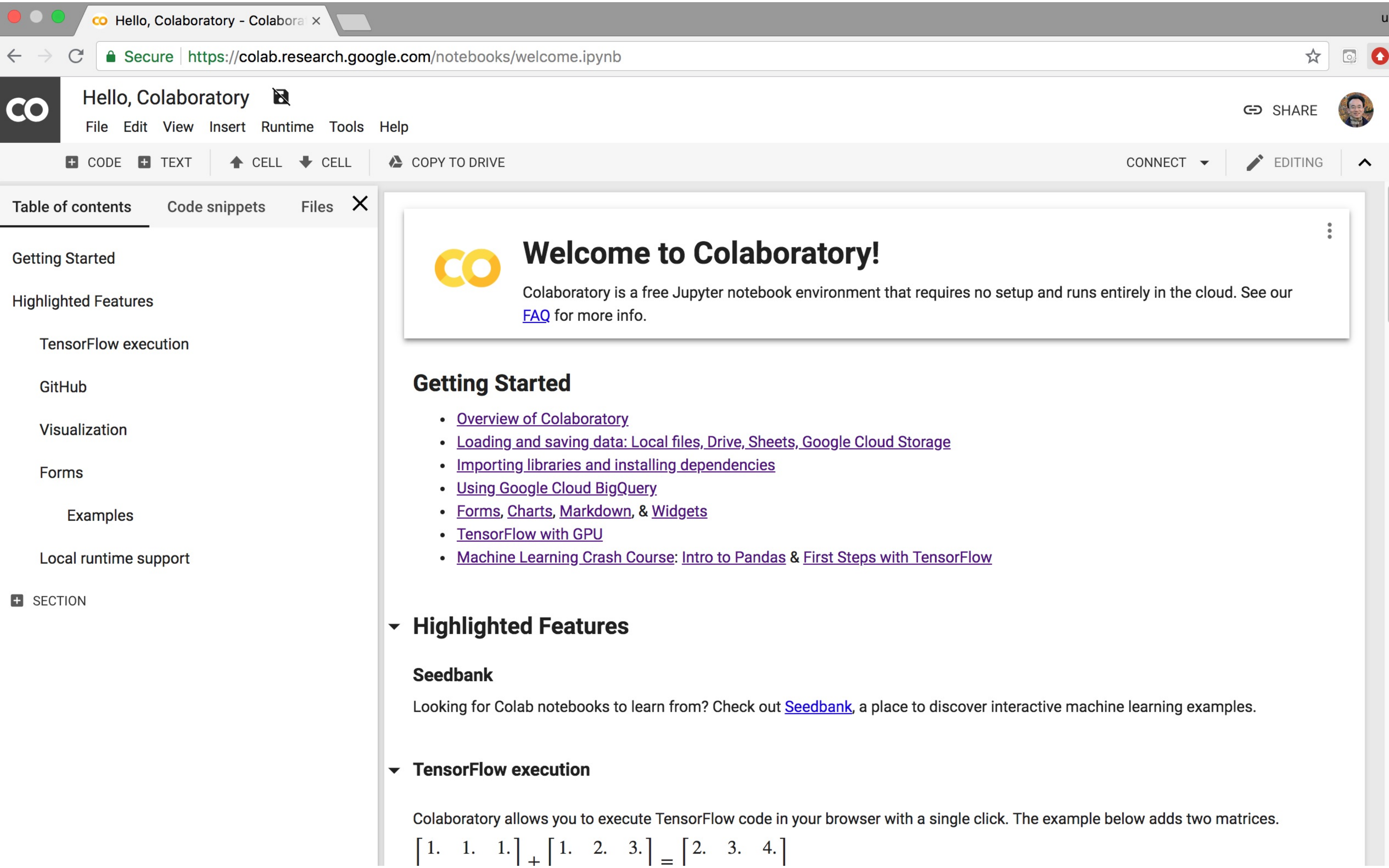
Machine learning and Cloud AI



Google Cloud

Source: https://cloud.google.com/solutions/big-data/overview/machine-learning-cloud-ai/

Google Colab



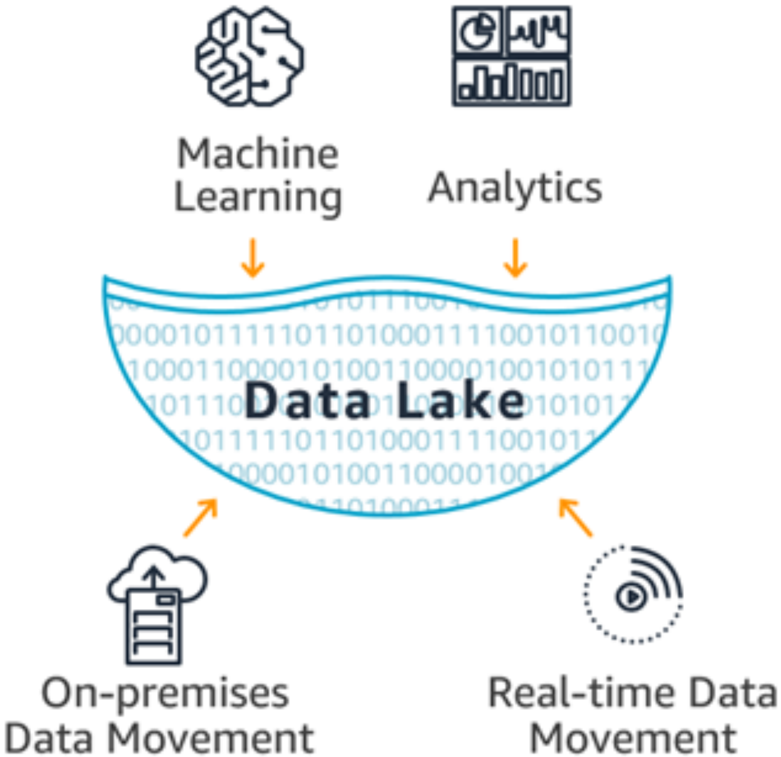
https://colab.research.google.com/notebooks/welcome.ipynb

Cloud Computing
AWS
Amazon Web Services

- Compute
- Storage
- Database
- Migration
- Networking & Content Delivery
- Developer Tools
- Management Tools
- Media Services
- Security, Identity & Compliance
- Analytics
- Machine Learning
- Mobile Services
- AR & VR
- Application Integration
- Customer Engagement
- Business Productivity
- Desktop & App Streaming
- Internet of Things
- Game Development
- AWS Cost Management

Source: https://aws.amazon.com/

Data Lakes and Analytics on AWS



Data Movement

Import your data from on-premises, and in real-time.

Data Lake

Store any type of data securely, from gigabytes to exabytes.

Analytics

Analyze your data with a broad selection of analytic tools and engines.

Machine Learning

Forecast future outcomes, and prescribe actions.

Source: https://aws.amazon.com/big-data/datalakes-and-analytics/

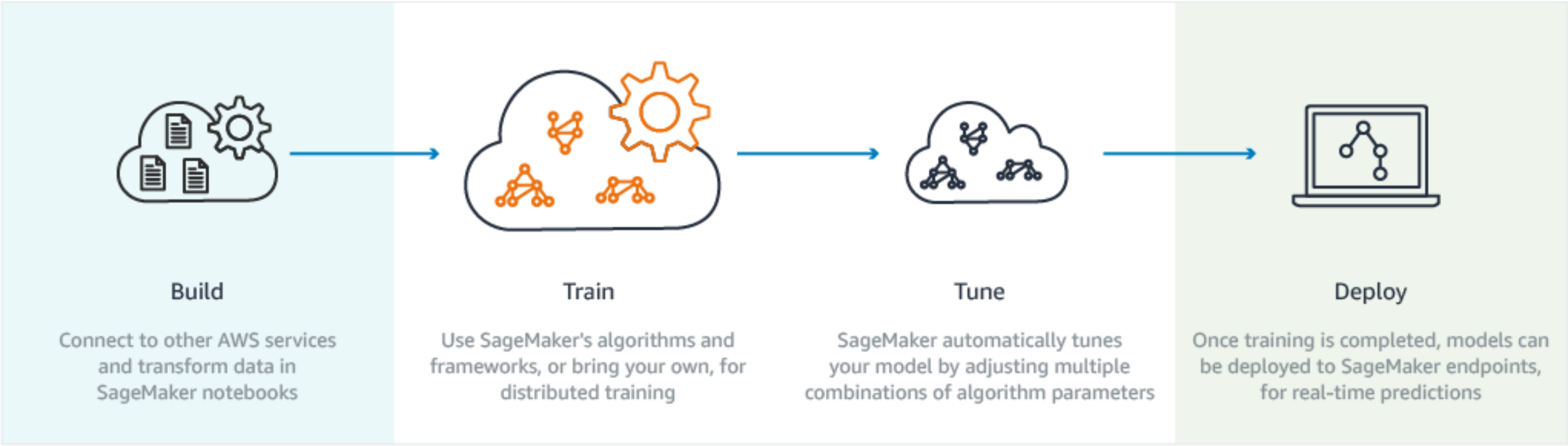
AWS Products
Analytics

- **Amazon Athena**
  - Query data in S3 using SQL
- **Amazon CloudSearch**
  - Managed search service
- **Amazon EMR**
  - Hosted Hadoop framework
- **Amazon Elasticsearch Service**
  - Run and scale Elasticsearch clusters
- **Amazon Kinesis**
  - Analyze real-time video and data streams
- **Amazon Redshift**
  - Fast, simple, cost-effective data warehousing
- **Amazon QuickSight**
  - Fast business analytics service
- **AWS Data Pipeline**
  - Orchestration service for periodic, data-driven workflows
- **AWS Glue**
  - Prepare and load data

Source: https://aws.amazon.com/

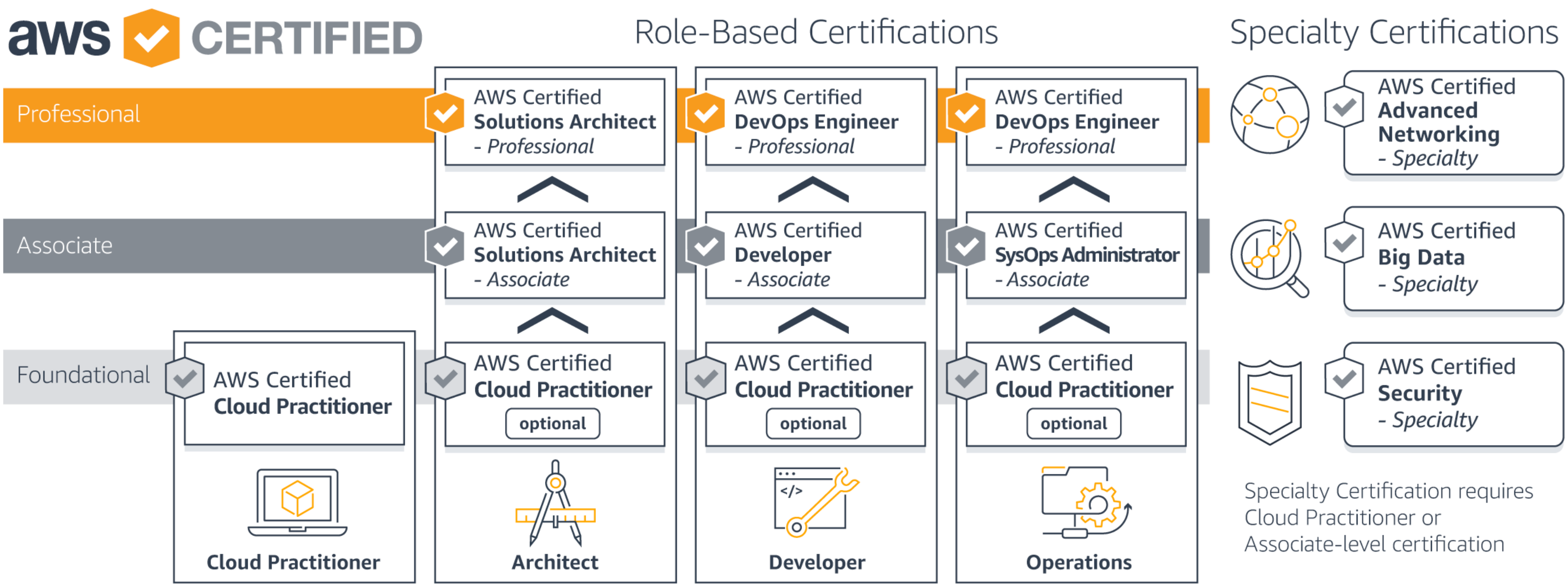
Machine Learning on AWS

Machine learning in the hands of every developer and data scientist



Source: https://aws.amazon.com/machine-learning/

Cloud Computing
AWS Cloud Practitioner
AWS Solutions Architect
AWS Certified Big Data Specialty

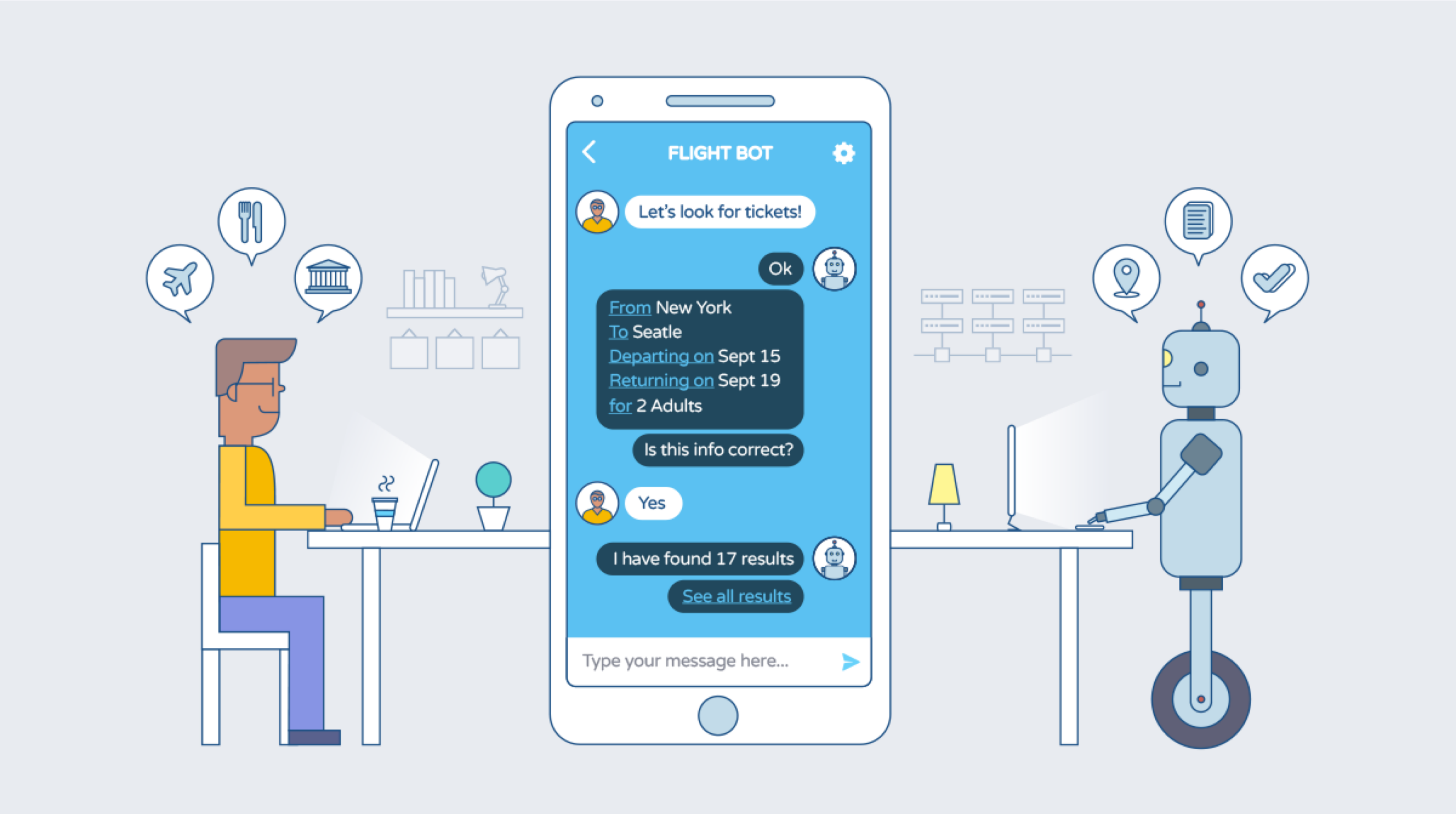


Source: https://aws.amazon.com/certification/

Short Text Conversation (STC)

AI and Dialogue System

Chatbot



Source: https://www.mdsdecoded.com/blog/the-rise-of-chatbots/

Can machines think?

(Alan Turing ,1950)

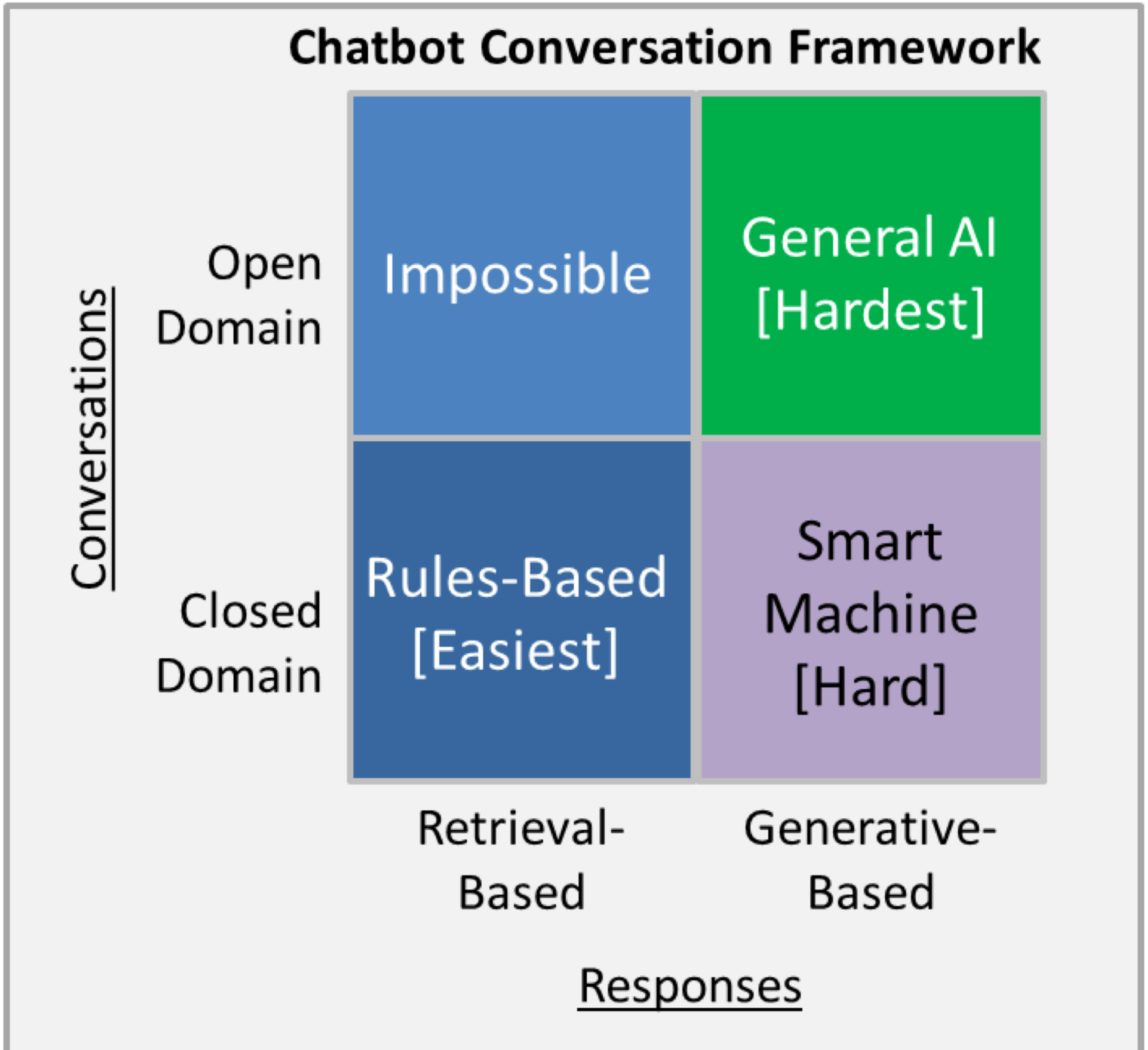
Source: Cahn, Jack. "CHATBOT: Architecture, Design, & Development." PhD diss., University of Pennsylvania, 2017.

Chatbot

“online human-computer dialog system with natural language.”

Source: Cahn, Jack. "CHATBOT: Architecture, Design, & Development." PhD diss., University of Pennsylvania, 2017.

Chatbot Conversation Framework

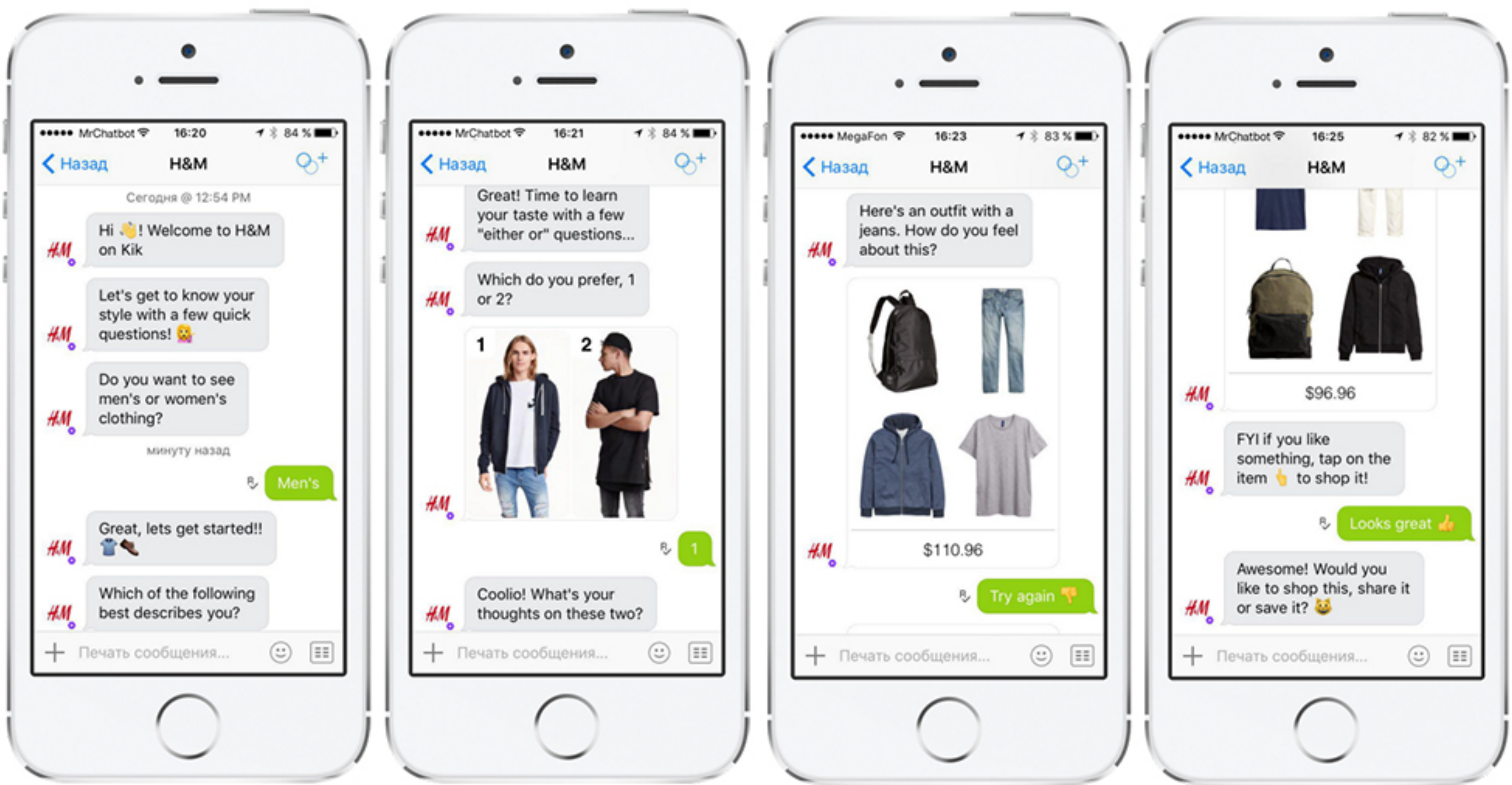


Source: https://chatbotslife.com/ultimate-guide-to-leveraging-nlp-machine-learning-for-you-chatbot-531ff2dd870c

From
E-Commerce
to
Conversational Commerce:
Chatbots
and
Virtual Assistants

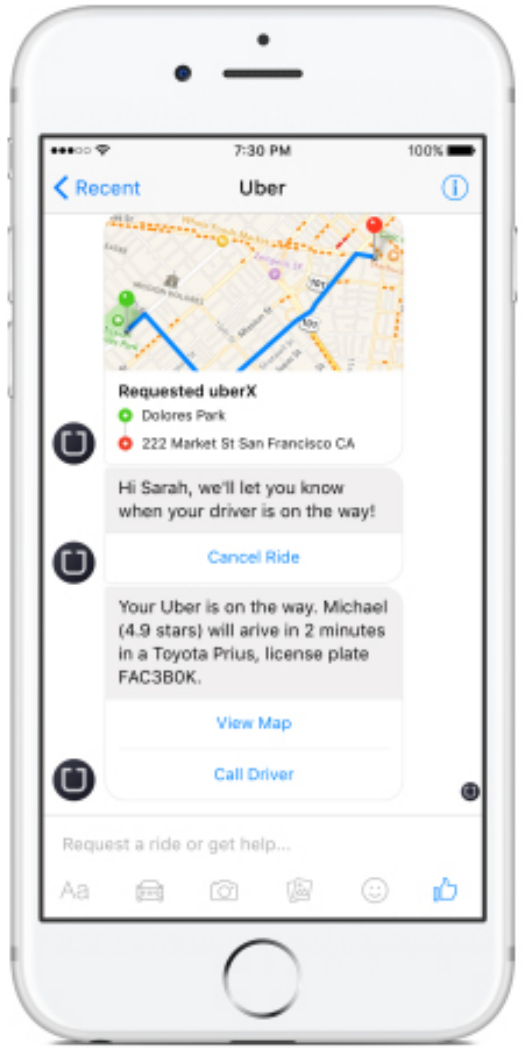
Source: http://www.guided-selling.org/from-e-commerce-to-conversational-commerce/

H&M’s chatbot on Kik



Source: http://www.guided-selling.org/from-e-commerce-to-conversational-commerce/

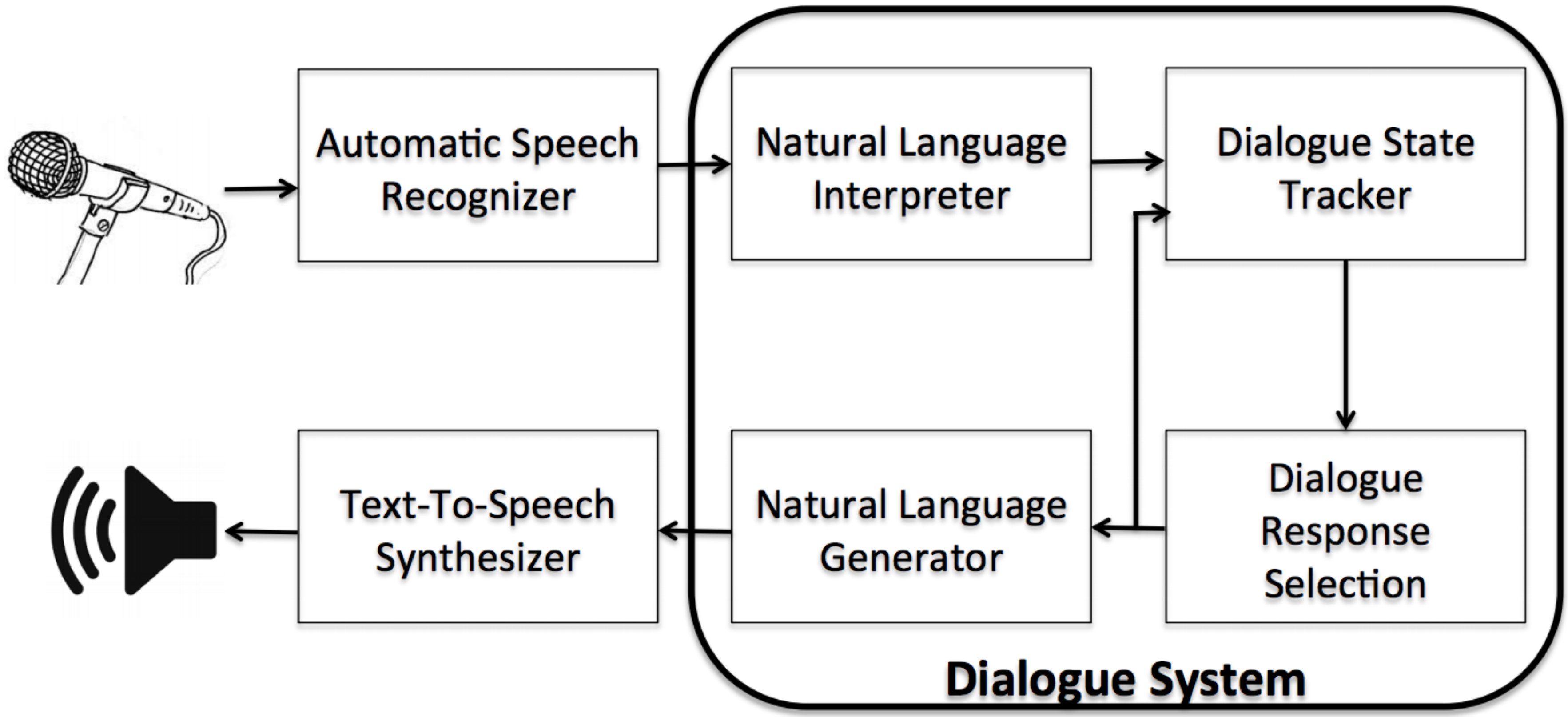
Uber’s chatbot on Facebook’s messenger



Uber’s chatbot on Facebook’s messenger
- one main benefit: it loads much faster than the Uber app

Source: http://www.guided-selling.org/from-e-commerce-to-conversational-commerce/

Dialogue System



Source: Serban, I. V., Lowe, R., Charlin, L., & Pineau, J. (2015). A survey of available corpora for building data-driven dialogue systems. *arXiv preprint arXiv:1512.05742*.

Short Text Conversation Task (STC-3)
Chinese Emotional Conversation Generation (CECG) Subtask

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge.html

NTCIR Short Text Conversation STC-1, STC-2, STC-3

	Japanese	Chinese	English
NTCIR-12 STC-1 22 active participants	Twitter, Retrieval	Weibo, Retrieval	
NTCIR-13 STC-2 27 active participants	Yahoo! News, Retrieval+ Generation	Weibo, Retrieval+ Generation	
NTCIR-14 STC-3		Weibo, Generation for given emotion categories Weibo+English translations, distribution estimation for subjective annotations	Weibo+English translations, distribution estimation for subjective annotations

Chinese Emotional Conversation Generation (CECG) subtask

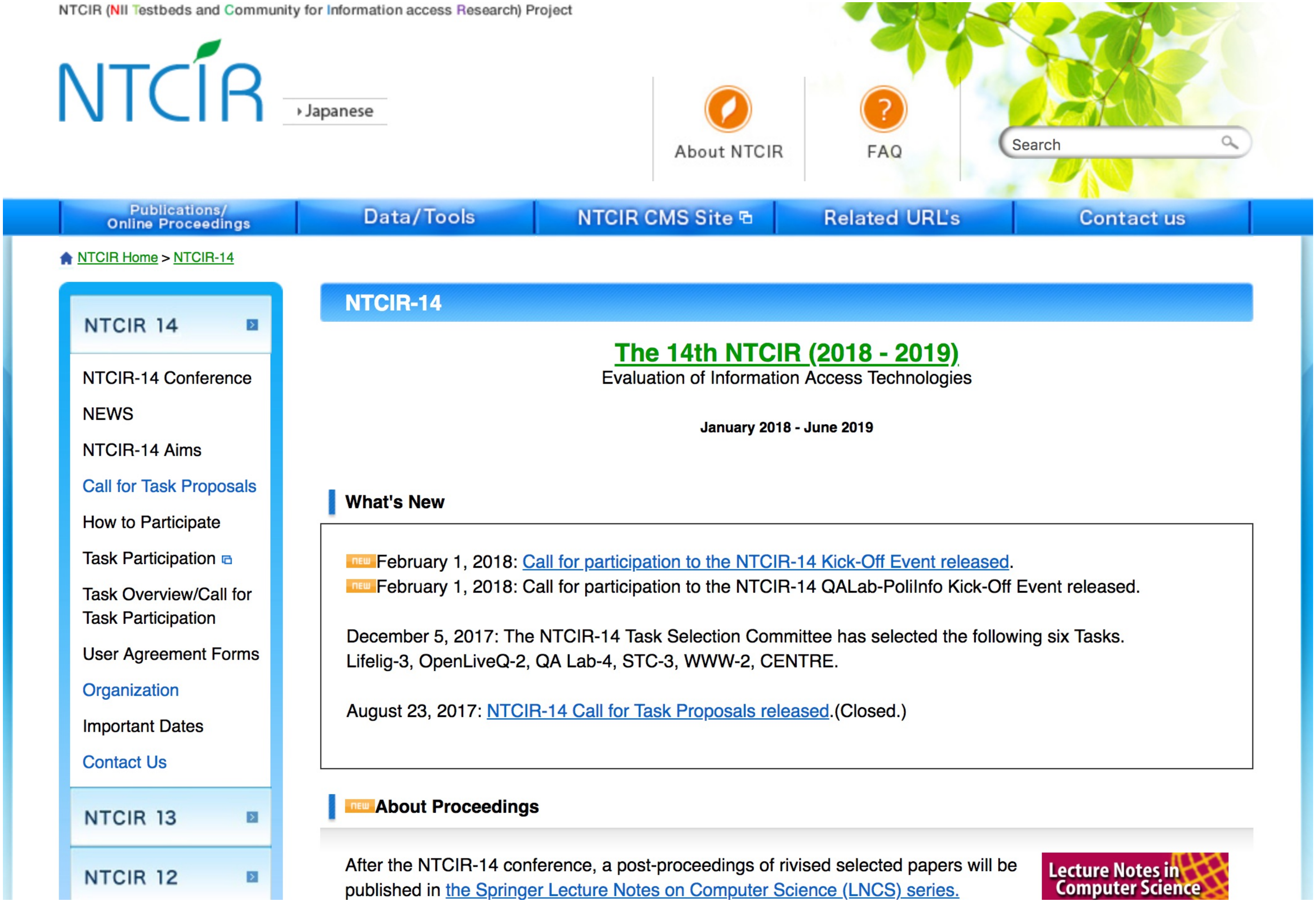
Dialogue Quality (DQ) and Nugget Detection (ND) subtasks

Single-turn, Non task-oriented

Multi-turn, task-oriented (helpdesk)

Source: https://waseda.app.box.com/v/STC3atNTCIR-14

The 14th NTCIR (2018 - 2019)



http://research.nii.ac.jp/ntcir/ntcir-14/index.html

NTCIR-14
Short Text Conversation Task (STC-3)

NTCIR-14 Short Text Conversation Task (STC-3)

- NTCIR
- Twitter: @ntcirstc
- STC-3@NTCIR-14

Welcome to the top page of STC-3@NTCIR-14!
STC-3 offers three subtasks:

- Chinese Emotional Conversation Generation (CECG) Subtask
- Dialogue Quality (DQ) Subtask (for Chinese and English)
- Nugget Detection (ND) Subtask (for Chinese and English)

Key dates for DQ and ND Subtasks

Feb-Mar 2018Crawling Chinese test data from Weibo

Oct 2017-Jan 2018	Training data translation into English
Apr-Jun, 2018	Test data translation into English
Jul-Aug 2018	Training/test data annotation
Aug 31, 2018	STC-3 task registrations due (CECG, DQ, ND)
Sep 1, 2018	Training data with annotations released
Nov 1, 2018	Test data released
Nov 30, 2018	Run submissions due
Dec 20, 2018	Results and draft overview released to participants
Feb 1, 2019	Participant papers due
Mar 1, 2019	Acceptance notification
Mar 20, 2019	All camera-ready papers due
Jun 2019	NTCIR-14 Conference@NII

NTCIR-14 STC-3

Short Text Conversation Task (STC-3)

Chinese Emotional Conversation Generation (CECG) Subtask

NTCIR

Short Text Conversation Task (STC-3)

Chinese Emotional Conversation Generation (CECG) Subtask

- Home
- Task Definition
- Dataset Description
- Evaluation Metric
- Time Schedule
- Copy Rights & Contacts

Links

- NTCIR-14
- NTCIR-14 STC-3
- NLPCC 2017

Call for Participation

In recent years, there has been a rising tendency in AI research to enhance Human-Computer Interaction by humanizing machines. However, to create a robot capable of acting and talking with a user at the human level requires the robot to understand human cognitive behaviors, while one of the most important human behaviors is expressing and understanding emotions and affects. As a vital part of human intelligence, emotional intelligence is defined as the ability to perceive, integrate, understand, and regulate emotions. Though a variety of models have been proposed for conversation generation from large-scale social data, it is still quite challenging (and yet to be addressed) to generate emotional responses.

In this challenge, participants are expected to generate Chinese responses that are not only appropriate in content but also adequate in emotion, which is quite important for building an empathic chatting machine. For instance, if user says "My cat died yesterday", the most appropriate response may be "It's so sad, so sorry to hear that" to express sadness, but also could be "Bad things always happen, I hope you will be happy soon"to express comfort.

Previous Evaluation Challenge at NLPCC 2017

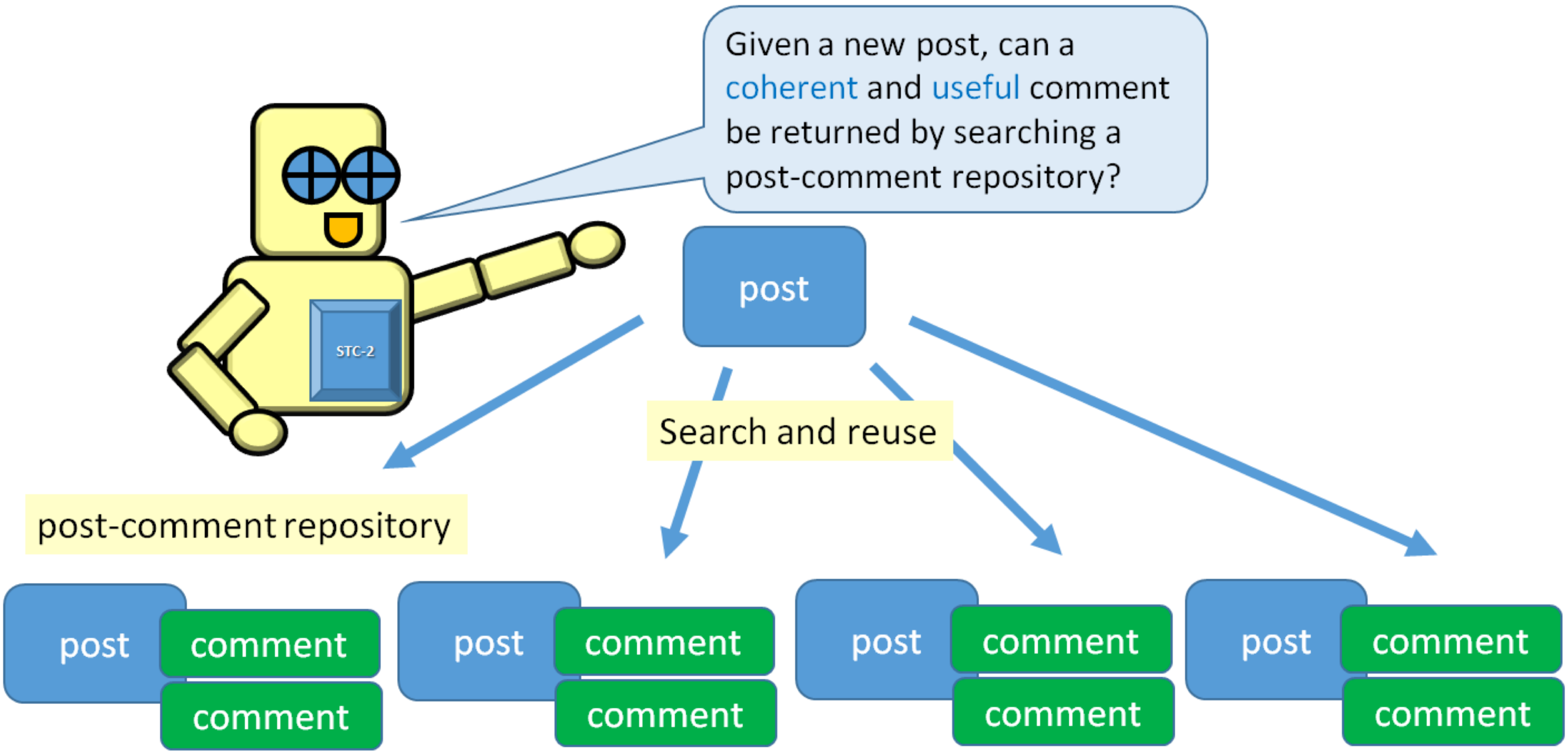
Overview of the NLPCC 2017 Shared Task: Emotion Generation Challenge

AI lab. of Computer Science, Tsinghua University, Beijing 100084, China.

http://www.aihuang.org/p/challenge.html

Short Text Conversation (NTCIR-13 STC2) Retrieval-based

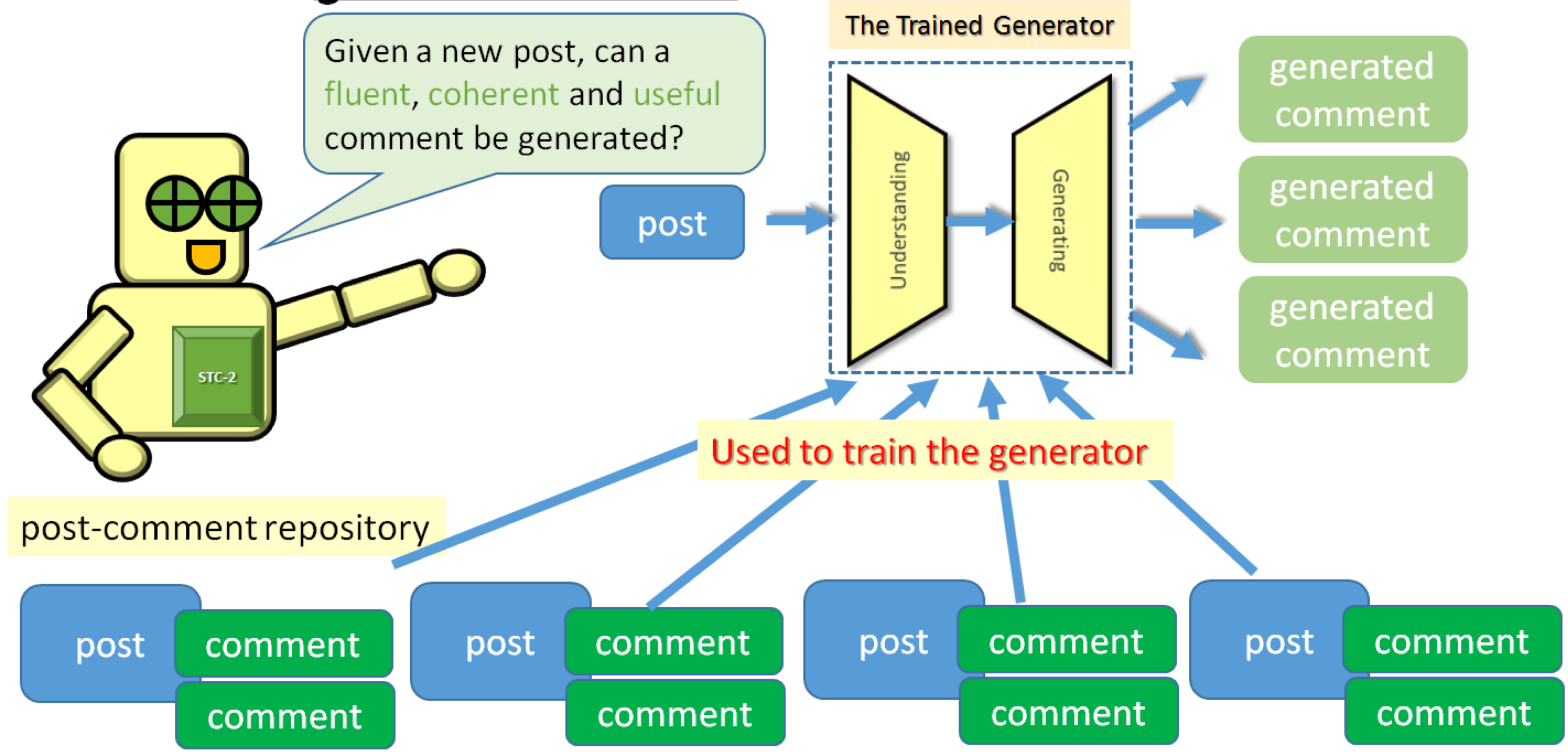
retrieval-based method



Source: http://ntcirstc.noahlab.com.hk/STC2/stc-cn.htm

Short Text Conversation (NTCIR-13 STC2) Generation-based

generation-based method



Source: http://ntcirstc.noahlab.com.hk/STC2/stc-cn.htm

Short Text Conversation (STC-3)

- Emotional Conversation Generation
- Dialogue Quality
- Nugget Detection subtasks
  using Chinese and English dialogue data

Source: http://research.nii.ac.jp/ntcir/ntcir-14/tasks.html

NTCIR-14
Short Text Conversation Task (STC-3)

- Three new subtasks
  - Chinese Emotional Conversation Generation (CECG)
  - Dialogue Quality (DQ) (for Chinese and English)
  - Nugget Detection (ND) (for Chinese and English)

Source: http://research.nii.ac.jp/ntcir/ntcir-14/tasks.html

NTCIR-14 STC-3 CECG Time Schedule

- CECG
  - Aug 31, 2018 CECG registration due
  - Sep 15, 2018 Training data released
  - Nov 15, 2018 Test data released
  - Nov 25, 2018 Run submissions due
- ALL
  - Dec 20, 2018 Evaluation results and draft overview released
  - Feb 1, 2019 Participant paper submissions due
  - Mar 1, 2019 Acceptance notification
  - Mar 20, 2019 All camera-ready papers due
  - Jun 2019 NTCIR-14 Conference & EVIA 2019 in NII, Tokyo

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/time_schedule/

Given post:

爱狗还会做饭的男人，最帅了！

The man who cooks and loves dogs is very handsome!

Response	Chinese / Translated English	Emotion Class	Coherence and Fluency	Emotion Consistency	Label
Response 1	会做饭的男人是很帅的啊。 The man who cooks is handsome.	喜欢 Like	Yes	Yes	2
Response 2	哈哈，我也觉得。 Haha, I feel the same way.	喜欢 Like	Yes	No	1
Response 3	这是哪部电影里的？ Which movie is this from?	厌恶 Disgust	No	Yes	0
Response 4	哈哈,你也是。 Haha, the same to you.	喜欢 Like	No	No	0
Response 5	我爱你会。 I love you, will you.	厌恶 Disgust	No	Yes	0
Response 6	这是同主义同的道！ This is the same way of the same doctrine!	厌恶 Disgust	No	No	0

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/evaluation_metric/

NLPCC Emotion Classification Challenge 6 classes (64%)

- 0: Other
- 1: Like
- 2: Sadness
- 3: Disgust
- 4: Anger
- 5: Happiness

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/dataset_description/

Chinese Emotional Conversation Generation (CECG)

Post (Given)	Emotion Category (Given)	Response (to be generated)
爱狗还会做饭的男人，最帅了！ The man who cooks and loves dogs is very handsome!	喜欢 Like	会做饭的男人是很帅的啊。 The man who cooks is handsome.
爸爸，明天是六一啦，带我出去玩吧！ Daddy, tomorrow is June 1th Children's Day. Please take me out to play!	喜欢 Like	好，一定带。 Yes, of course.
被特别在乎的人忽略，会很难过，而装做不在乎，会更难过。 It's sad to be ignored by someone that you really care, and it's harder to pretend that you don't care.	悲伤 Sadness	因为在乎，所以在乎。 I am concerned with it, because I care about it.
冰天雪地的哈尔滨！你要冻死我咋地？！？！ Harbin is in the snow! Are you going to freeze me to death?!?!	喜欢 Like	注意保暖，小心别感冒! Keep warm. Watch out for getting cold!
本来想学一把沧桑，结果令我更忧伤。 I wanted to learn the vicissitudes of life, but I became sadder.	悲伤 Sadness	这是一个悲伤的故事。 It a sad story.
今天一整天都在海边训练。虽然很累，但还是很开心的！ I have been training at the seaside all day. Though very tired, I still very happy!	快乐 Happiness	是的呢，开心！ Yeah, happy!

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/task_definition/

Chinese Emotional Conversation Generation (CECG) Dataset

- 1,110,000 Weibo post-response pairs
  - [[[post,post_label],[response,response_label]], [[post,post_label],[response,response_label]],...].

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/dataset_description/

Fluency judgement on responses with repetitive words

Response	Fluency
悲哀，悲哀，悲哀。	Yes
飘逸，飘逸，飘逸！	Yes
好可爱，好可爱！	Yes
哈哈哈，当然啦	Yes
疯了。疯了。全疯了	Yes
对的对的对的对的对的对的对的对的对的对的对的	No
我骗骗骗骗我骗你信不信不信不信我反正信了	No
中国的教育	No
矮贱贱贱贱贱贱贱贱，你爱你，我也不爱你了	No
不关我事不关我事不关我事不关我事不关我事！	No

Source: Huang, Minlie, Zuoxian Ye, and Hao Zhou. "Overview of the NLPCC 2017 Shared Task: Emotion Generation Challenge." In *National CCF Conference on Natural Language Processing and Chinese Computing (NLPCC)*, pp. 926-936. Springer, Cham, 2017.

Sample responses generated by Seq2Seq and ECM (Emotional Chatting Machine)

Post	Model		Response(Original)	Response(Translated)
承诺，就是一个骗子说给一个傻子听的。 A promise is what a cheater says to a fool.	Seq2Seq		承诺，就是一个骗子说给一个傻子听的。	A promise is what a cheater says to a fool.
	ECM	Like	承诺是一个专一的傻子	Promise is a single-minded fool
		Happy	我就是一个开心的傻子	I'm just a happy fool
		Sad	我就是一个悲伤的傻子	I am a sad fool
		Disgust	一个骗子说给一个傻子听	A cheater says to a fool
		Angry	不管你信不信，反正我信了！	Whatever you believe or not, I trust in it!
知道真相的我眼泪笑下来 Knowing the truth, I laughed myself to tears	Seq2Seq		知道真相的我眼泪掉下来。	After knowing the truth, I burst into tears.
	ECM	Like	太可爱了，太可爱了！	Too lovely, it's too lovely.
		Happy	笑得我眼泪都出来了。	I can't help laugh myself to tears.
		Sad	知道真相的我眼泪掉下来。	After knowing the truth, I burst into tears.
		Disgust	这是真的还是假的？	Is this true?
		Angry	什么意思？没看懂。	What do you mean? I don't understand.

Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Sample responses generated by Seq2Seq and ECM (Emotional Chatting Machine)

Post	Model		Response(Original)
承诺，就是一个骗子说给一个傻子听的。 A promise is what a cheater says to a fool.	Seq2Seq		承诺，就是一个骗子说给一个傻子听的。
	ECM	Like	承诺是一个专一的傻子
		Happy	我就是一个开心的傻子
		Sad	我就是一个悲伤的傻子
		Disgust	一个骗子说给一个傻子听
		Angry	不管你信不信，反正我信了！

Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Emotional Short Text Conversation (ESTC) Dataset

Training	Posts	217,905	
	Responses	Angry	234,635
		Disgust	689,295
		Happy	306,364
		Like	1,226,954
		Sad	537,028
		Other	1,365,371
Validation	Posts	1,000	
Test	Posts	1,000	

Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

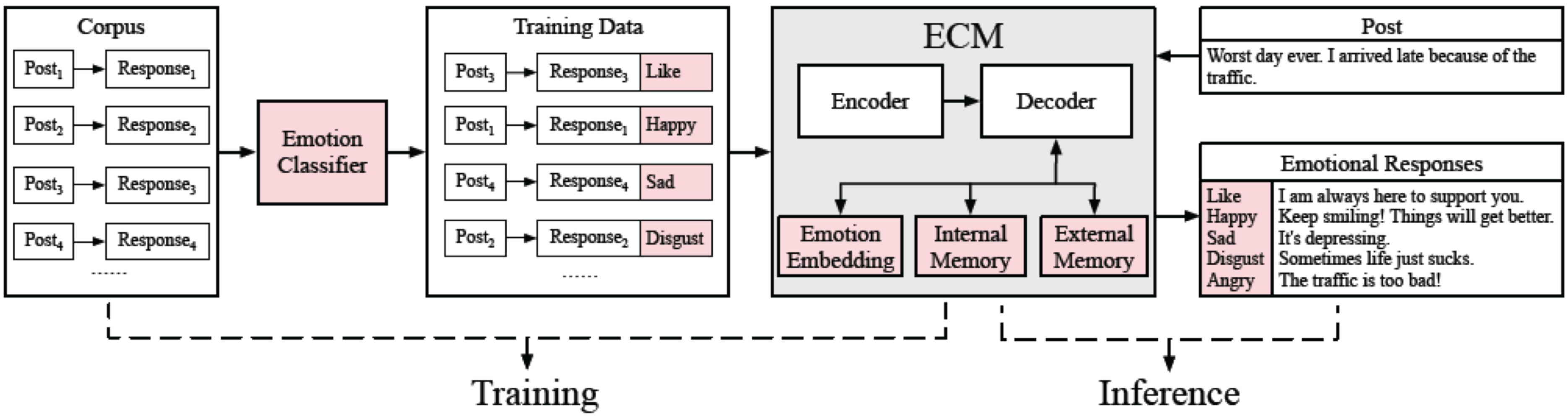
Conversations with/without considering emotion

Emotional Chatting Machine (ECM)

- User: Worst day ever. I arrived late because of the traffic.
  1. Basic Seq2Seq: You were late.
  2. ECM (Like): I am always here to support you.
  3. ECM (Happy): Keep smiling! Things will get better.
  4. ECM (Sad): It’s depressing.
  5. ECM (Disgust): Sometimes life just sucks.
  6. ECM (Angry): The traffic is too bad!

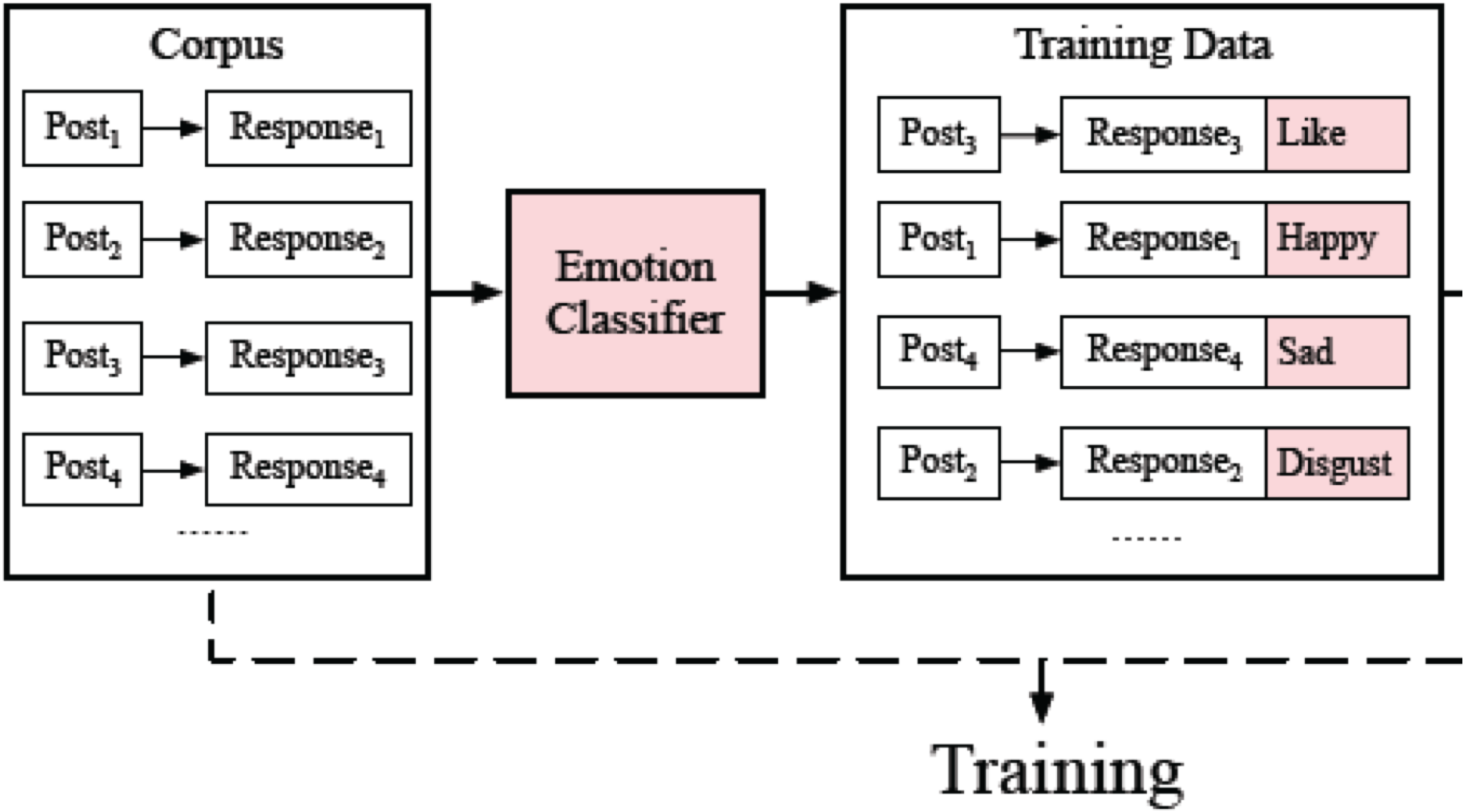
Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Overview of Emotional Chatting Machine (ECM)



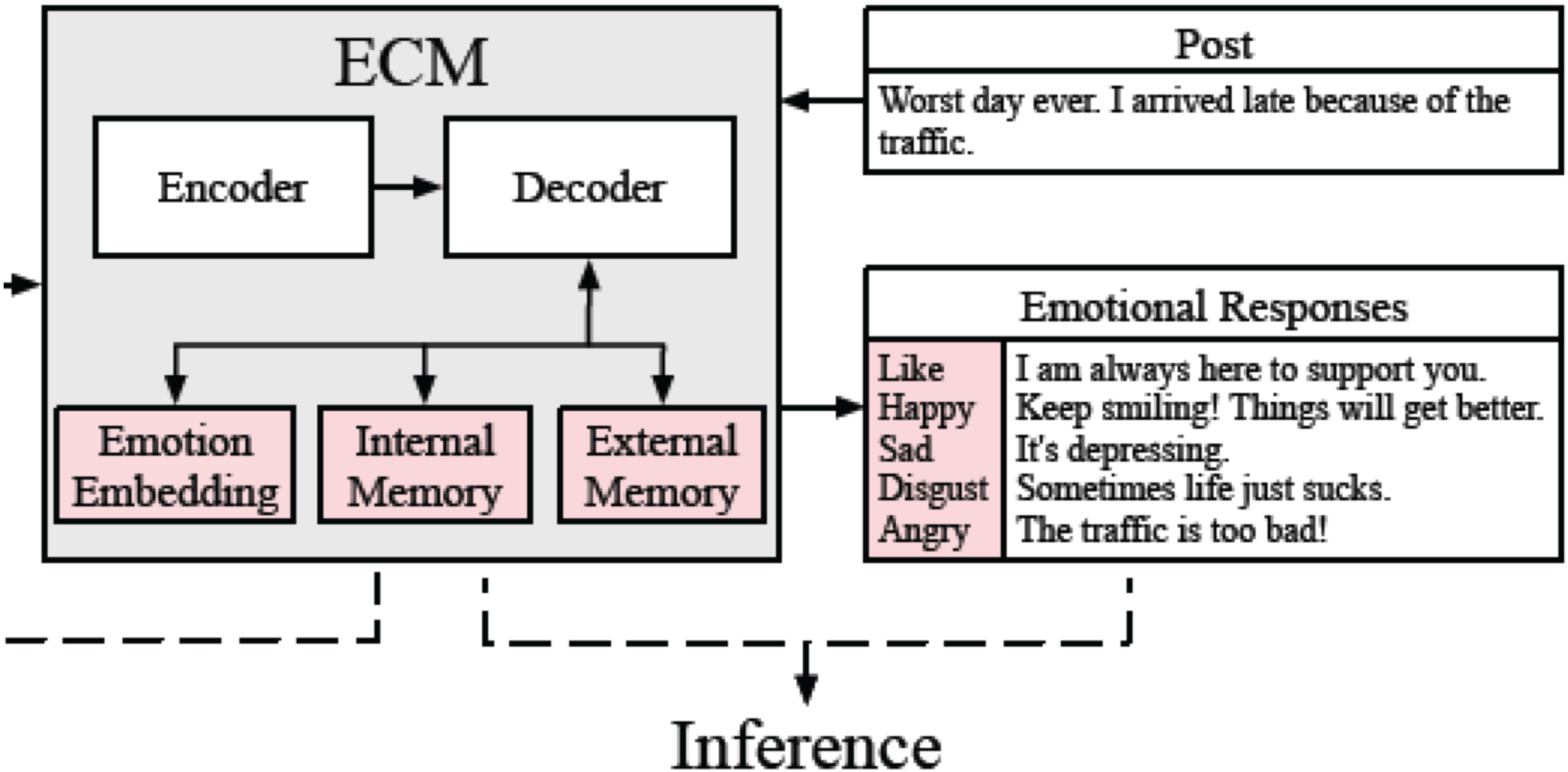
Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Overview of Emotional Chatting Machine (ECM)



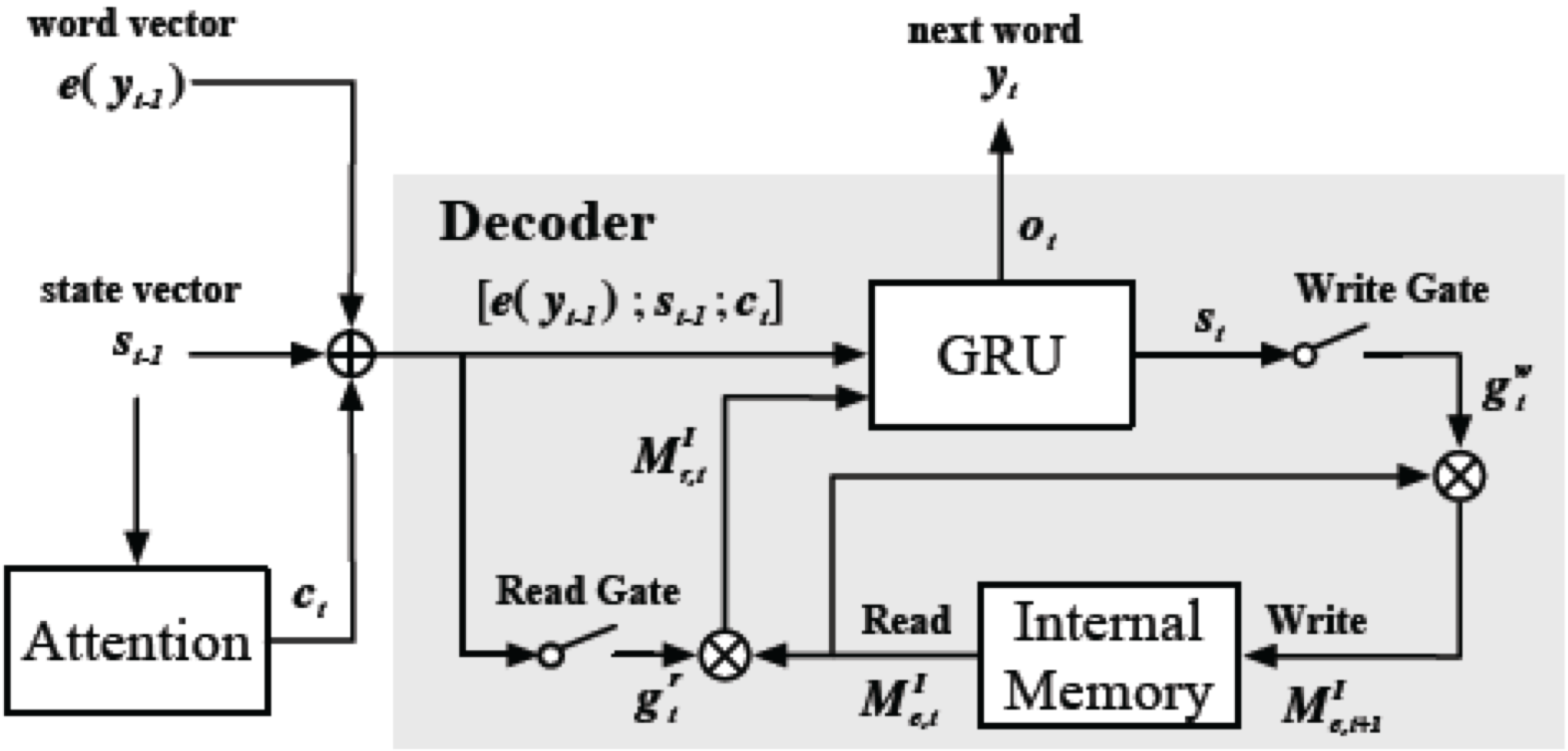
Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Overview of Emotional Chatting Machine (ECM)



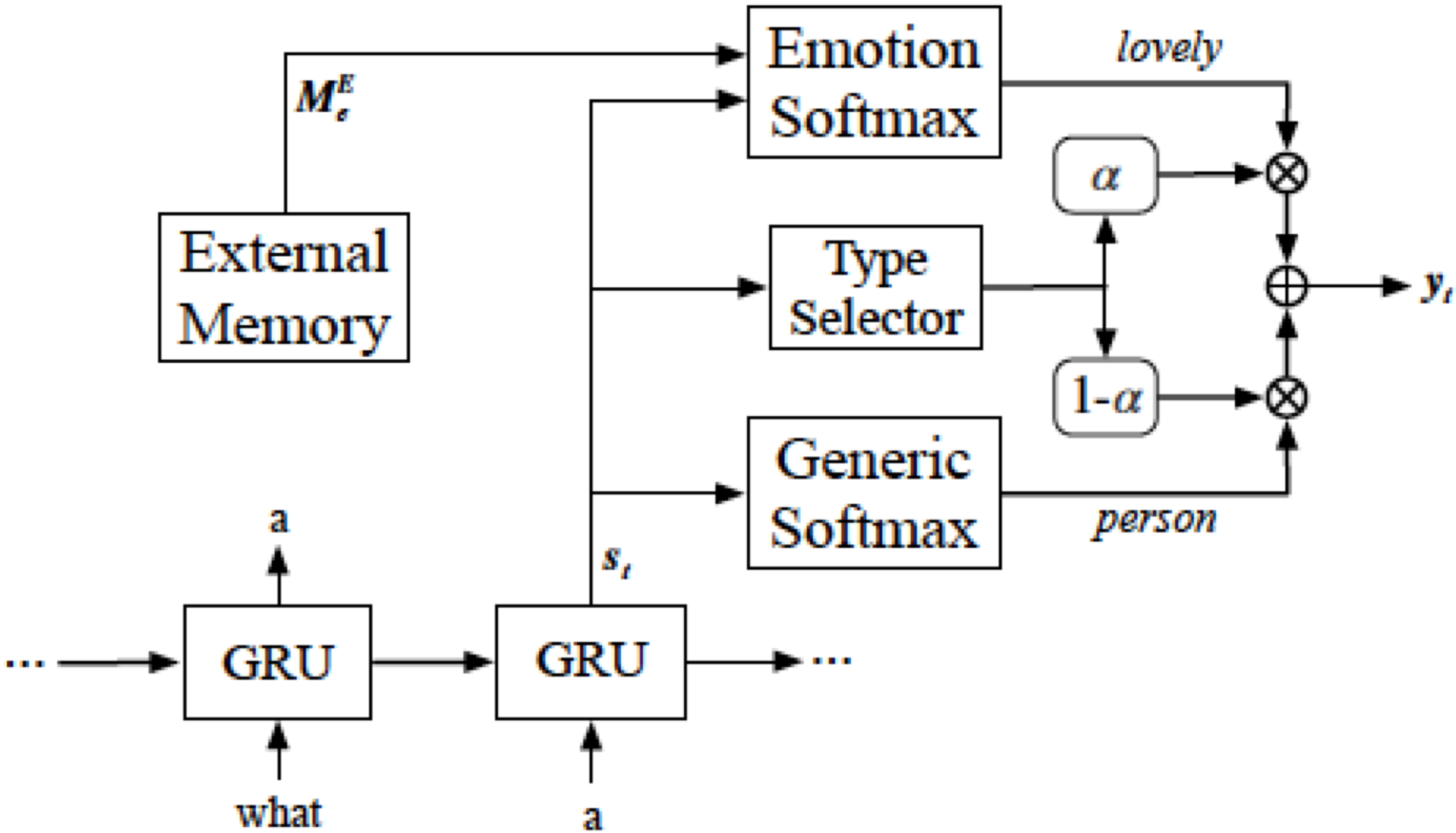
Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Data flow of the decoder with an internal memory



Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Data flow of the decoder with an external memory



Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Sample responses generated by Seq2Seq and ECM (Emotional Chatting Machine)

Post	Model		Response(Original)
知道真相的我眼泪笑下来 Knowing the truth, I laughed myself to tears	Seq2Seq		知道真相的我眼泪掉下来。
	ECM	Like	太可爱了，太可爱了！
		Happy	笑得我眼泪都出来了。
		Sad	知道真相的我眼泪掉下来。
		Disgust	这是真的还是假的？
		Angry	什么意思？没看懂。

Source: Zhou, Hao, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. "Emotional chatting machine: emotional conversation generation with internal and external memory." *arXiv preprint arXiv:1704.01074* (2017).

Chinese Emotional Conversation Generation (CECG) Evaluation Metric

- Emotion Consistency
  - whether the emotion class of a generated response is the same as the pre-specified class.
- Coherence
  - whether the response is appropriate in terms of both logically coherent and topic relevant content.
- Fluency
  - whether the response is fluent in grammar and acceptable as a natural language response.

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/evaluation_metric/

Chinese Emotional Conversation Generation (CECG) Evaluation Metric

IF Coherence and Fluency

    IF Emotion Consistency

        LABEL 2

    ELSE

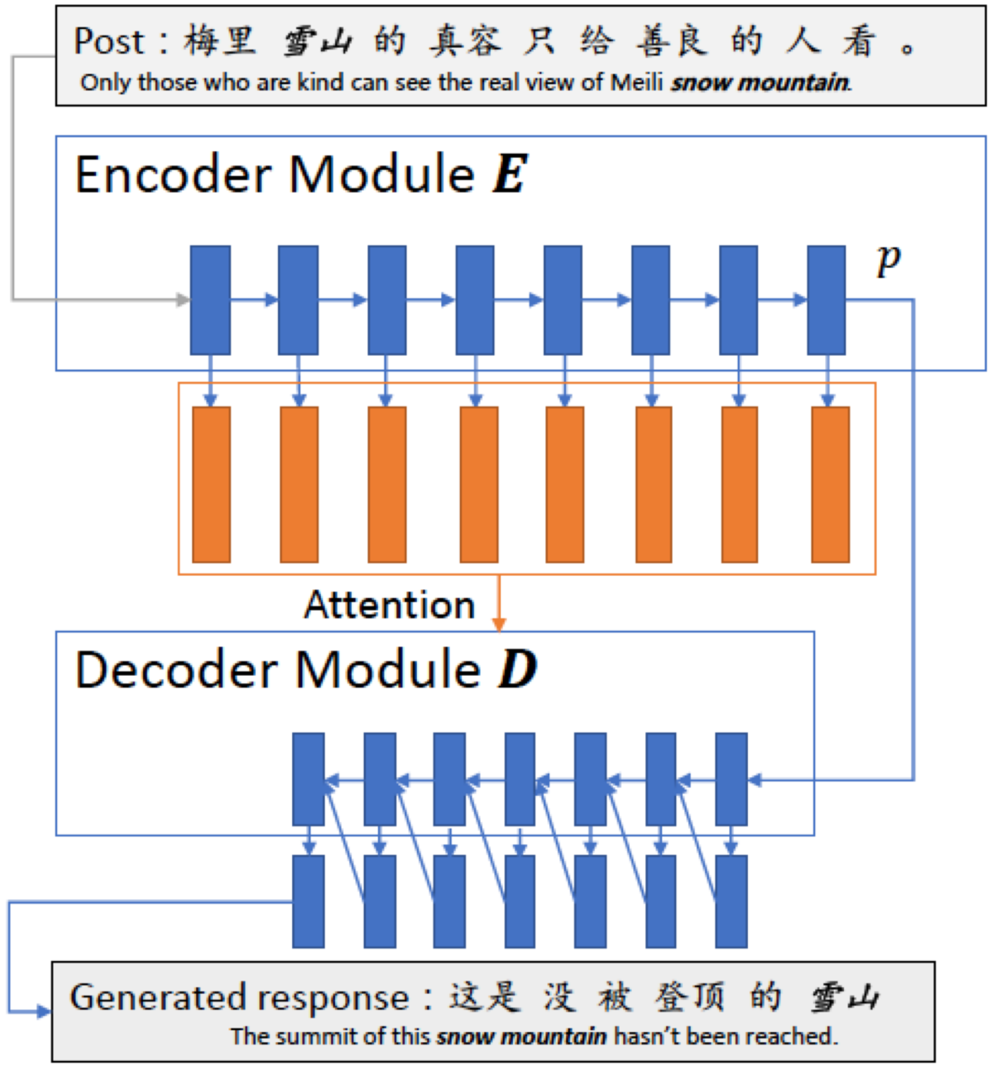
        LABEL 1

ELSE

    LABEL 0

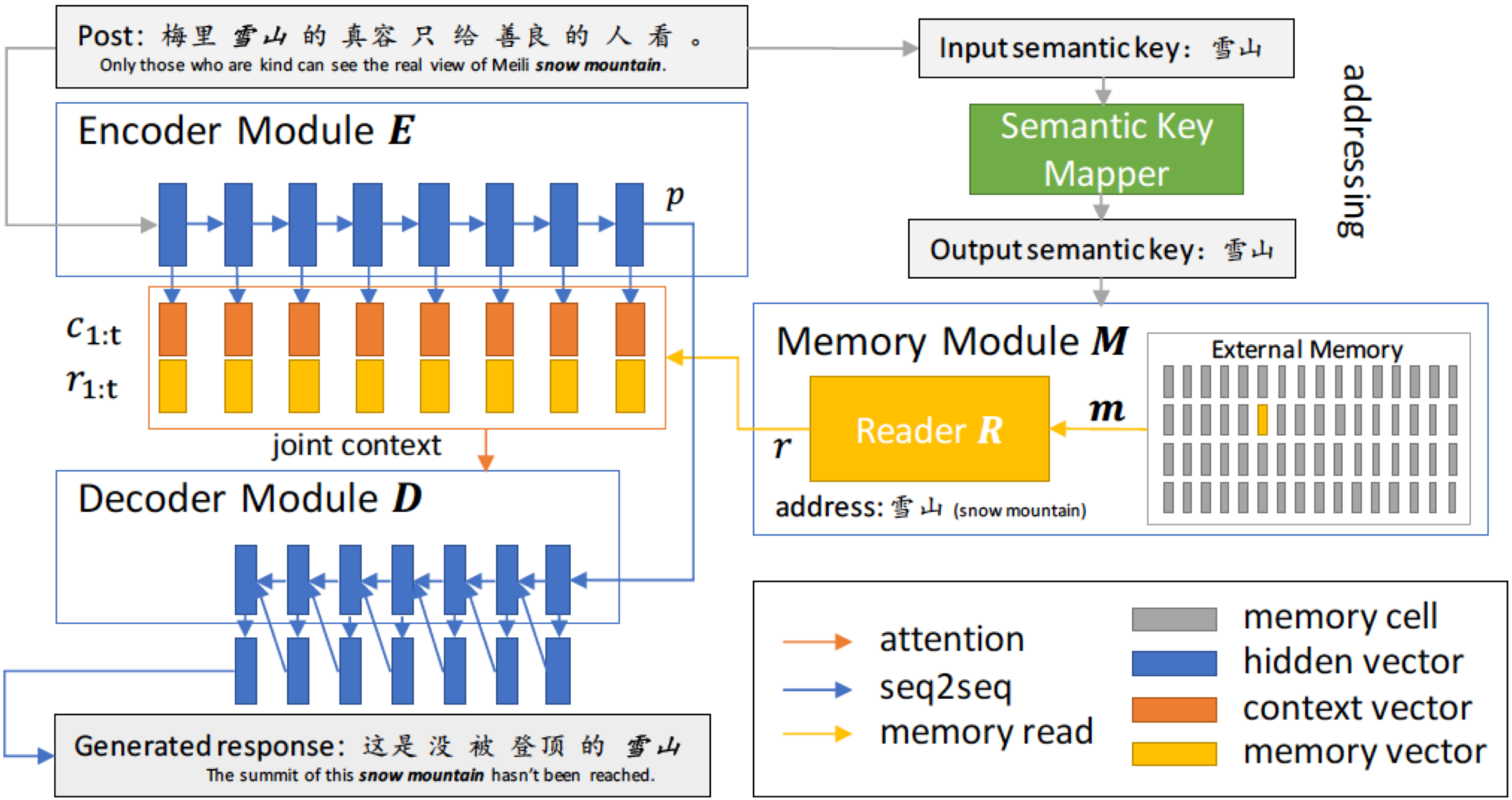
Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/evaluation_metric/

Sequence-to-sequence Learning with Attention for Generation-based STC



Source: Yu, Kai, Zijian Zhao, Xueyang Wu, Hongtao Lin, and Xuan Liu. "Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2018).

General Framework of Controllable Short-Text-Conversation Generation with External Memory



Source: Yu, Kai, Zijian Zhao, Xueyang Wu, Hongtao Lin, and Xuan Liu. "Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2018).

Controllable Short Text Conversation Examples

Post	李娜太牛了!中国第一个世界网球大满贯冠军! Li Na is great! China's first World Tennis Grand Slam champion!	
NRM	李娜是世界冠军! Li Na is the world champion!	
MrRNN	李娜加油,加油! Come on, Li Na. Come on!	
ESED	Key-Word	Response
	世界 world	李娜是世界上最棒的! Li Na is the best in the world!
	中国 China	李娜是中国的骄傲! Li Na is the pride of China!
	网球 tennis	中国网球公开赛,加油! China Tennis Open, come on!

Post	感人器官捐献广告,请感谢那些给你第二次生命的人. A touching organ donation advertisement, please thank those who gave you a second life.	
NRM	给我第二次生命的人。 A man who gives me a second life.	
MrRNN	感谢每一个人! Thanks for everyone!	
ESED	Key-Word	Response
	生命 life	感谢生命给予的一切! Thanks for everything given by life!
	器官 organ	感谢那些器官捐赠者。 Thanks for those organ donors.
	广告 advertisement	很感人的广告! A very touching advertisement!

Source: Yu, Kai, Zijian Zhao, Xueyang Wu, Hongtao Lin, and Xuan Liu. "Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2018).

Comments Generated Using Different Semantic key Mapping Methods

Post	美好的社会应该由善良的人组成。 A good society should be composed of people of kindness.	
Mapping-Method	Key-Word	Response
Identity Mapper	善良 kindness	善良是一种信仰。 Kindness is a belief.
Synonym	友善 friendliness	善良是一种美德。 Kindness is a virtue.
Antonym	丑恶 evil	没有丑恶的社会,更可怕。 A society without evils is more terrifying.
Embedding	乐观 optimism	乐观是一种生活态度。 Optimism is an attitude of life.
Common-Reply	真 really	真的吗? Really?
	对 right	对的,对的。 Right, right.
Random	人 person	每个人都有善良的一面。 Every person has a side of kindness.
	中国 China	中国的贫富差距太大了。 The gap between the wealthy and the poor in China is too big.

Source: Yu, Kai, Zijian Zhao, Xueyang Wu, Hongtao Lin, and Xuan Liu. "Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2018).

Generated Responses of Knowledge Introduction by External Memory

Post	你最喜欢那个星座? What is your favorite constellation?	
Semantic-Key	Key-Word	Response
	摩羯座 Capricorn	我最喜欢摩羯座。 I love Capricorn.
	射手座 Sagittarius	我是射手座,我喜欢。 I am a Sagittarius and I like it.
Post	今天周杰伦的电影,太好看了。 Jay Chou's movie is so nice today.	
Semantic-Key	Key-Word	Response
	歌 song	杰伦的歌也很好听。 Jay Chou's songs are nice too.
Post	梵高的画,很漂亮。 Van Gogh's paintings are very beautiful.	
Semantic-Key	Key-Word	Response
	向日葵 sunflower	梵高笔下的向日葵。 The Sunflower by Van Gogh.
	荷兰 Holland	梵高的画,很漂亮。 Van Gogh's paintings are very beautiful.

Source: Yu, Kai, Zijian Zhao, Xueyang Wu, Hongtao Lin, and Xuan Liu. "Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2018).

Welcome to NTCIR-14 STC-3

- CECG
  - Aug 31, 2018 CECG registration due
  - Sep 15, 2018 Training data released
  - Nov 15, 2018 Test data released
  - Nov 25, 2018 Run submissions due
- ALL
  - Dec 20, 2018 Evaluation results and draft overview released
  - Feb 1, 2019 Participant paper submissions due
  - Mar 1, 2019 Acceptance notification
  - Mar 20, 2019 All camera-ready papers due
  - Jun 2019 NTCIR-14 Conference & EVIA 2019 in NII, Tokyo

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge/time_schedule/

Summary

- AI
- Big Data
- Cloud Computing

References

- Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson.
- Jared Dean (2014), Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, Wiley.
- Mehmet Kaya, Jalal Kawash, Suheil Khoury, and Min-Yuh Day (2018), Social Network Based Big Data Analysis and Applications, Lecture Notes in Social Networks, Springer International Publishing.
- Varun Grover, Roger HL Chiang, Ting-Peng Liang, and Dongsong Zhang (2018), "Creating Strategic Business Value from Big Data Analytics: A Research Framework", Journal of Management Information Systems, 35, no. 2, pp. 388-423.
- Ting-Peng Liang and Yu-Hsi Liu (2018), "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Expert Systems with Applications, 111, no. 30, pp. 2-10.
- Stuart Russell and Peter Norvig (2016) , Artificial Intelligence: A Modern Approach, 3rd Edition, Pearson International.
- Javier Mata, Ignacio de Miguel, Ramón J. Durán, Noemí Merayo, Sandeep Kumar Singh, Admela Jukan, and Mohit Chamania (2018), "Artificial intelligence (AI) methods in optical networks: A comprehensive survey", Optical Switching and Networking, 28, pp. 43-57
- Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications.