Session 1

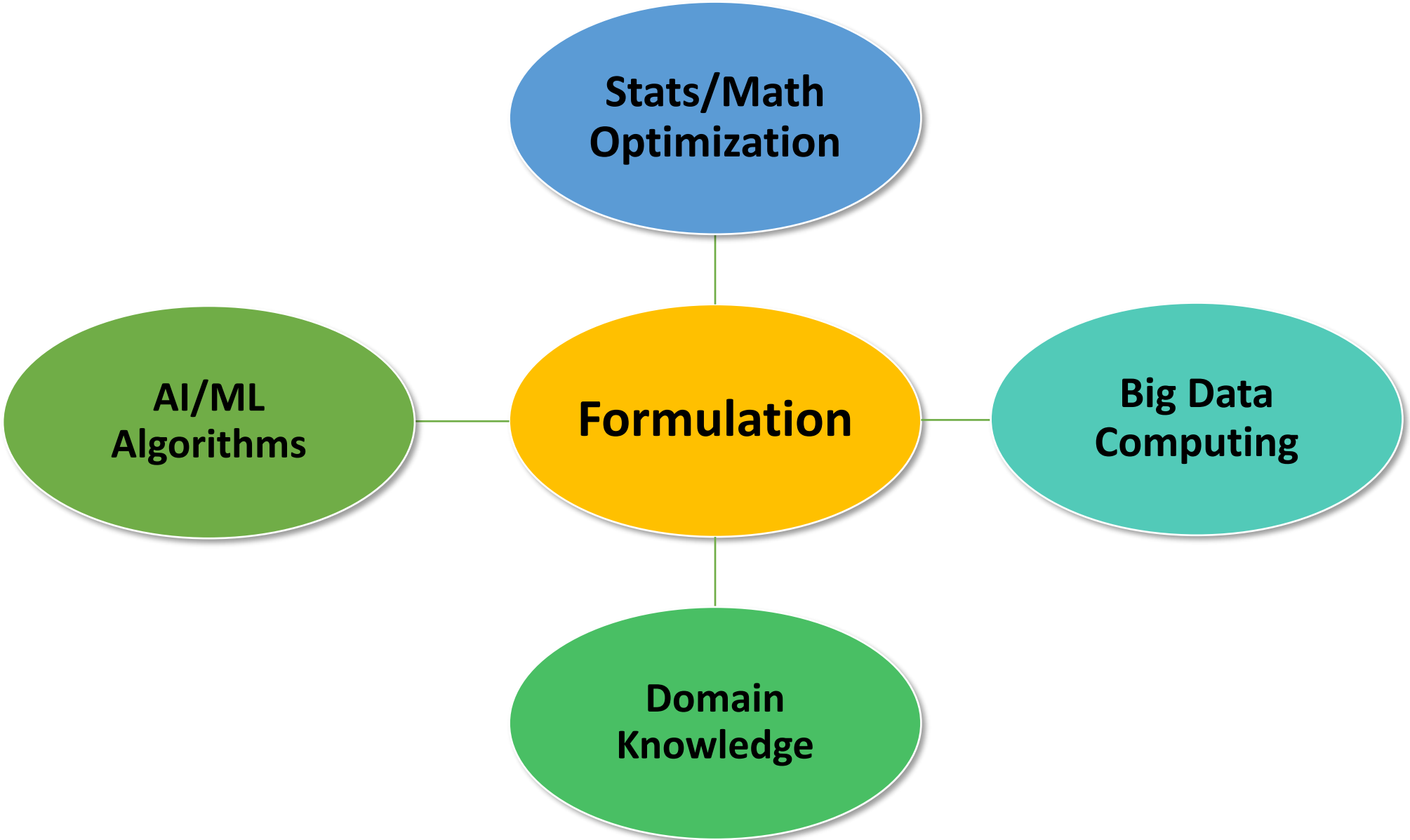
INTRODUCTION TO MACHINE LEARNING

Dr. Shailesh Kumar

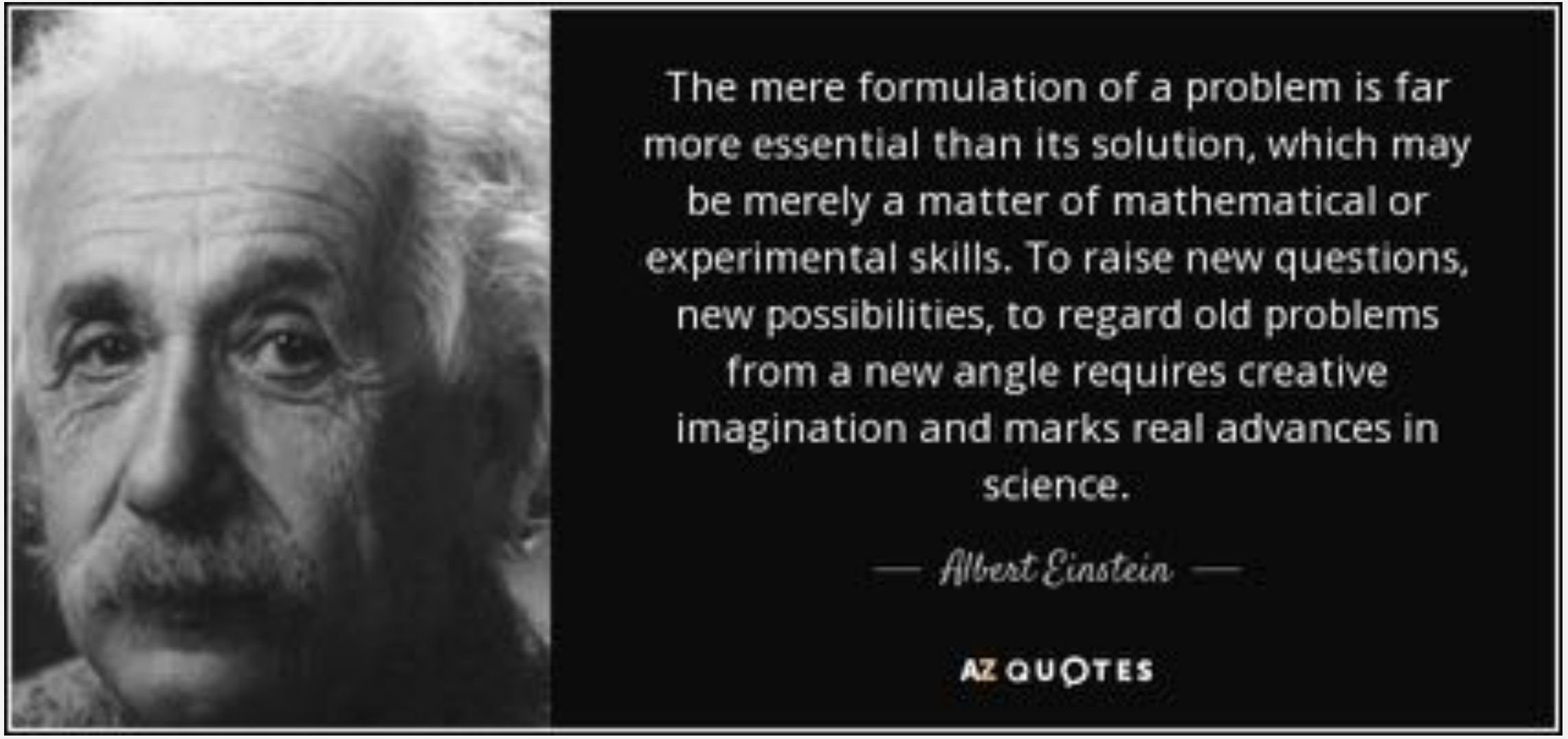
What makes a great **PRODUCT** today?

Interface	Intuitive |Functional | Elegant
Infrastructure	Storage |Computation |Network
Intelligence	Learn |Predict | Adapt | Evolve
Integrated	Connected | Complex | Ecosystem

What **SKILLS** future data scientists need?



FORMULATION is the Key Skill in Data Science



What **ATTITUDE** good data scientists have?

The Most elementary and valuable statement in Science – the beginning of Wisdom is –

“I do not know!”

Evolution of Technology

What is **TECHNOLOGY**?

"Our **technology**, our machines, is **part of our humanity**. We created them to **extend ourselves**, and that is what is **unique** about human beings!".

-- Ray Kurzweil

CONVERGENCE of Technologies

INTERNET

CONNECTIVITY

DEVICES

CLOUD

ML / DL / AI

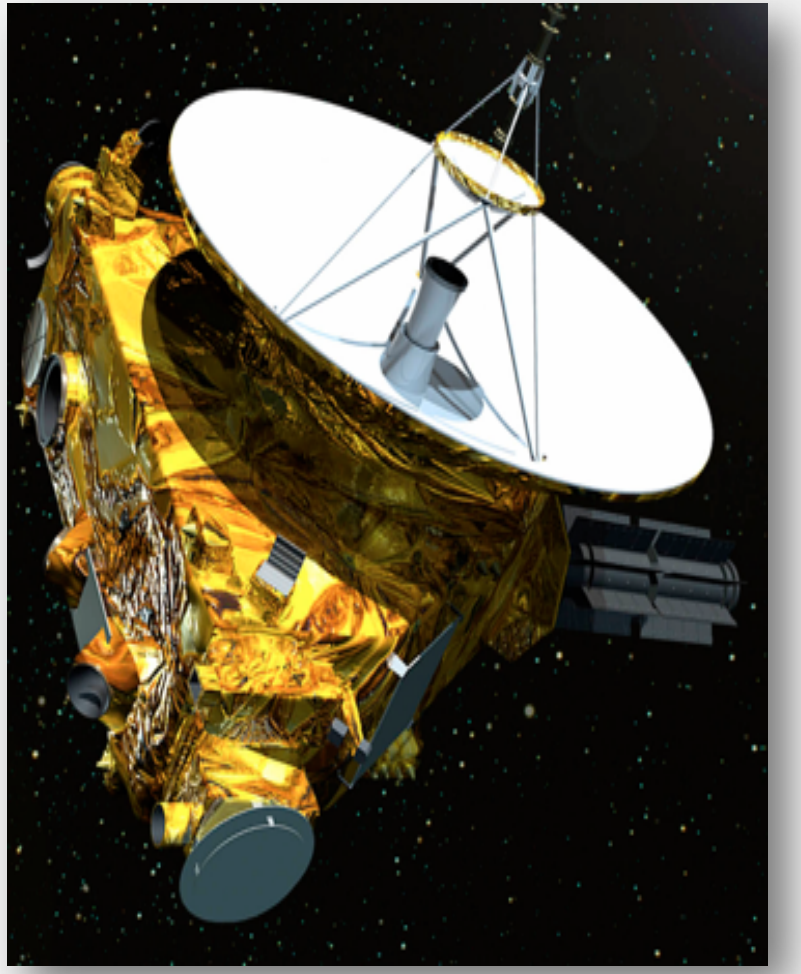
INTERNET OF THINGS

All **ASPECTS** of Technology are evolving

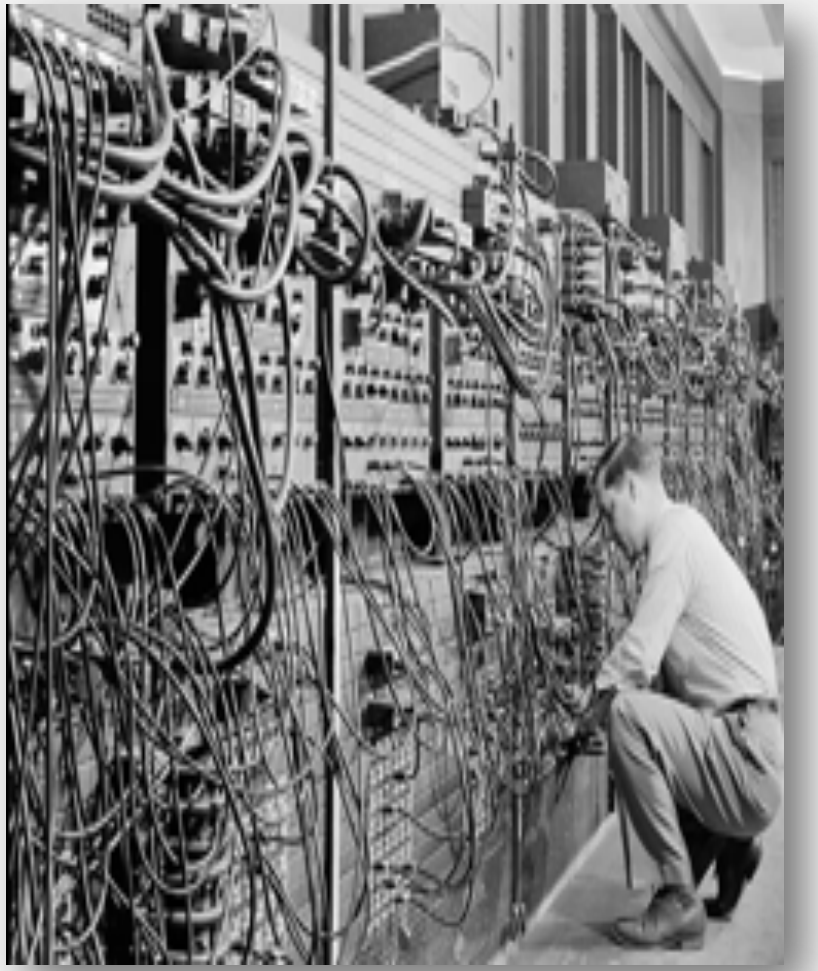
Tools	Computing	Interfaces
Sensors	Data	Intelligence
Products	AI Thinking	Data Science Skills

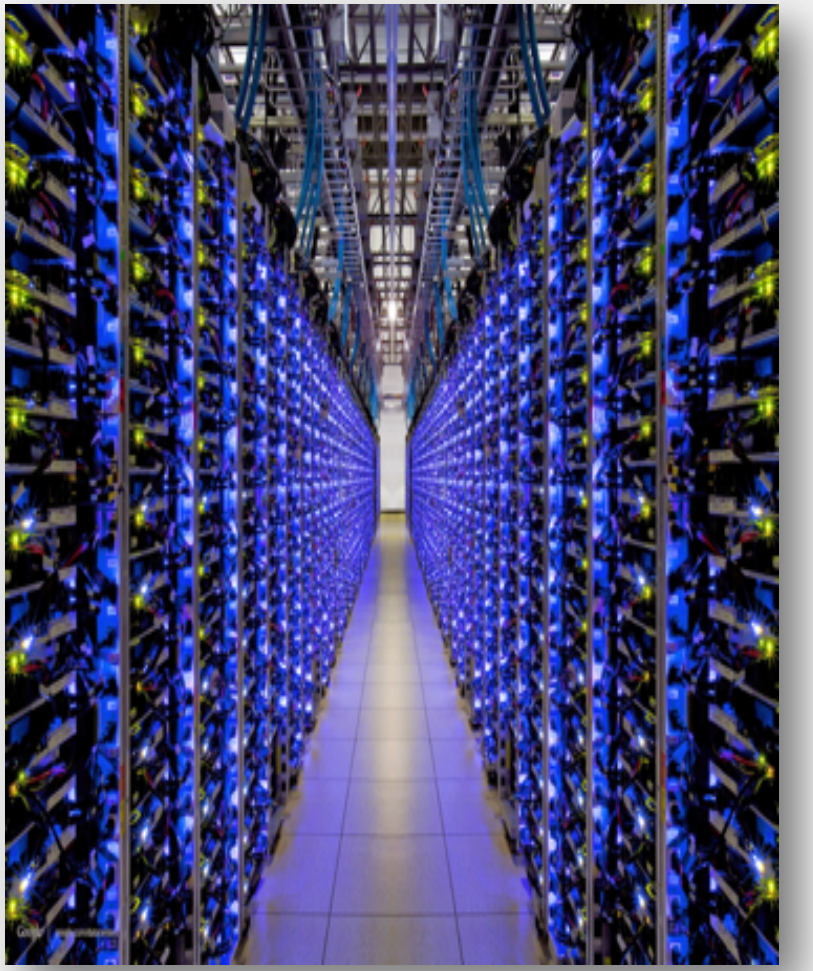
TOOLS are evolving

SENSORS are evolving

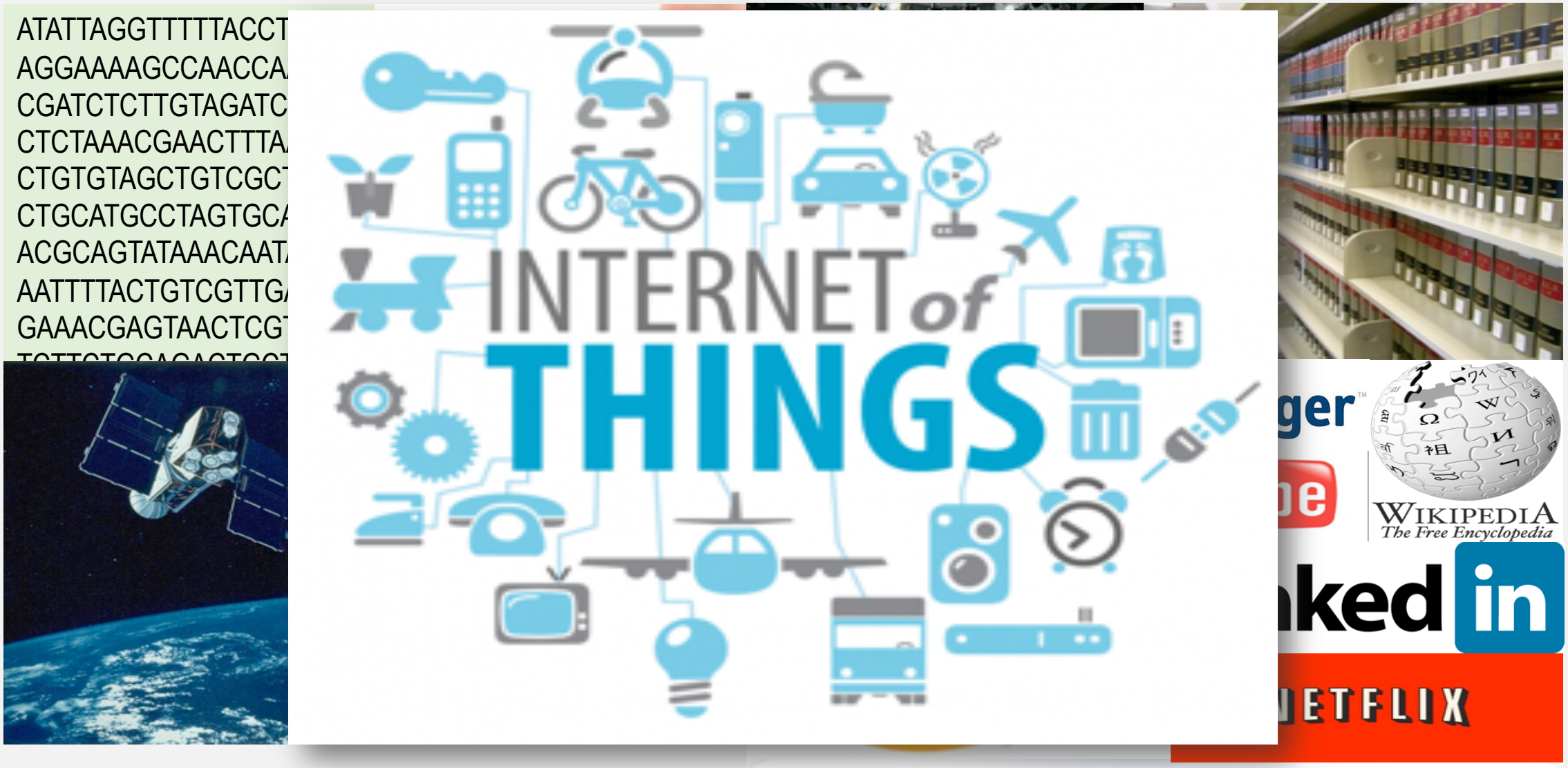


COMPUTING is evolving



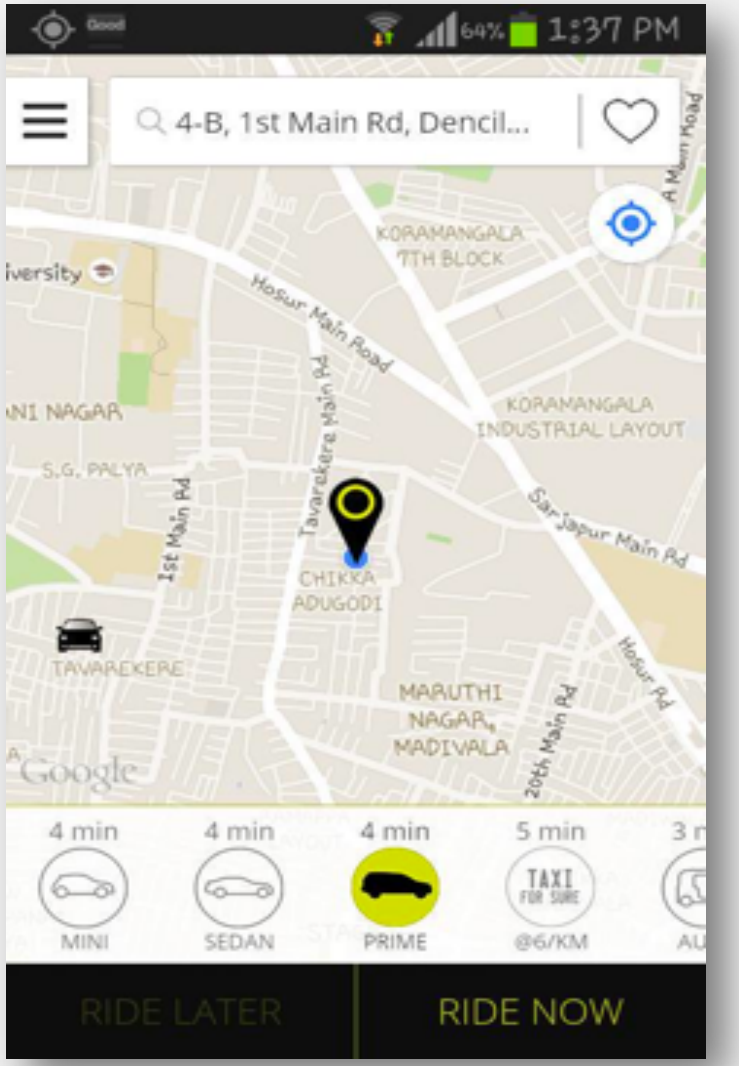


DATA is evolving

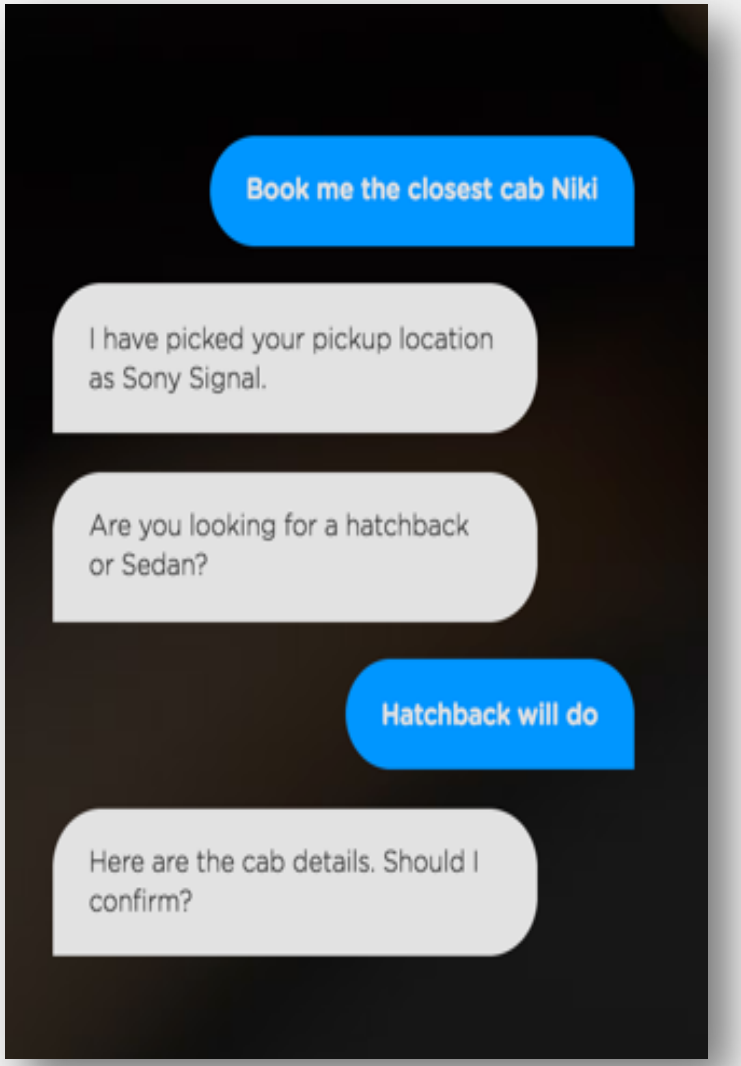


INTERFACES are evolving

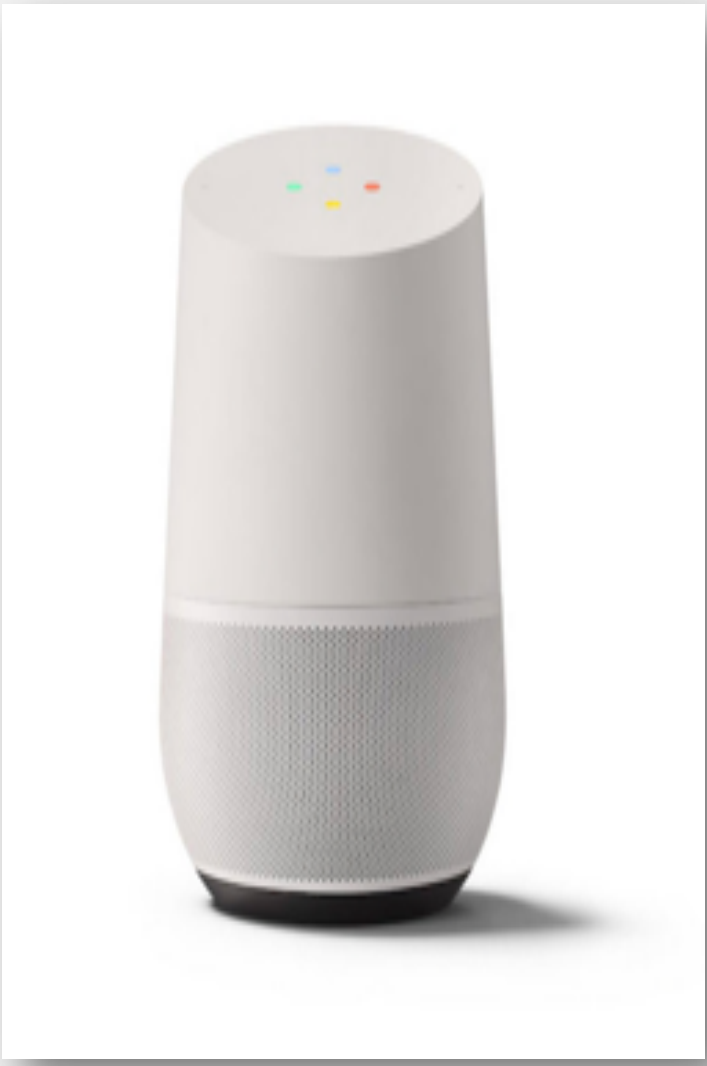
TOUCH



CONVERSATIONAL



SPEECH



PRODUCTS are evolving

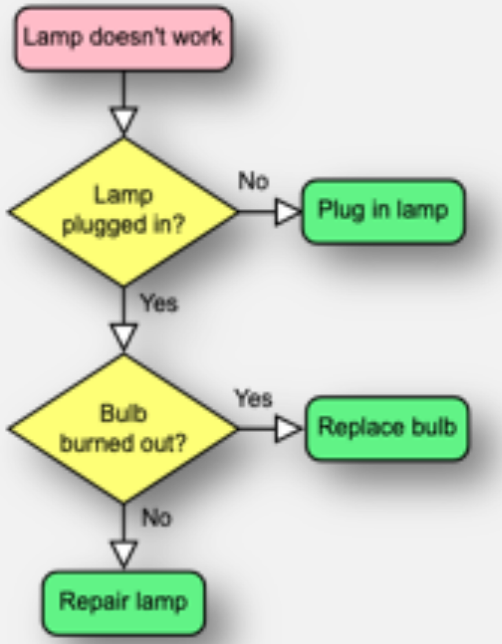
Google in

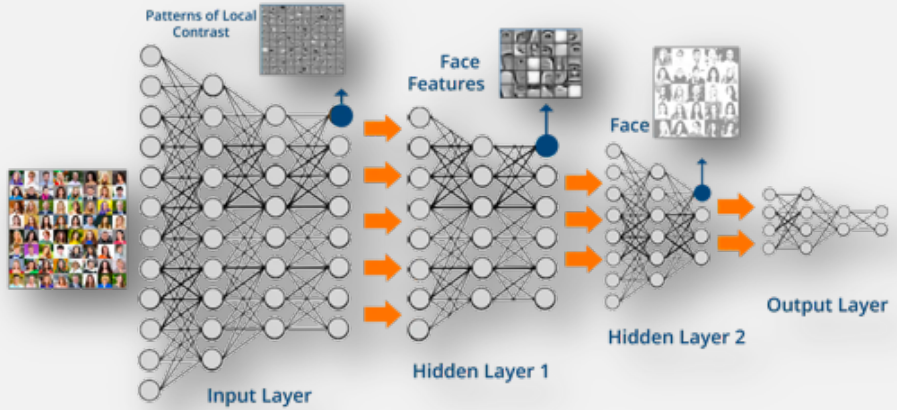
YouTube

PRODUCT 1.0 (Physical | Reliable)
PROGRAMMED INTELLIGENCE

PRODUCT 2.0 (Digital | Scalable)
CROWD INTELLIGENCE

PRODUCT 3.0 (Hybrid | Integrated)
EMERGENT INTELLIGENCE





- System of **SENSORS**
- System of **INTERACTIONS**
- System of **RECORDS**
- System of **INTELLIGENCE**
- System of **OPERATIONS**
- System of **METRICS**

Ecosystem Products of the **FUTURE**

Philosophy of AI

Understanding the Nature of the Mind

ML Goal 1 | Unravelling the **NATURE** of the Mind

CONSCIOUSNESS….Awareness? Sentience?

CREATIVITY…….Synthesis? Manifestation?

THINKING……Reasoning? Deduction?

UNDERSTANDING…….Semantics? Analysis?

LEARNING……Syntax? Prediction?

What is **INTELLIGENCE**?

Generalization: Ability to **predict** or assign a label to a "**new**" observation based on the **model** built from **past experience**

What is **UNDERSTANDING**?

Understanding: Ability to assign **meaning** to all the **parts** simultaneously based on the **context** so the **whole** makes sens!

What is the **MEANING** of “meaning”?

Disambiguation

I was **right** to avoid a **suit** against **apple**

Man in blue **suit** on my **right** was drinking **apple** juice

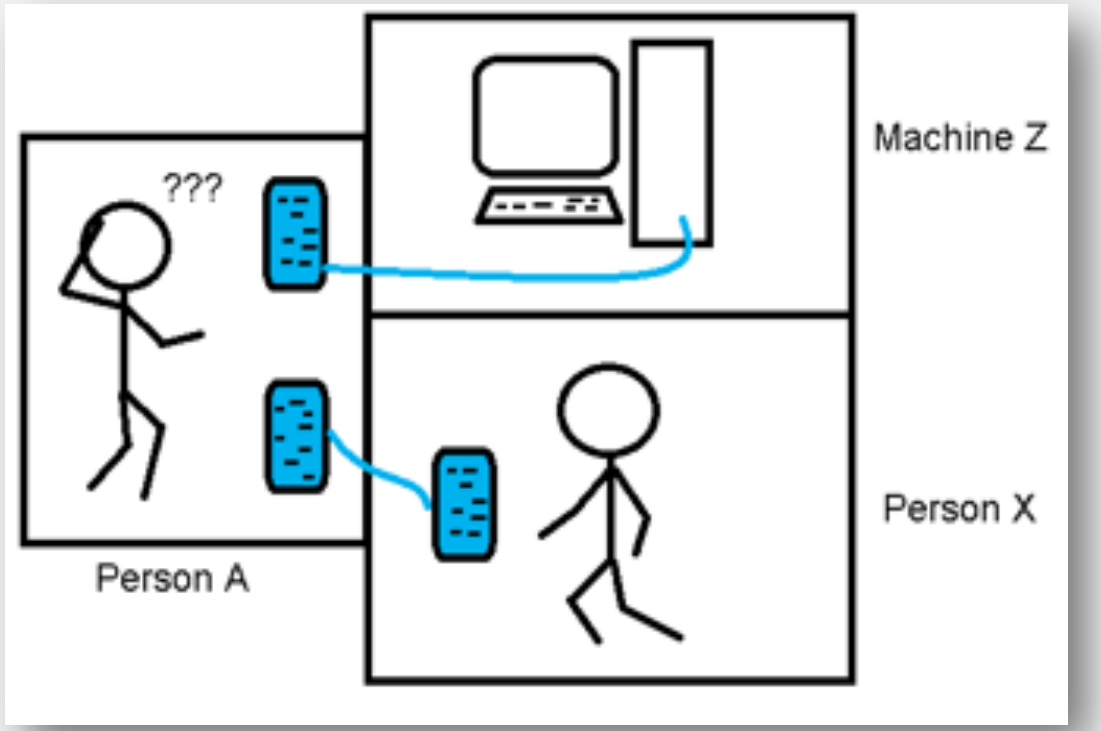
Equivalencing

Apple filed a **suit charging orange** of **illegal behavior**

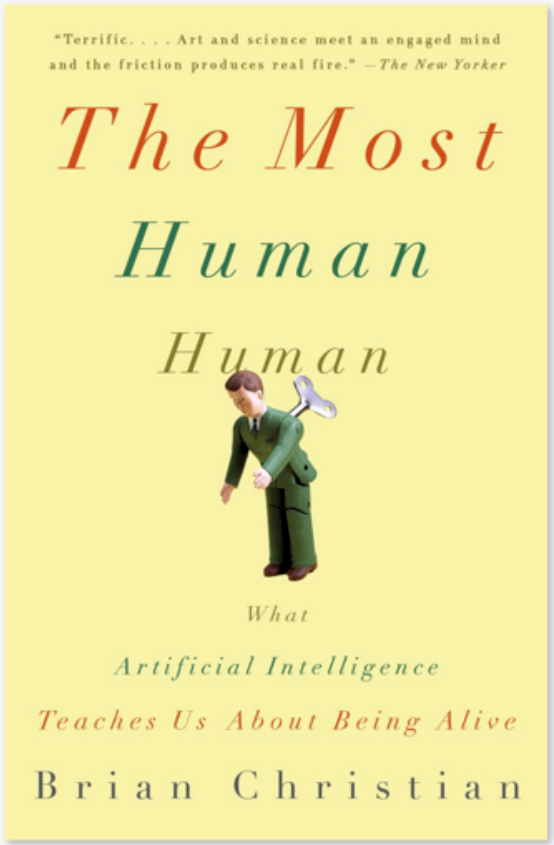
Orange submitted a **case accusing apple** of **unauthorized conduct**

A **TEST** of Intelligence

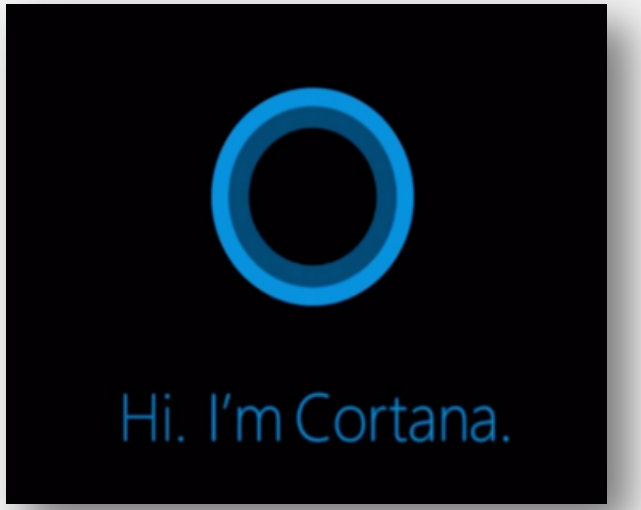
"A computer would deserve to be called intelligent if it could deceive a human into believing that it was human". *-- Alan Turing*



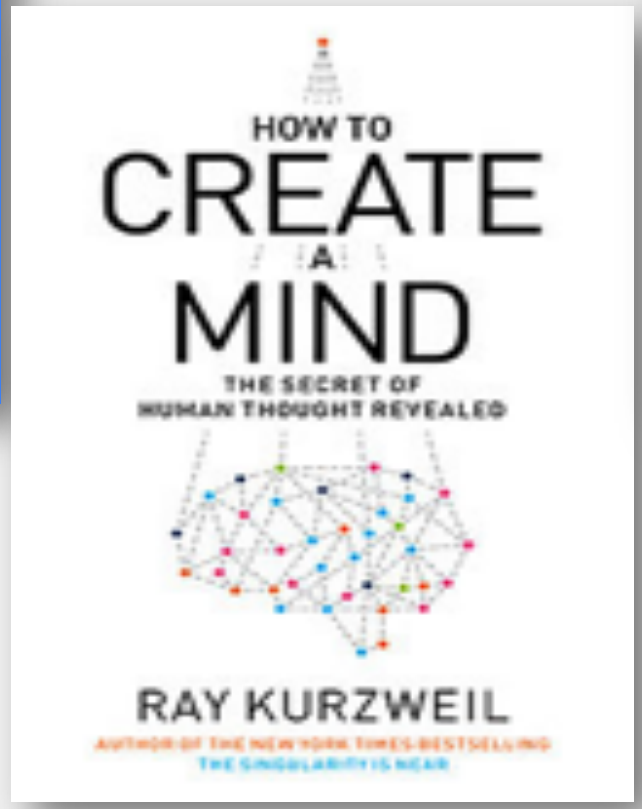
The Turing Test

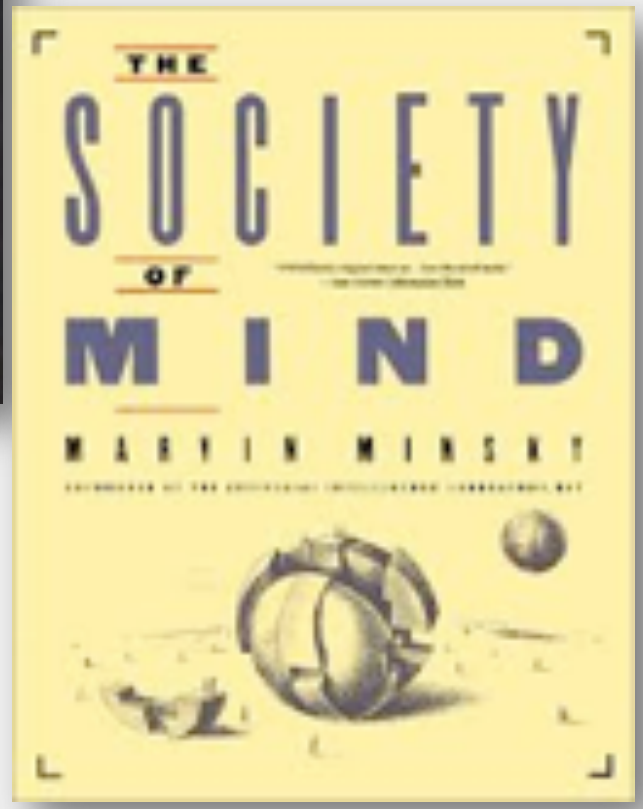


CHATBOT Season!



Further Readings on the **PHILOSOPHY** of AI

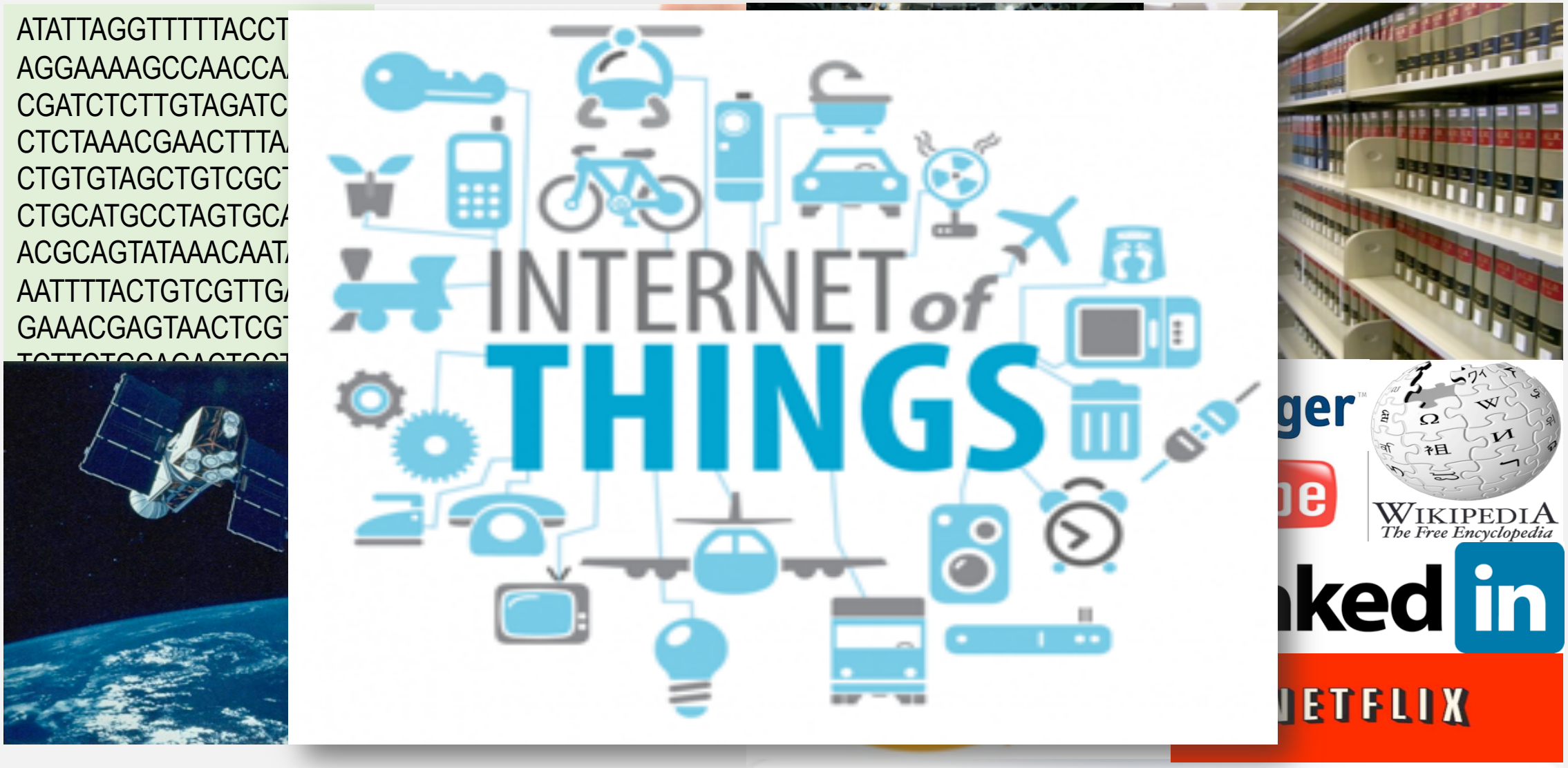




Application of AI

Making “Better” Decisions

Drowning in **DATA**, Starving for **KNOWLEDGE**



How BIG is Big Data?

600 million

tweets per **DAY**

100 hours of

videos per **MINUTE**

800+ websites

per **MINUTE**

100 TB of data

uploaded **DAILY**

3.5 Billion

Queries PER **DAY**

30 Million

Orders per **DAY**

How did we get here?

- **Collection** – better “sensors”
  - Scanner, Point of Sale, Medical Sensors, Crowd Sourcing,…
- **Communication** – faster “transfer”
  - Network infrastructure, Compression Technologies
- **Storage** – cheaper “warehouses”
  - Cloud based storage, large warehouses, NoSQL databases
- **Computation** – massive “distributed computing”
  - Distributed and Parallel Computing, Mapreduce/Hadoop
- **Decisions** – intelligent “decisions”
  - Advances in Machine Learning and Artificial Intelligence

DECISIONS are everywhere

- Which **position** should a **Page/Ad** be shown for a **query**?
- Should the **home loan/credit card transaction** be approved?
- Which **Video** to **show next** on **YouTube**?
- Who might user X **connect/follow** on **LinkedIn/Twitter**?
- What **articles/movies/songs** customer X might be interested in?
- Which **coupons/offers** will a **customer respond** to here/now?
- Where to **open** the next store? How to **price** the next gadget?
- Which **medicine/treatment** will **cure** this disease?
- Who will be the right **person (resume)** for this **Job (description)**?
- Which **crop** is best for this **soil/weather**?
- Which **concepts** is this **student** ready to learn next?

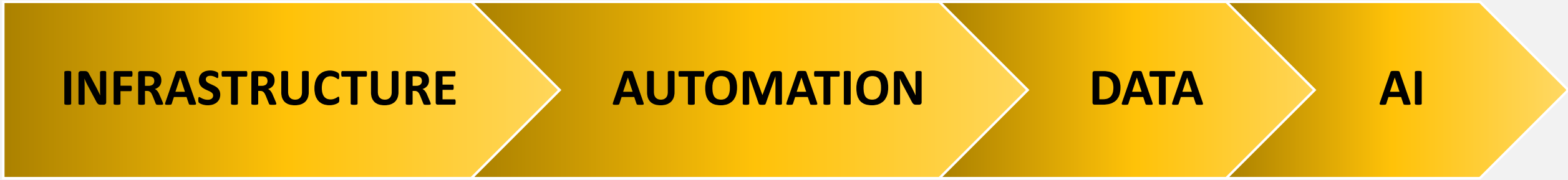
ML Goal 2 | Making “better” **DECISIONS**

- **BROADCAST** → **PERSONALIZED** decisions
  - Avoids “death by **average**”
- **BATCH** → **REAL-TIME** decisions
  - Avoids “death by **delay**”
- **REACTIVE** → **PROACTIVE** decisions
  - Avoids “death by **ignorance**”
- **TACTICAL** → **STRATEGIC** decisions
  - Avoids “death by **nearsightedness**”

AI Mindset

AI > IT

A Natural **PROGRESSION** of everything…

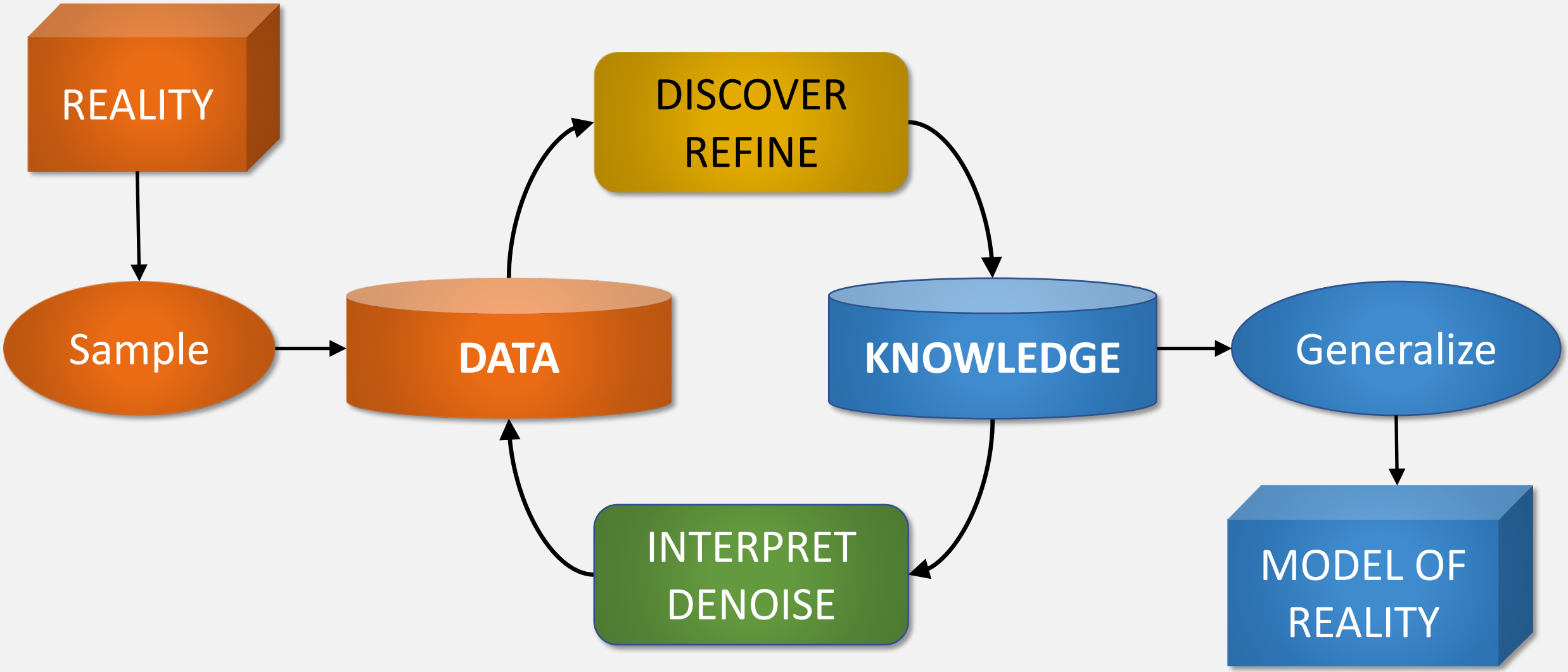


How do we go from **DATA RICH** to **AI-FIRST**?

Four **STAGES** of AI Adoption

EXPLORING ALL FUTURES (artificial intelligence)	How to maximize her life-time value?
CREATING A NEW FUTURE (data sciences)	What can we do about it?
PREDICTING THE FUTURE (machine learning)	Is this customer about to Churn? Why?
REPORTING THE PAST (business intelligence)	What are my customers doing?

AI is a **LEARNING LOOP**

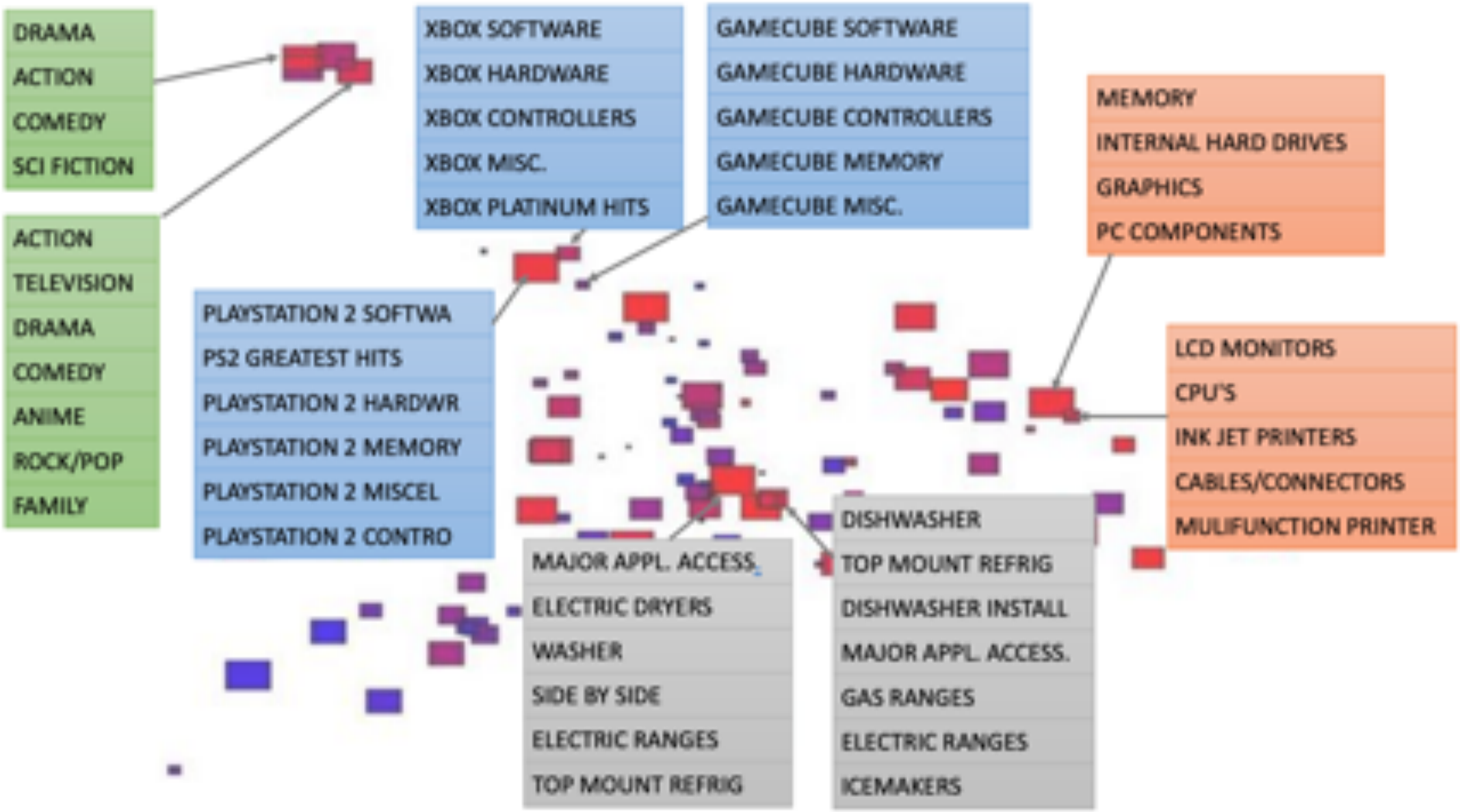


A Zoo of Algorithms

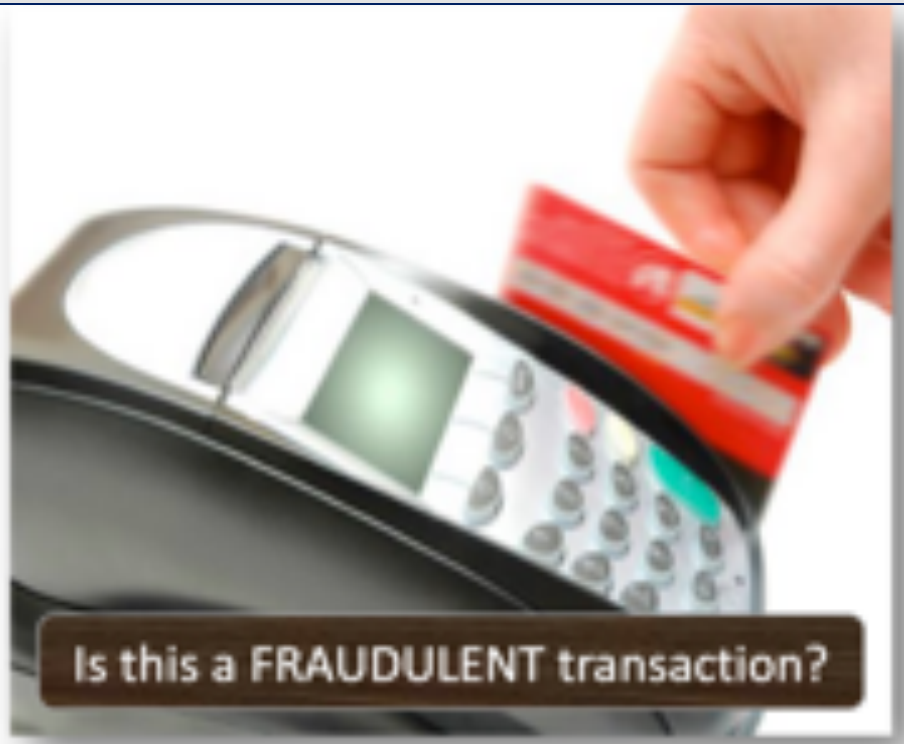
AI is at a 4-year old!

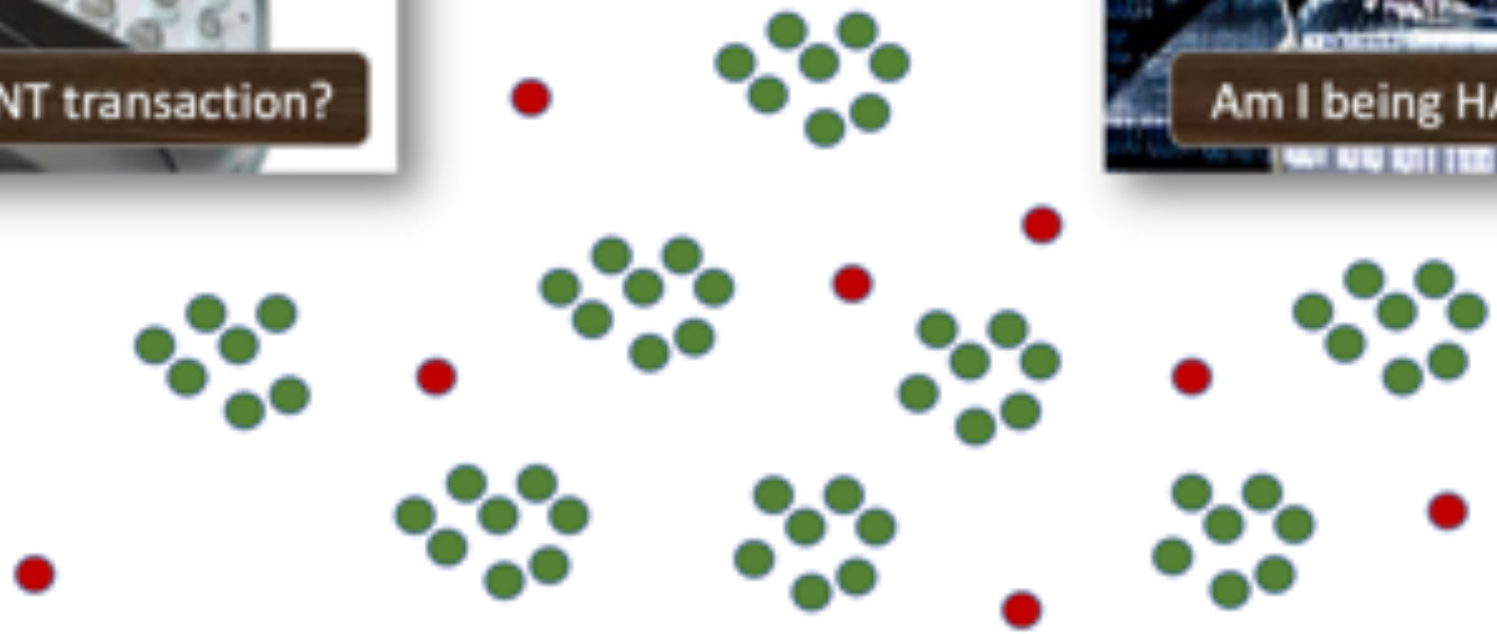
Data	Insights
Rules	Models
Hearing	Listening
Looking	Seeing
Reading	Understanding
Moves	Strategies

CLUSTERING paradigm

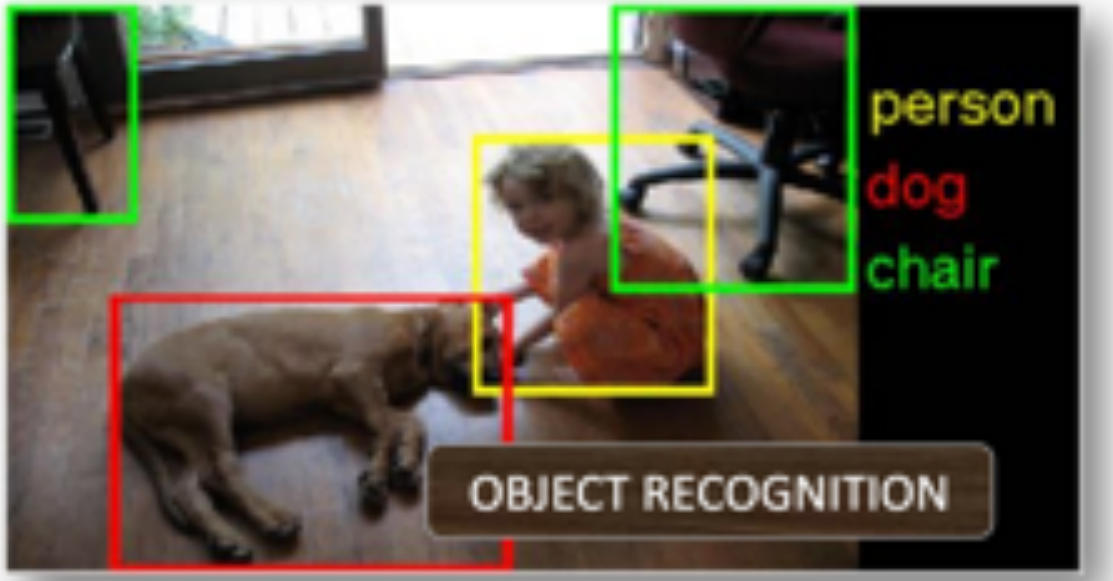


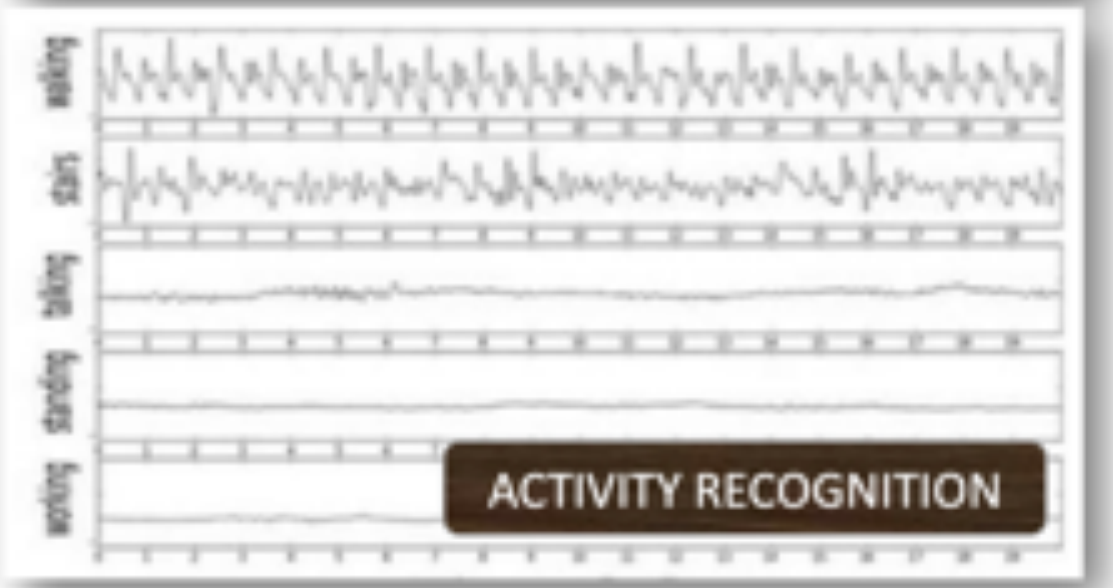
OUTLIER DETECTION paradigm





INTERPRETATION paradigm





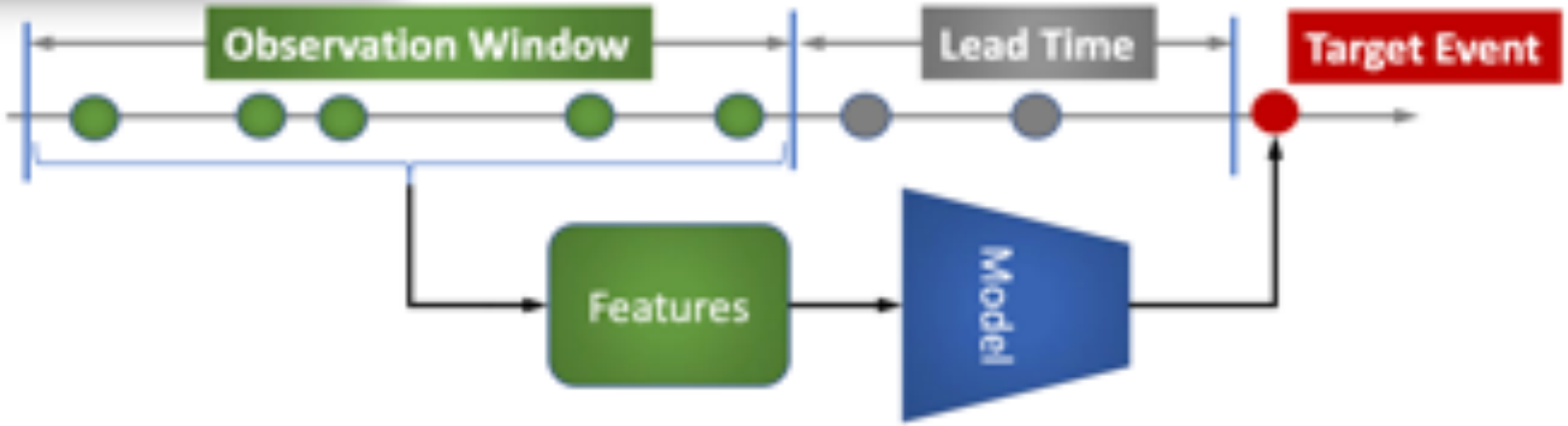
Entity | Aspect | Sentiment

"I had ordered this today ie. on 14'May at 12:10 AM and got it delivered by 01:00 PM using same day delivery option. **Flipkart** as usual **delivers within no time** and **really worth it** paying Rs.140 for **same day delivery**.

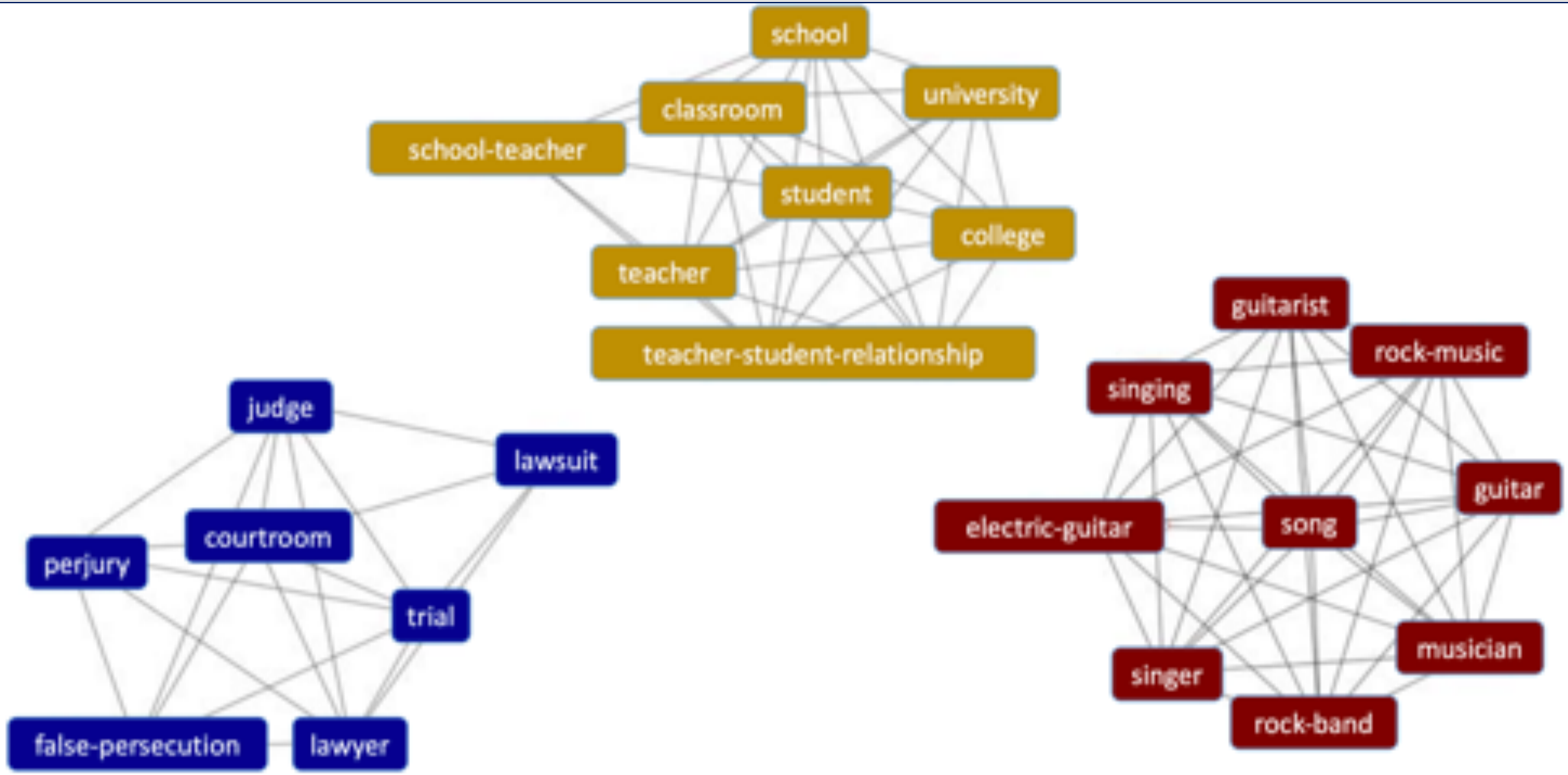
Coming to **Moto E**, I am sure no other branded mobiles are available at this **price range** and with this specs. All the sepcs related to **RAM, processor, Gorilla glass** and mostly **screen resolution** are **good**. **I recommend this** personally **comparing the price and spec**. I feel it had most but only **flash light is missing which if present would have nothing to be improved**. It is really good and happy to share this."

SENTIMENT ANALYSIS

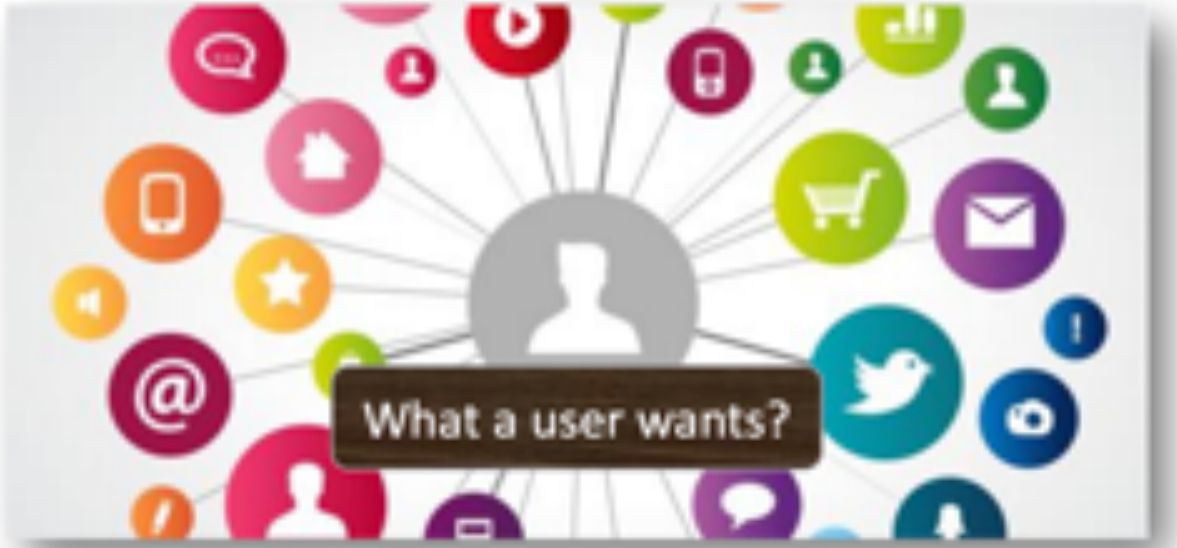
PREDICTION paradigms

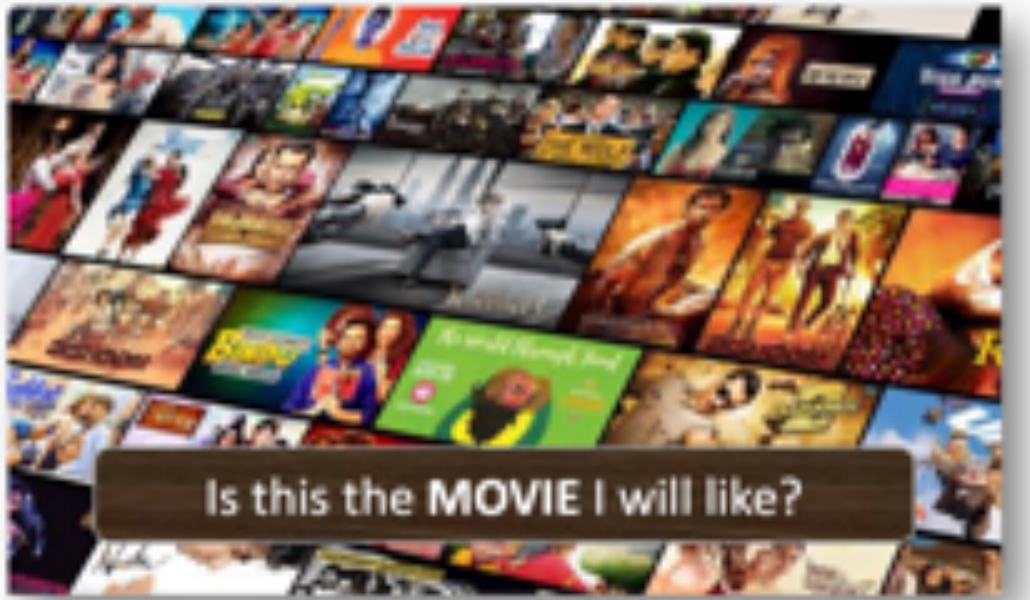


PATTERN RECOGNITION paradigms

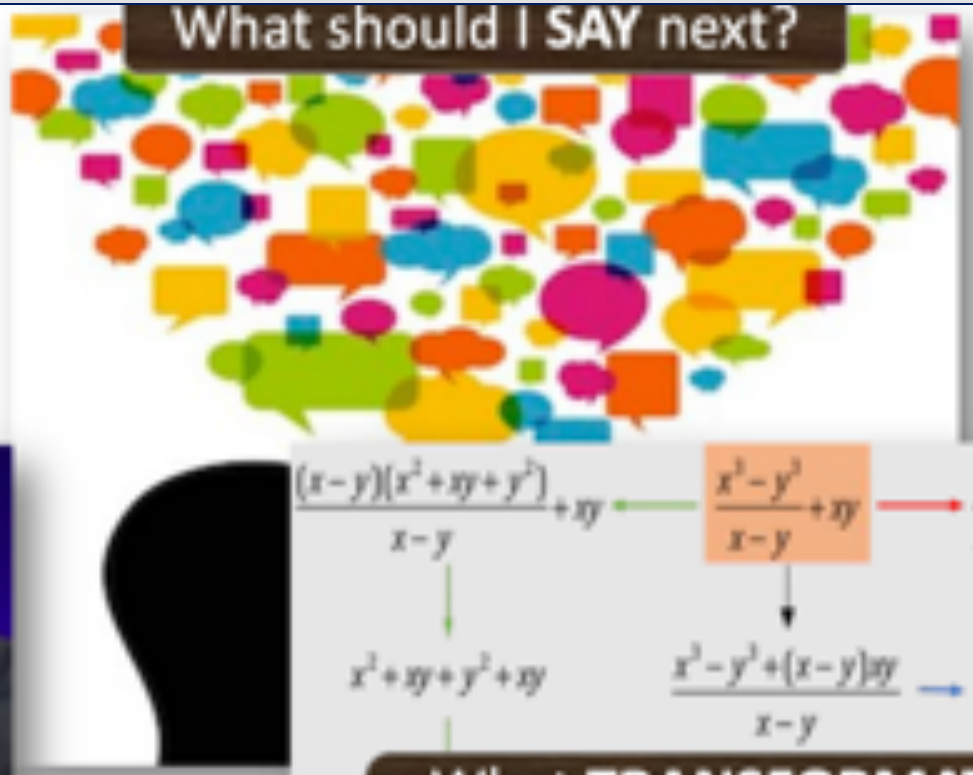


RECOMMENDATION ENGINE paradigm

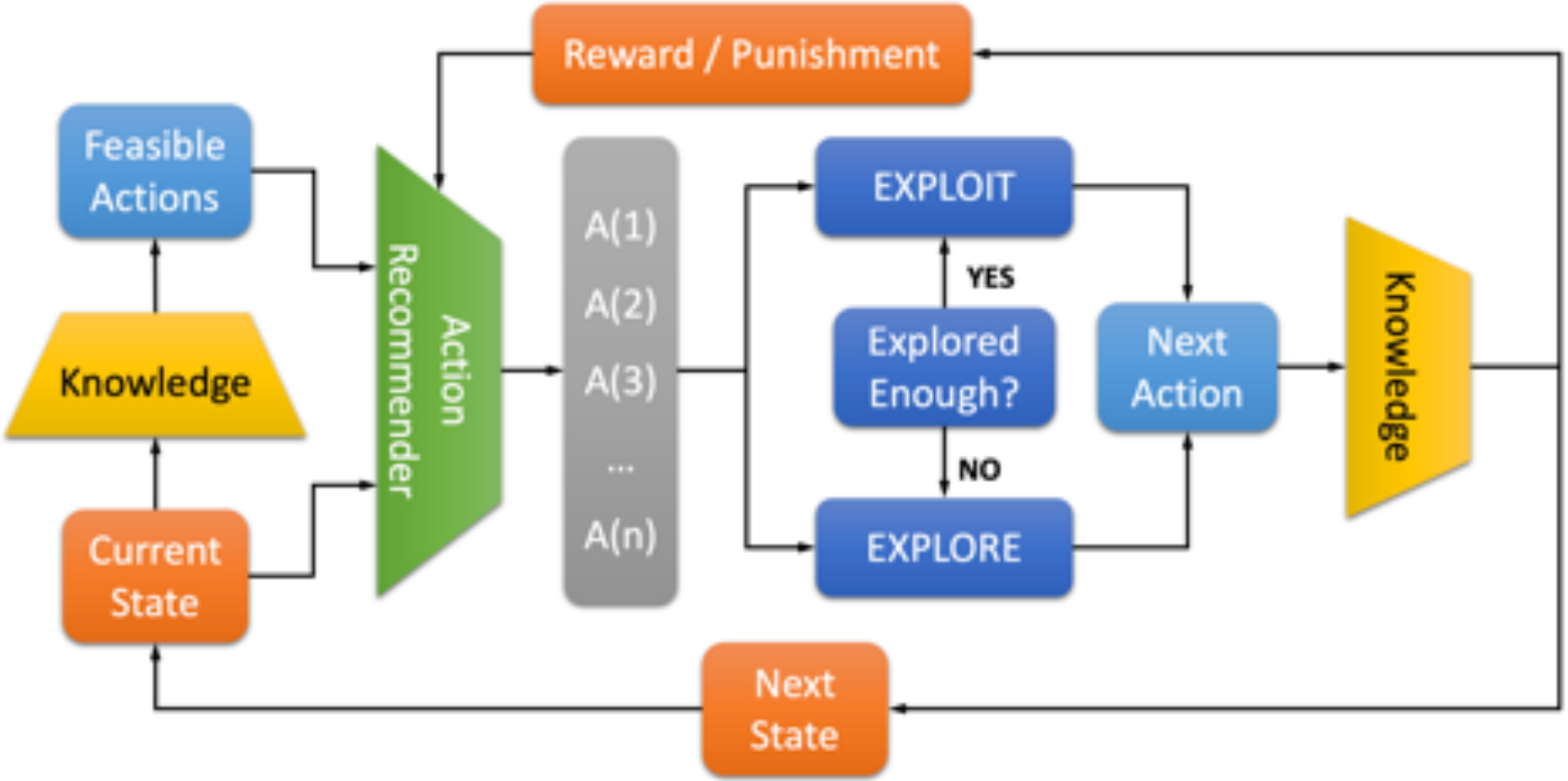




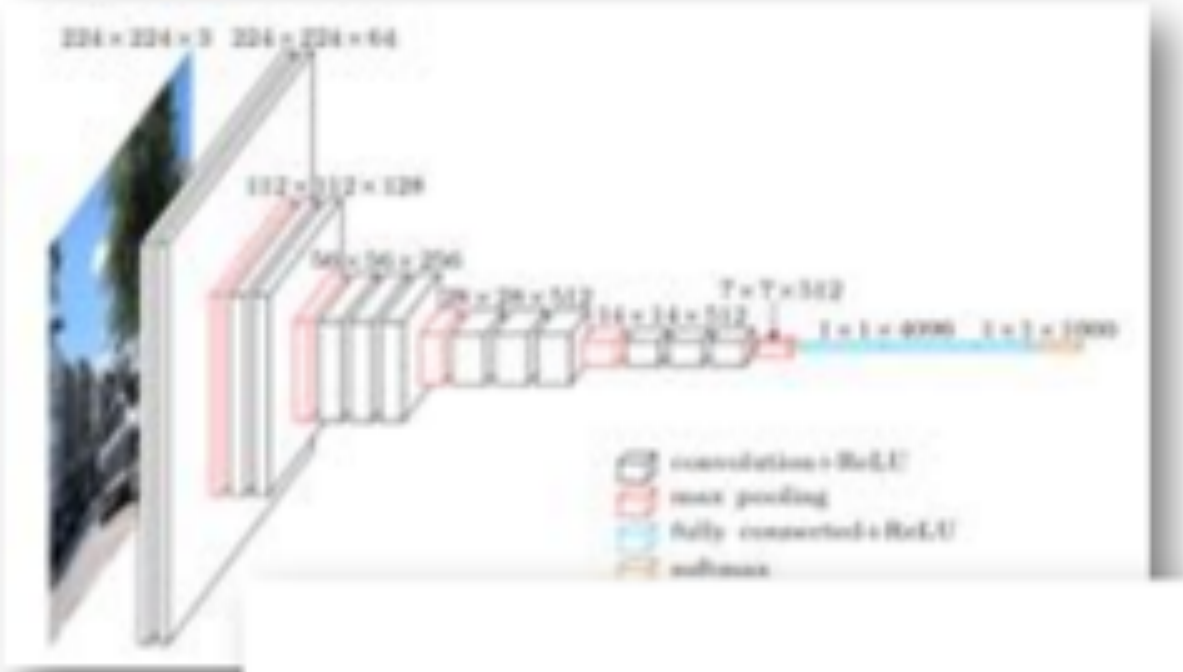
REASONING paradigm

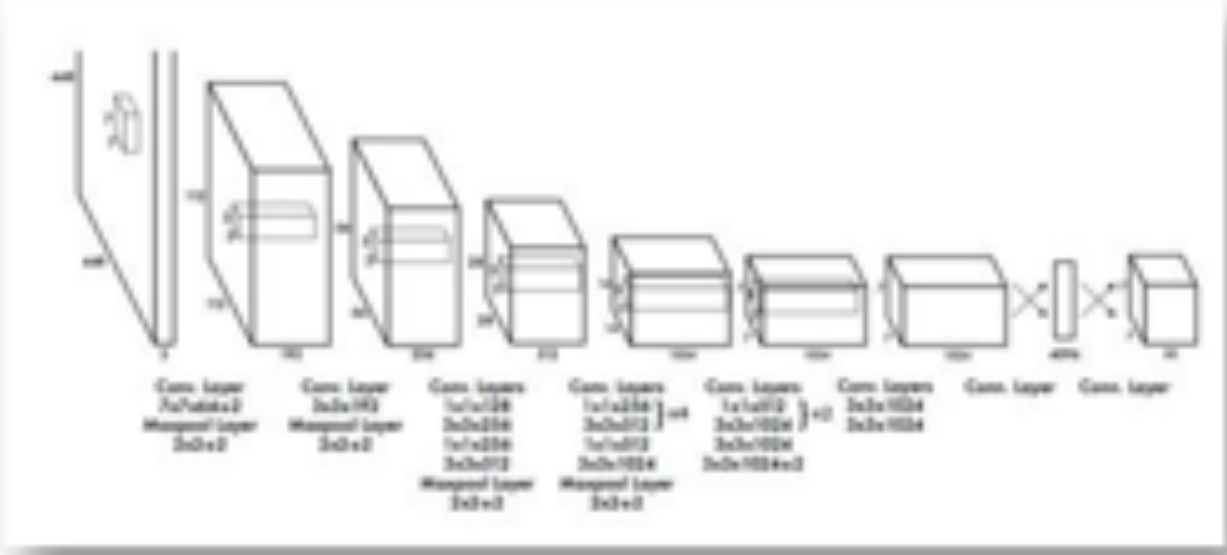


REINFORCEMENT LEARNING paradigm



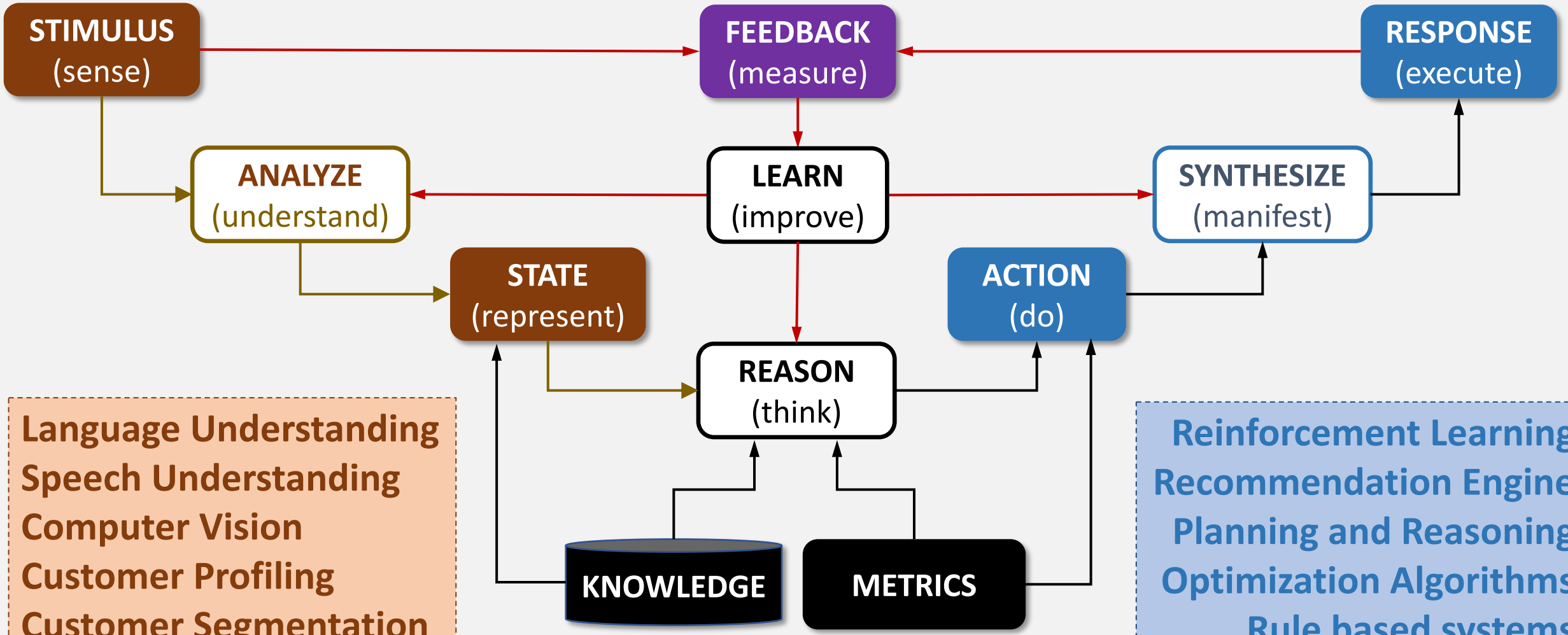
DEEP LEARNING paradigm





Templatization of AI

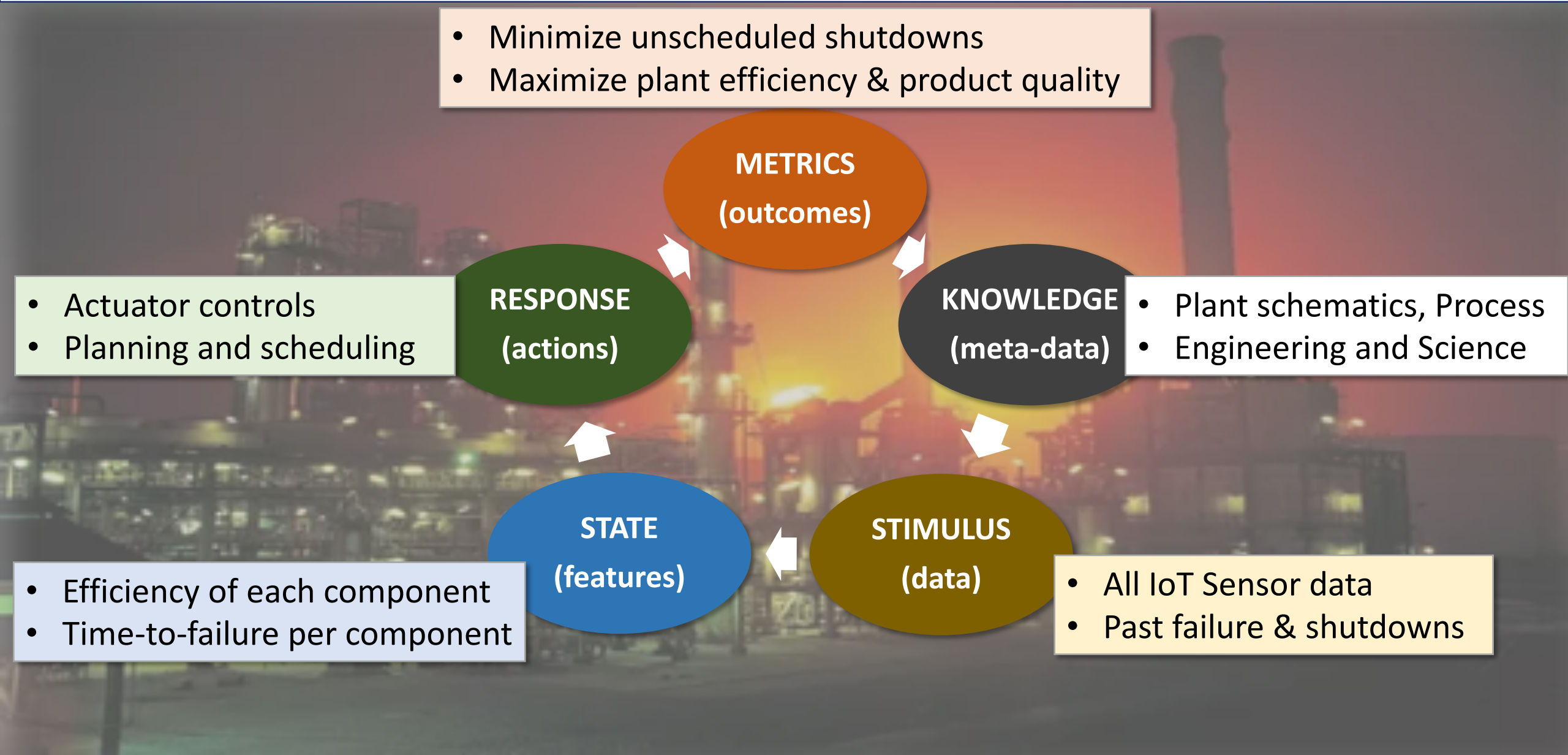
An **ARCHITECTURE** of AI



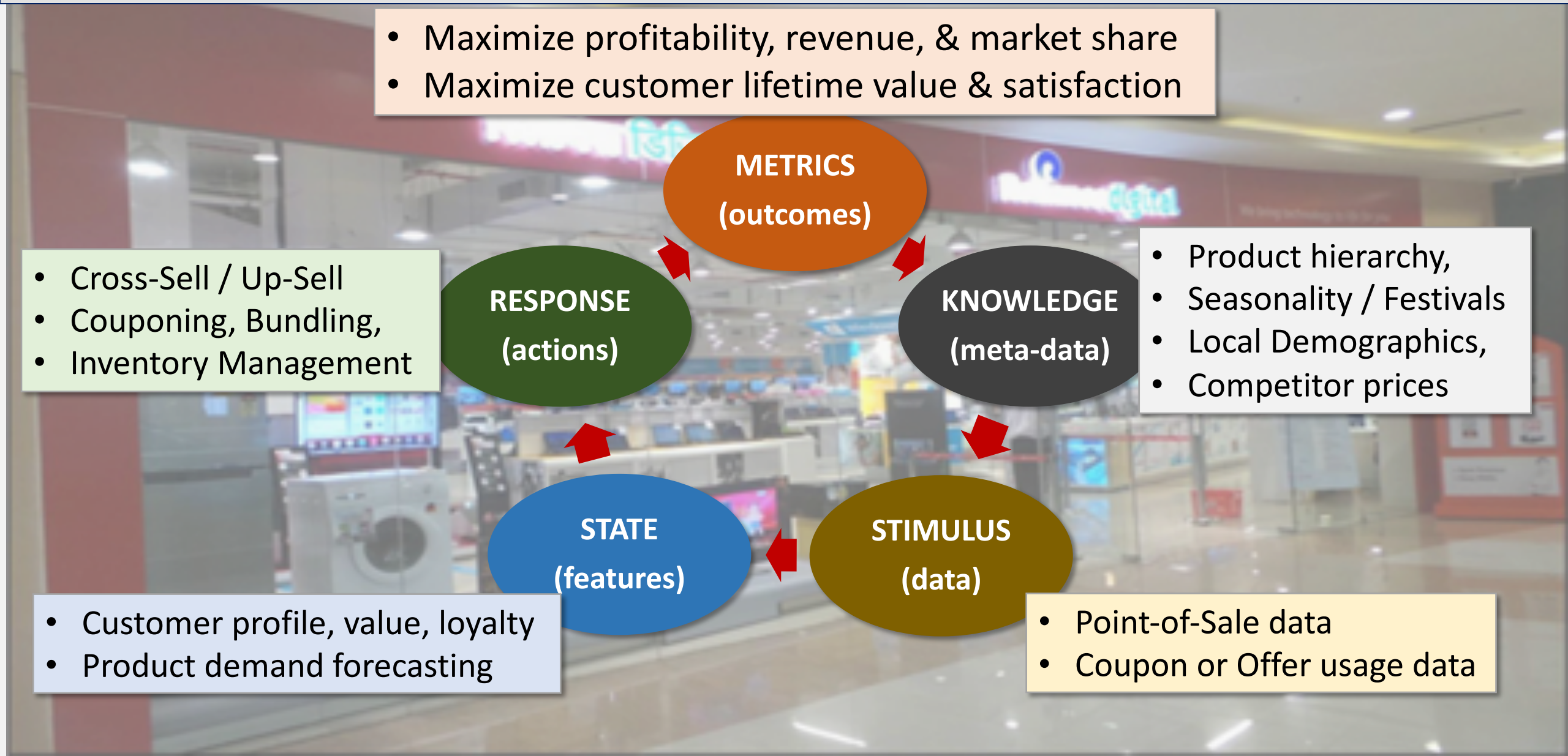
Language Understanding
Speech Understanding
Computer Vision
Customer Profiling
Customer Segmentation
Demand Forecasting

Reinforcement Learning
Recommendation Engine
Planning and Reasoning
Optimization Algorithms
Rule based systems
Business Logic

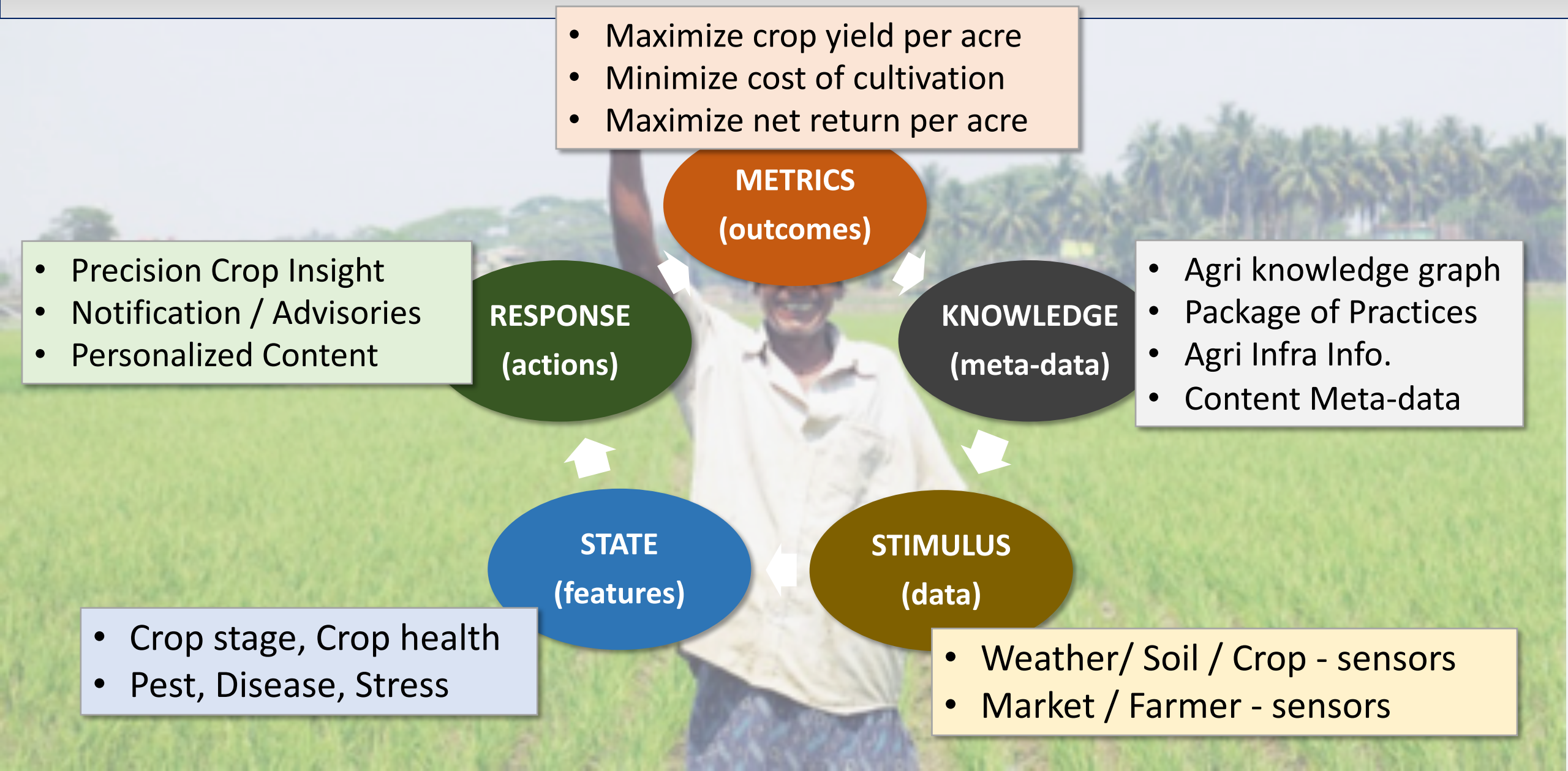
AI for **INDUSTRIAL IoT**– Self-optimizing Plants



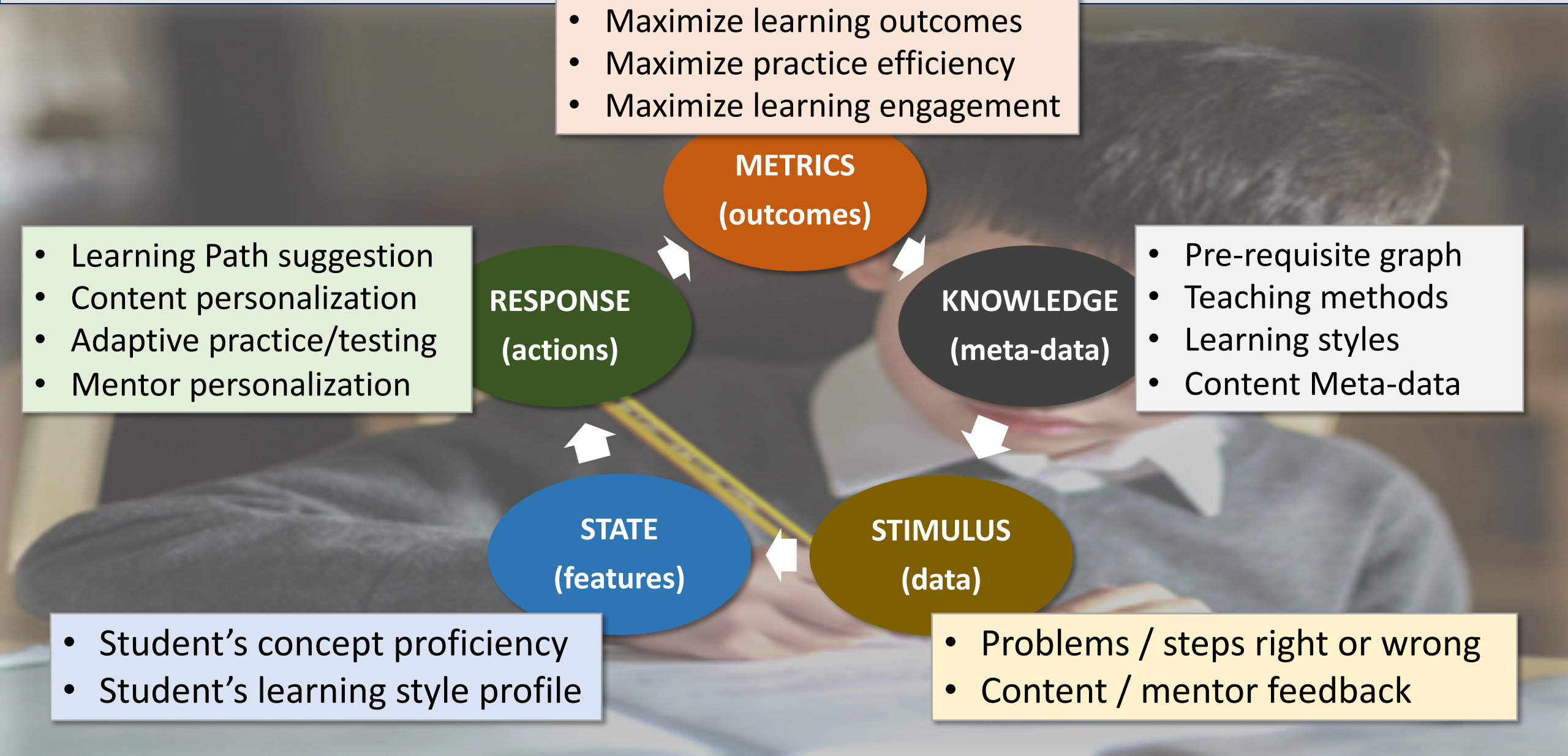
AI for **RETAIL** – Personalized, Intelligent Retailing



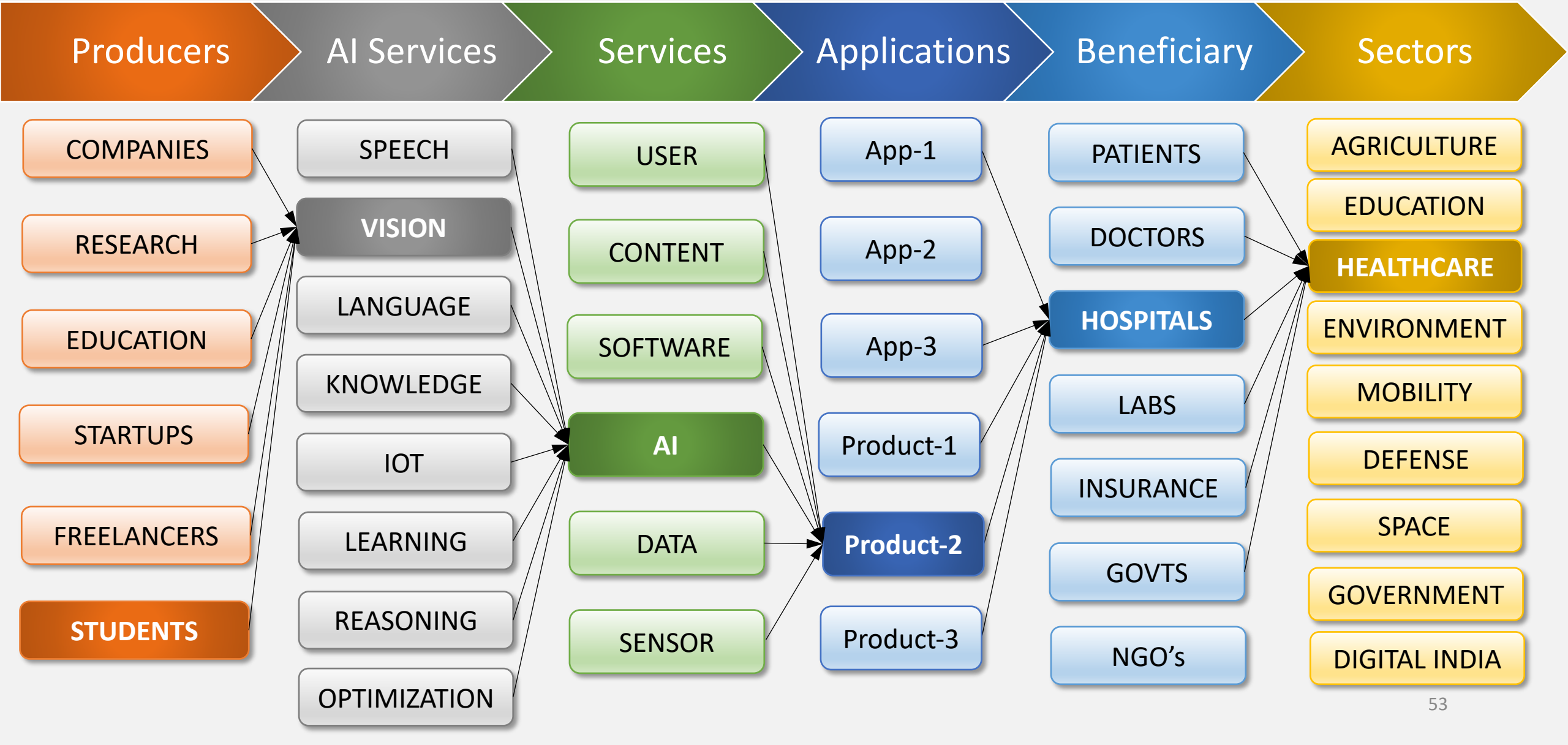
AI for **AGRICULTURE** | Personalized, Precision farming



AI for **EDUCATION** – Personalized education at scale



Democratizing AI for All

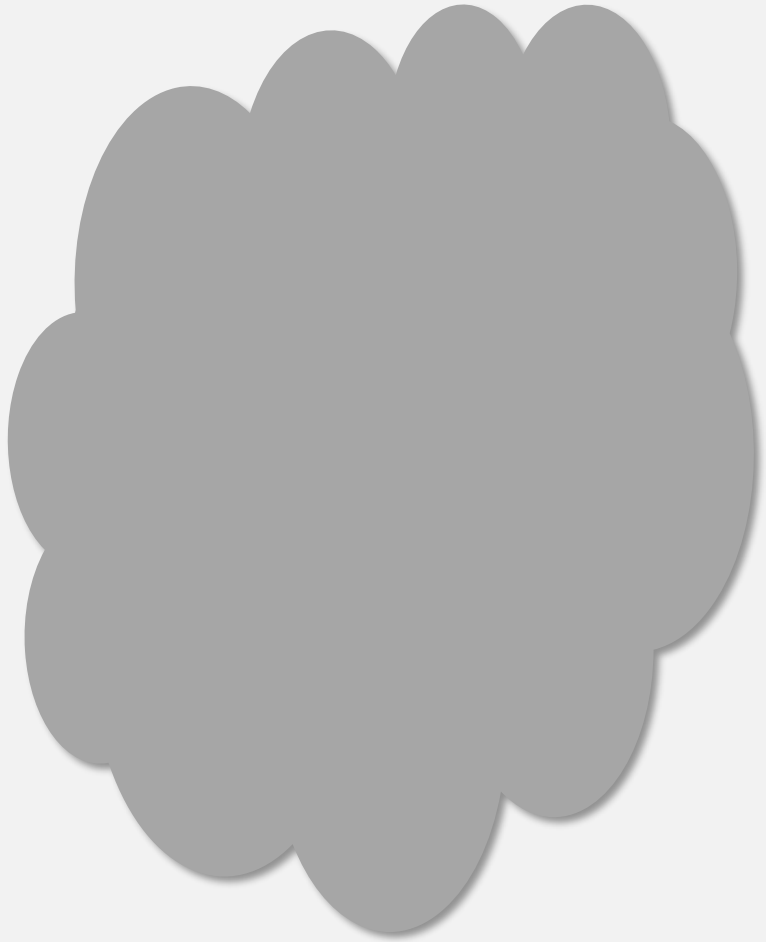


Course Outline

Machine Learning Paradigms

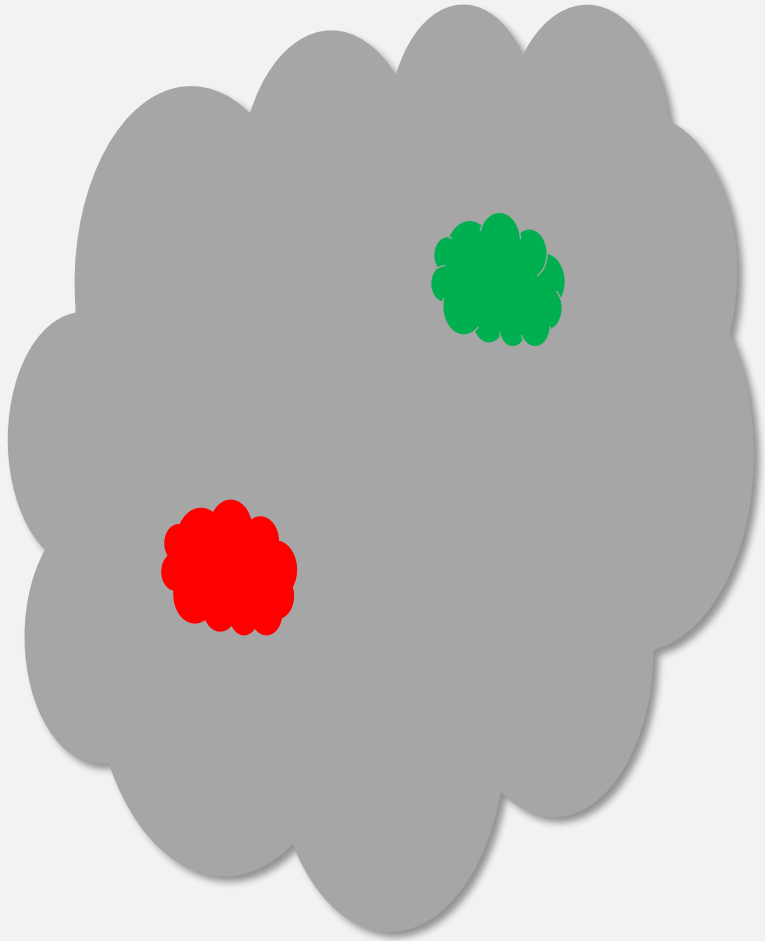
- **Unsupervised Learning**
  - Find **structure** in data. (Clusters, Density, Patterns)
- **Supervised Learning**
  - Find **mapping** between features to labels
- **Semi-Supervised Learning**
  - Using unlabeled data to improve Supervised Learning Models
- **Active Learning**
  - Find which model to get labeled next to maximize value
- **Reinforcement Learning**
  - Learn local/early strategies from global/delayed “rewards”

Un-labelled Data : Unsupervised Learning

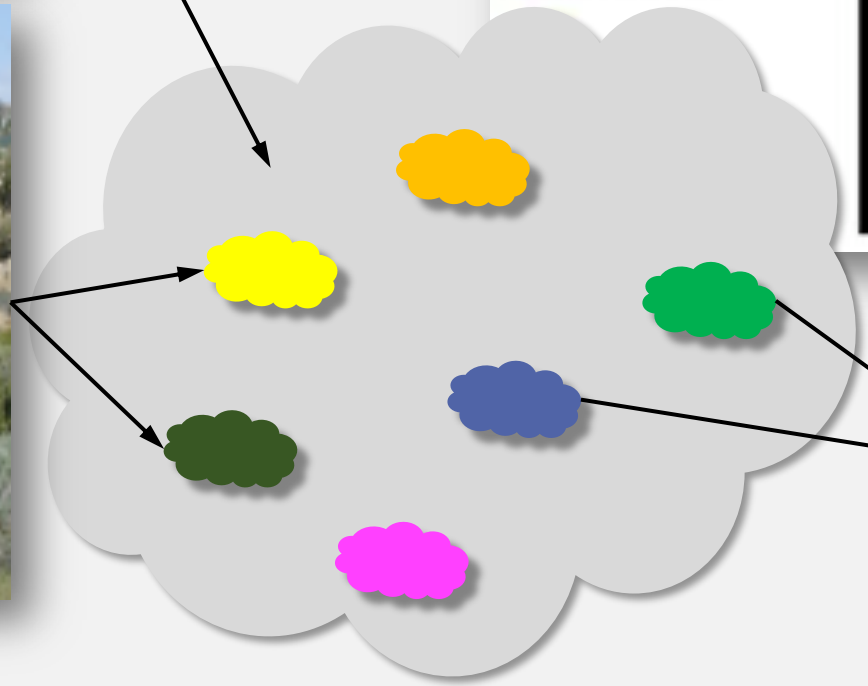


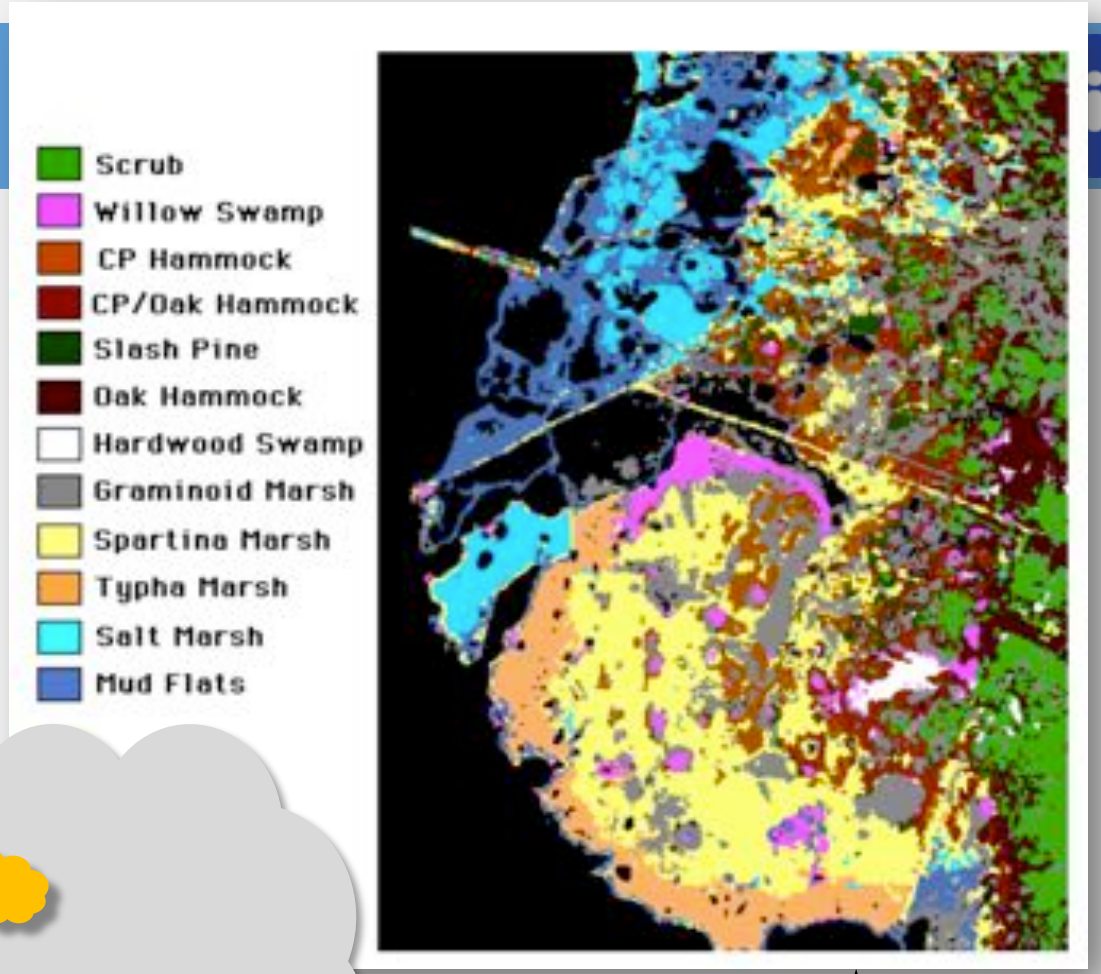
- **Web Pages**
- **YouTube Videos**
- **LinkedIn Profiles**
- **Flipkart Transactions**
- **Your E-mails**
- **Earth Pixels**
- **Gene Sequences**

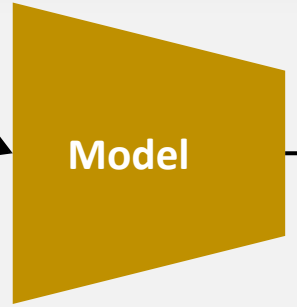
Labeled Data : Supervised Learning



- **Web Pages** → {Spam | Not-Spam}
- **YouTube Videos** → { Cartoon | not}
- **LinkedIn Profiles** → { Job | not-Job}
- **Flipkart Transactions** → {Churn | not}
- **Your E-mails** → {Main | Promo | Spam}
- **Earth Pixels** → {Land | Water | Crop}
- **Gene Sequences** → {Cancer | not}




Scrub
Willow Swamp
CP Hammock
CP/Oak Hammock
Slash Pine
Oak Hammock
Hardwood Swamp
Graminoid Marsh
Spartina Marsh
Typha Marsh
Salt Marsh
Mud Flats


Model

Supervised Learning Process

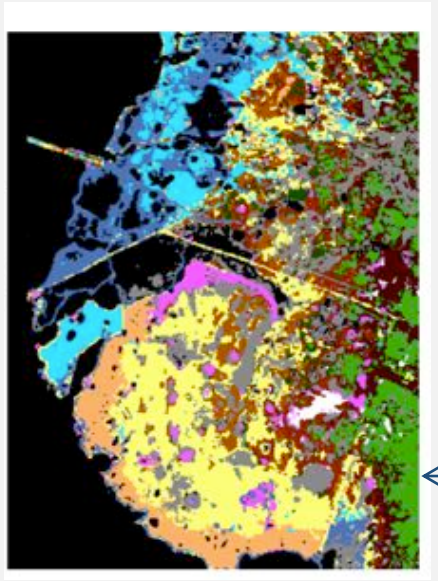
Collect Raw (Input) Data

Collect (Output) Ground Truth!

Engineer "***Predictive***" features

"**Train**" an ML "Model"

Choose Model type & complexity



Deploy Model. Make Predictions on Unlabeled data

Evaluate, Iterate, Improve Model

Supervised Learning Paradigms

- **Regression**
  - Predict a **numerical** **VALUE**.
- **Classification**
  - Predict a **categorical** **CLASS**.
- **Recommendation**
  - Predict user **PREFERENCE** from a **large pool** of options.
- **Retrieval**
  - Predict **RELEVANCE** of an **entity** to a “**query**”

Which Supervised Learning Problem is this?

- Which crop is best for this soil/weather?
- Which position should a Page/Ad be shown for a query?
- Should the home loan/credit card transaction be approved?
- Which Video to show next on YouTube?
- Who might user X connect/follow on LinkedIn/Twitter?
- Which product the customer will buy next?
- How much should Android One / iPhone 6 be sold for in India?
- Which medicine/treatment will cure this disease?
- Who will be the right person (resume) for this Job (JD)?
- What movies customer X might be interested in?

Which Supervised Learning Problem is this?

- Does this MRI scan show Cancer?
- What will this stock value be tomorrow?
- How much credit limit should I give a customer X?
- Is customer X about to Churn?
- What activity (*stealing* or *shopping*) is user doing in a Video?
- What objects (*car, tree, horse, …*) are present in an image?
- How many iPads will I sell in Mumbai between April – July?
- Is this a positive or negative TWEET about a person/product?
- How high the value of my house will go in 3 years?
- Which song are you humming the tune of?
- Who will be the best match for you (say shaadi.com)