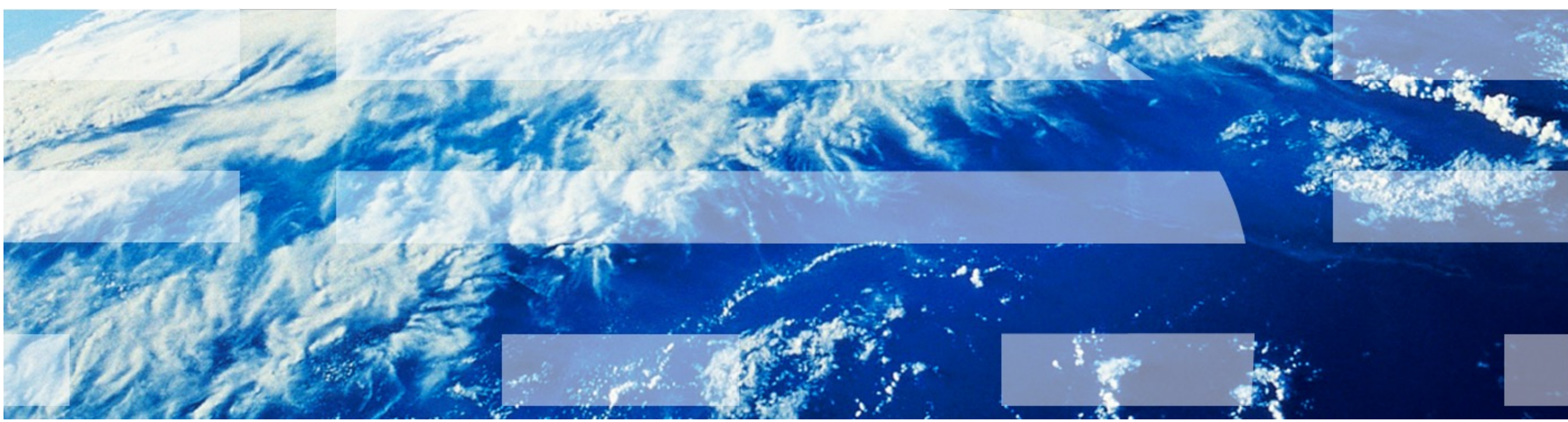
E6895 Advanced Big Data Analytics and AI Lecture 1:

Introduction of Advanced Big Data and AI

Ching-Yung Lin, Ph.D.

Adjunct Professor, Depts. of Electrical Engineering and Computer Science



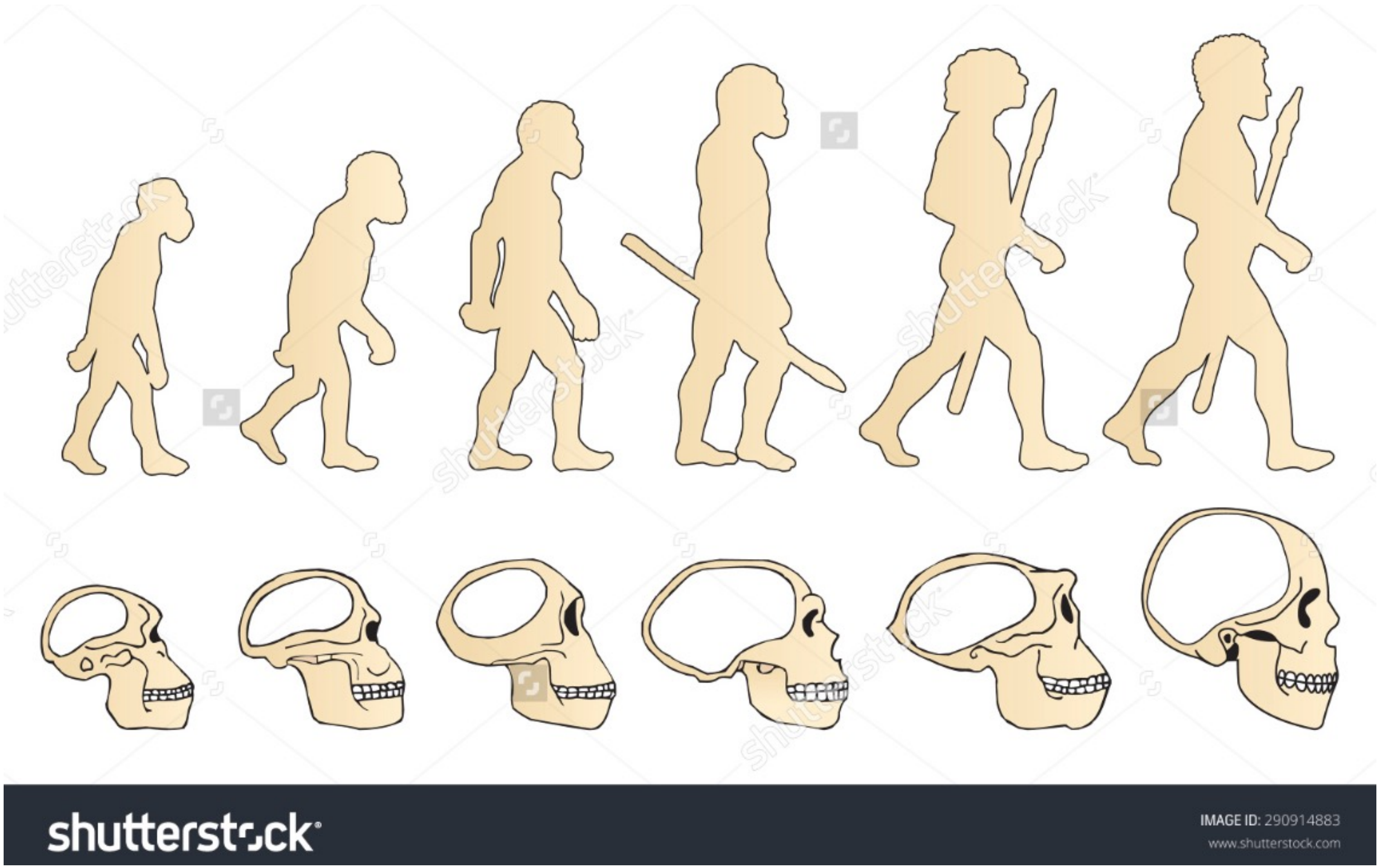
January 19, 2024

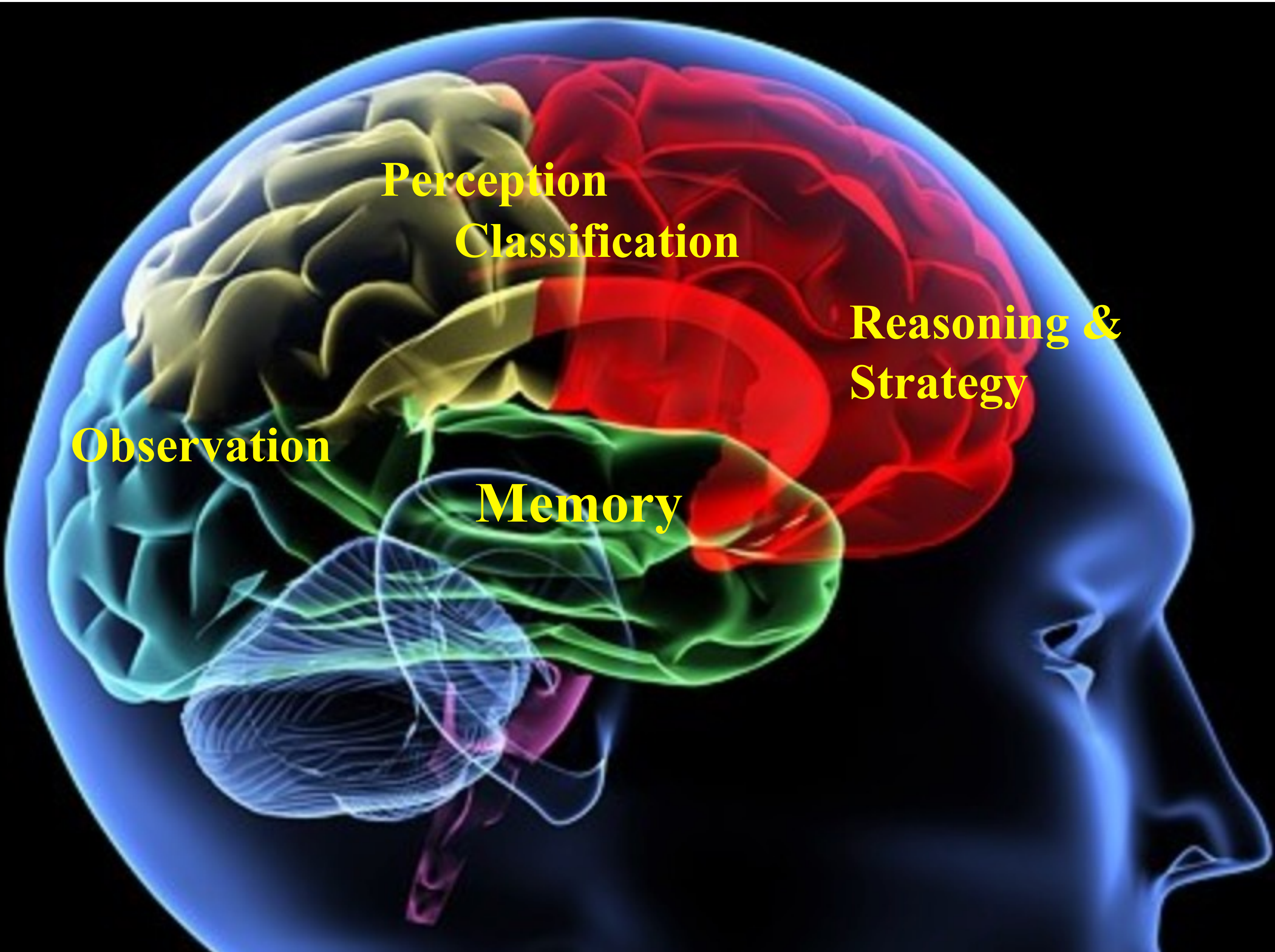

1997
IBM
GARRY
KASPAROV
DEEP
BLUE

2011


THINK
ΣΚΕΨΟΥ
सोचिए
$24,000
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$1,000
$77,147
Who is Bram Stoker?
$ 17,973
$21,600
WHO IS BRAM STOKER?
$5600

2015


shutterstock
IMAGE ID: 290914883
www.shutterstock.com


Perception
Classification
Reasoning &
Strategy
Observation
Memory

Do we need AI?

Who will be our caregiver?

“Single Child”
Finalist of the 26th National Photo Contest, China

Where to find Helps?

All Developed and some developing countries have been facing labor shortage crisis ➔ More and more serious everyday.

https://www.youtube.com/watch?v=BV8qFeZxZPE

2023 AI Summit at New York Javits Center

Graphen: The largest booth

World’s First AI Digital Human for Daily Life

Graphen Robotics

Meet Aiia

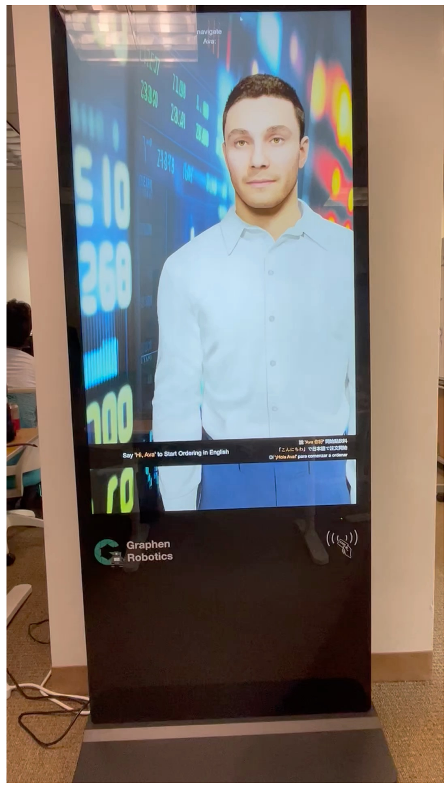
- Hardware-Software Integrated Local AI ‘Brain’.
- Privacy / Individual / Personal
- Speaks English, Chinese, Japanese, and Spanish
- Avatars with Personality & Emotion
- Eye Contact / Facial Expression
- Integrating with Payment, Mobile Apps, etc.

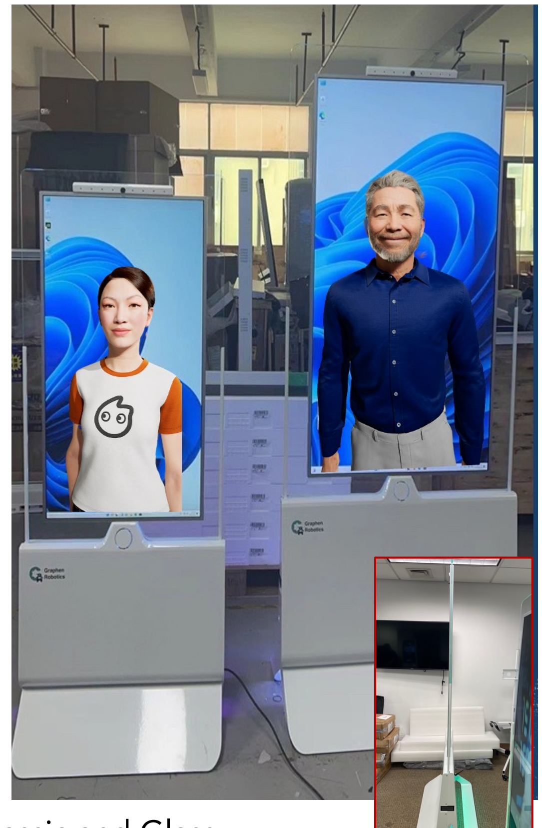
Graphen Robotics Hardware Systems

Aiia Kiosk

32″

Aiia Know



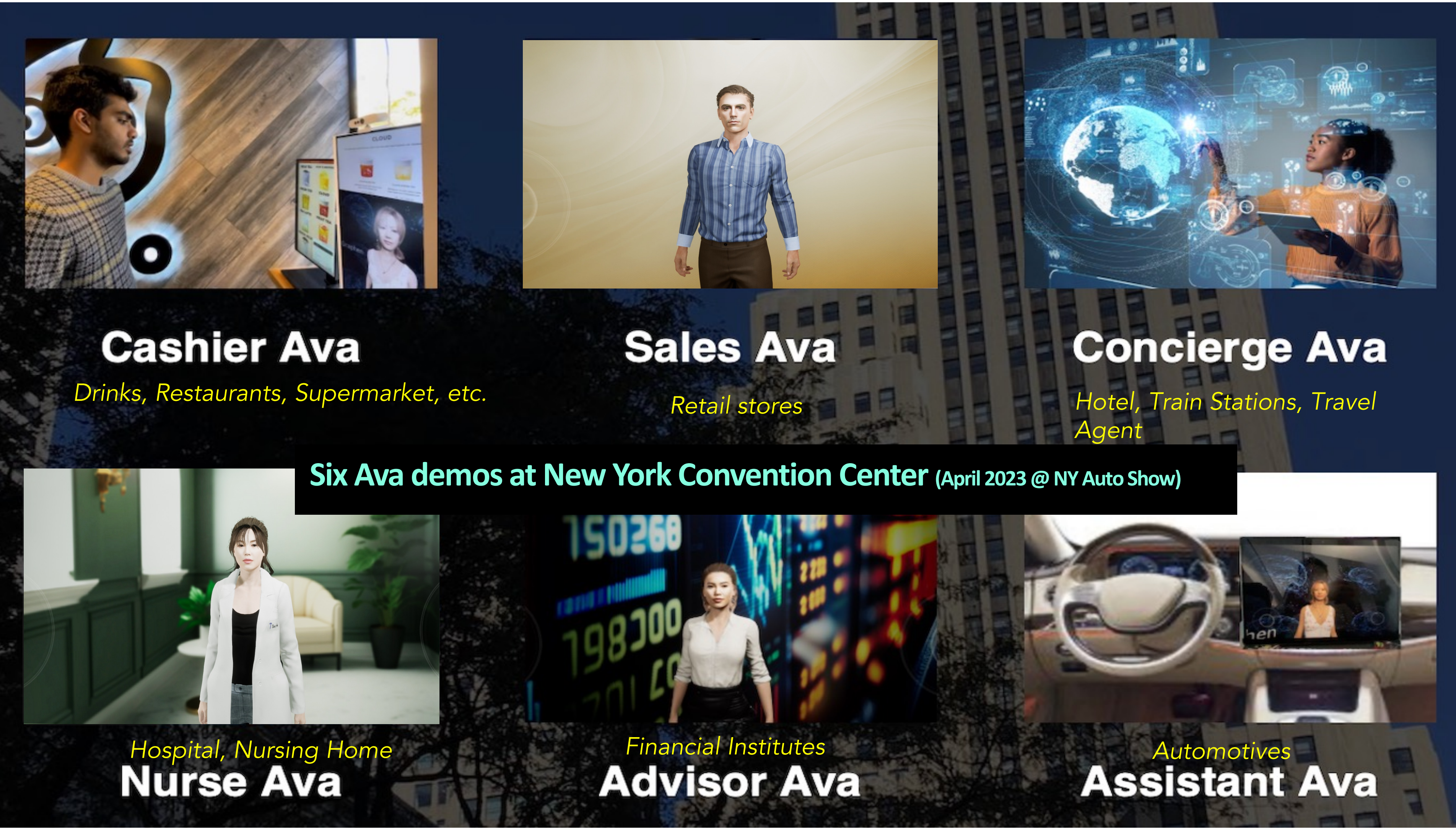


43″ and 55″; Classic and Glass

Aiia Robot (Adam)

32″

Digital Human Application Scenarios



Examples :

Instant reference tool for medication dosages, side effects, and interactions, reducing the risk of medication errors.

Patient education : helping nurses provide accurate, understandable explanations of medical conditions and treatments.

Question : What is the infusion time for 1 unit of Packed Red Blood Cells?

Aiia Nurse Assistant: PBRCs are a blood product used to replace erythrocytes; infusion time for 1 unit is usually between 2 and 4 hours.

Source: The answer is obtained by retrieving page 158 in the provided PDF, which is the RN Exam textbook.

→ Aiia answered 90% questions correctly in New York RN License Exam

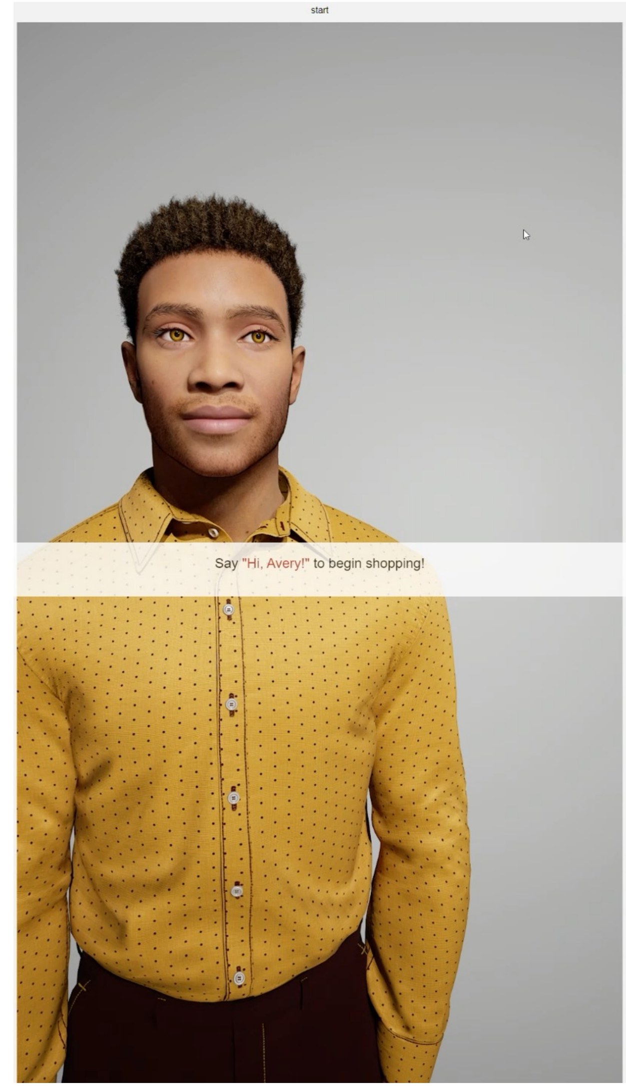
Digital Human Application Scenarios

Aiia Examples

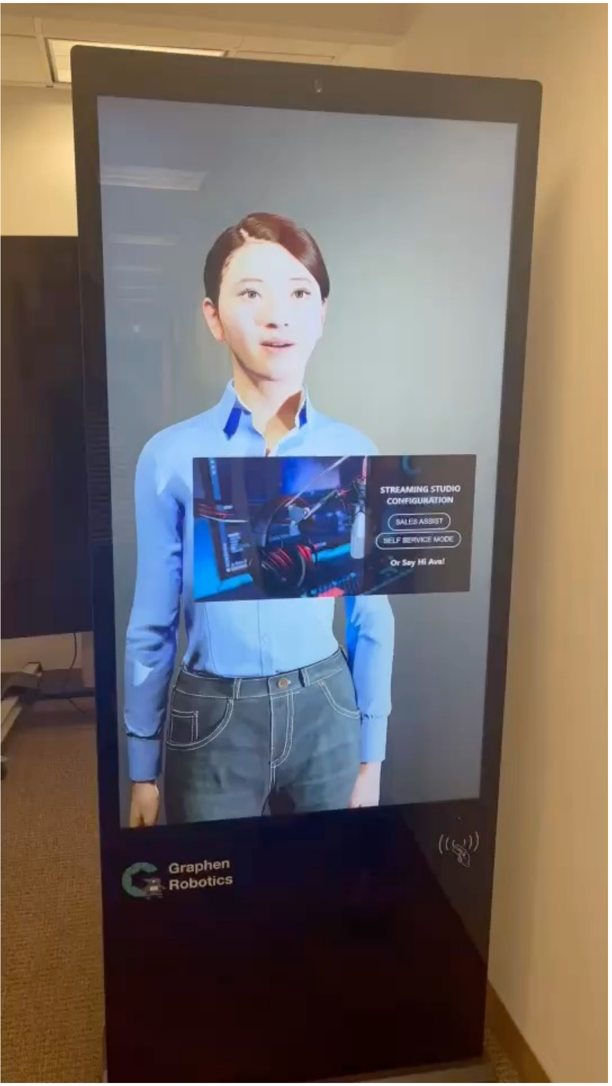
Aiia Financial Advisor

Retail Aiia

Graphen Robotics



Digital Human Application Scenarios



Aiia Examples

Aiia as Sales Assistant

Aiia helps ordering

Graphen Robotics

① Patient Care:

- **Do Routine Works for Nurse:** discharging info, facility info, forms to fill, patient and caregiver educations, social worker roles, etc.
- **Provide health-information** on how to cope with the situation by medical information, health information etc.,
- **Provision of information** on economics, travel, hobbies and preferences through conversations between patients and avatars
- **Brain vitalization** through enhanced daily conversation, maintenance and promotion of health, and **support as a "personal companion"**
- **Entertainment**

② Operational Support for Nurse Station:

- **Patient Personal information** (including gene information, treatment history, drug administration information, etc.)
- **Real-time sensor data monitoring** of patients (body temperature, blood pressure, sleep, turnover status, awake status, etc.)

Course Outline

Class Date	Class Number	Lecture Topics	Student Presentations
01/19/24	1	Introduction of Advanced Big Data and AI	
01/26/24	2	Big Data & AI Foundations	
02/02/24	3		Full-Brain AI (I) & Green Earth (I)
02/09/24	4		Financial Advisor (I) & Healthy Life (I)
02/16/24	5	Contextual Knowledge	Advanced AI Study
02/23/24	6	Machine Reasoning	Advanced AI Study
03/01/24	7		Full-Brain AI (II) & Green Earth (II)
03/08/24	8		Financial Advisor (II) & Healthy Life (II)
03/15/24		SPRING BREAK	
03/22/24	9	Language Models	Advanced AI Study
03/29/24	10	Cognition	Advanced AI Study
04/05/24	11		Full-Brain AI (III) & Green Earth (III)
04/12/24	12		Financial Advisor (III) & Healthy Life (III)
04/19/24	13	Edge AI	Advanced AI Study
04/26/24	14	Advanced Artificial Intelligence	Advanced AI Study
05/03/24	15		Final Project Workshop

Presentation Schedule shall be adjusted by on the distribution of tasks.

Course Grading

- Task: 45%
  - — Teamwork: 1 - 2 students per team
  - Choose a task from 60 potential tasks
  - Language Requirement: Python, JavaScript, C/C++, Java, Perl
  - 3 milestones (45%): Presentation, Slides, Report and Source Code

- Final Project: 30%
  - — Teamwork: 1 - 2 students per team
  - Building System
  - Final Report (paper, up to 12 pages)
  - Workshop Presentation and Online Video
  - Open Source

- Research Study: 15%
  - 3 research paper presentations related to Advanced AI: Slides

- Class Participation: 10%
  - Attendance
  - Discussion (Asking/Answering Questions)

Task Sign-Up Spreadsheet is available until midnight 1/26

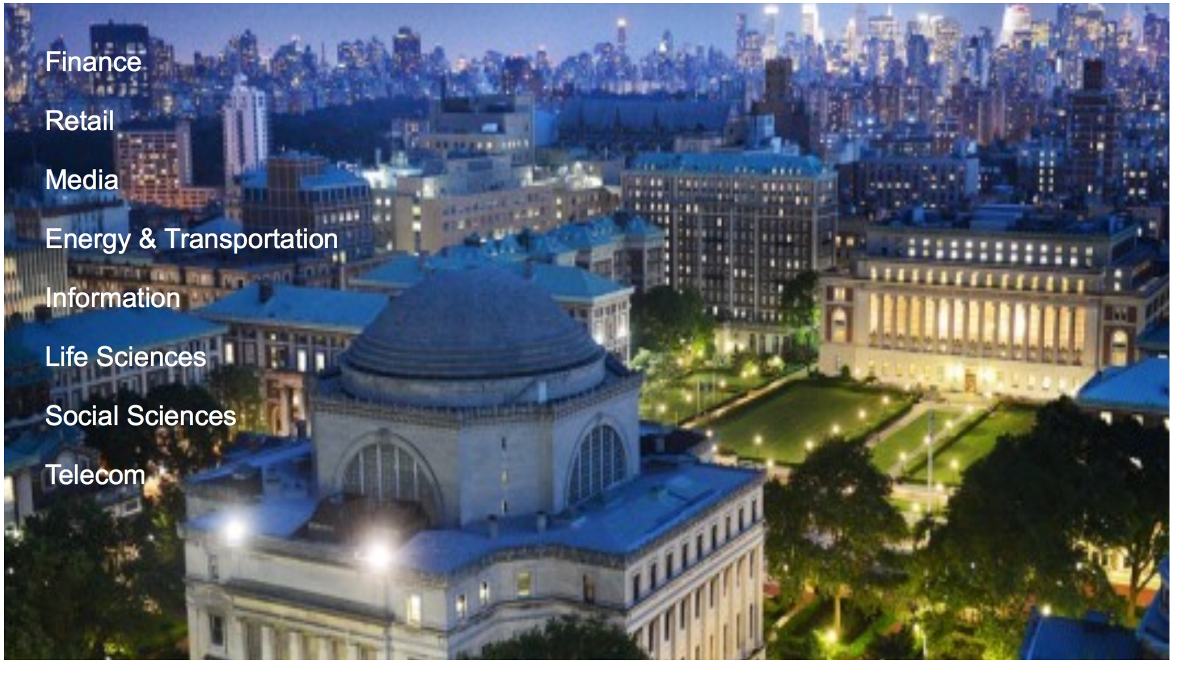
Course Information

- Website:

  http://www.ee.columbia.edu/~cylin/course/bigdata/

- Textbook:

  -- None, but reference book(s) and/or articles/papers will be provided each lecture.



Contact Information and TAs

- Professor Lin:
  - Office Hours and Location:

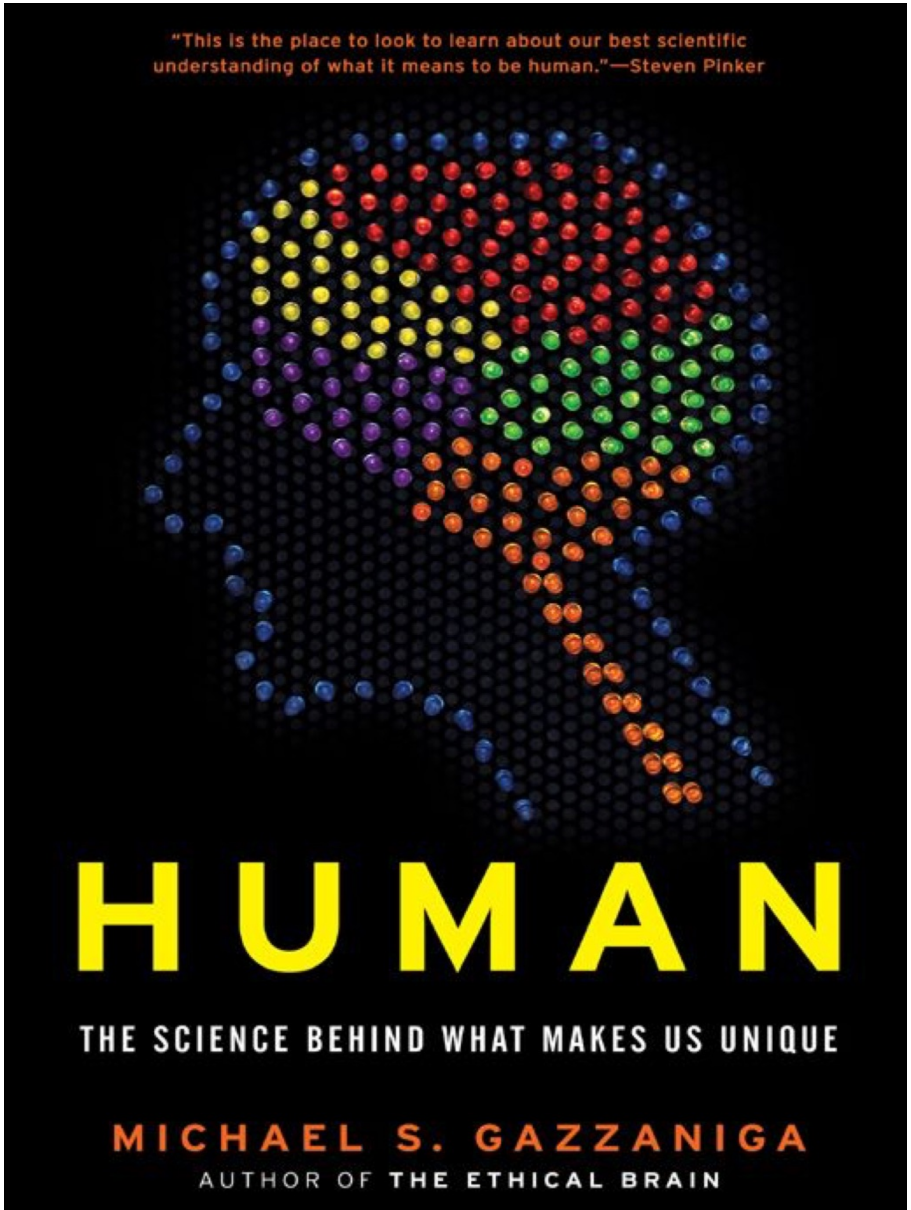
    Friday 9:30pm – 10:00pm (lecture room) or by appointment (500 Fifth Ave., Suite 2420, New York, NY 10110)
  - Contact: c.lin@columbia.edu

- TAs:
  - Shiyu Wang (sw3601)
  - TBD

Special Request — thanks:

- If you may not take the class, please do not sign up the task spreadsheet.
- Please remove your name from the task sign-up sheet immediately when you drop the class.
- Please drop the class as early as possible, if you are not planning to take the class.

Reference Book



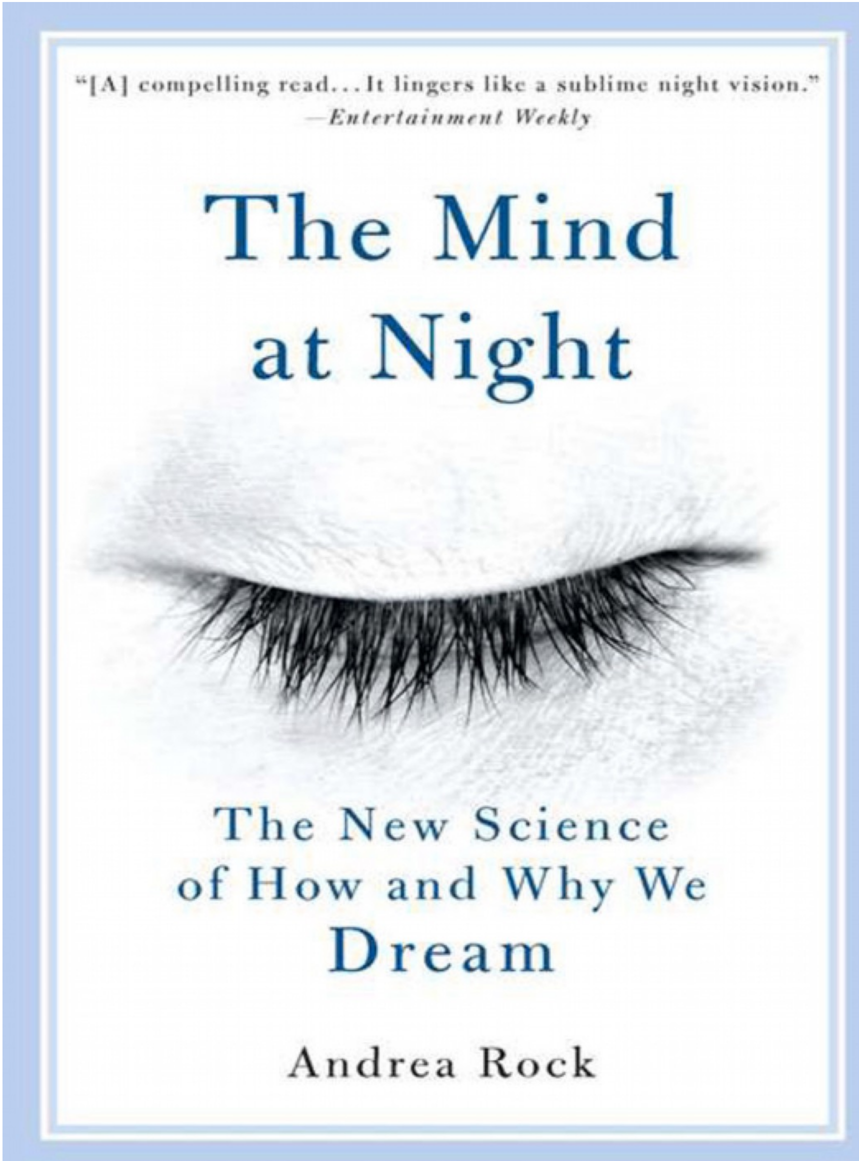
Part 1
The Basics of Human Life

Part 2
Navigating the Social World

Part 3
The Glory of Being Human

Part 4
Beyond Current Constraints

Reference Book



1 Rockettes, EEGs, and Banana Cream Pie
2 The Anti-Freud
3 Experiments of Nature
4 The Lesson of the Spiny Anteater
5 Rerunning the Maze
6 Nocturnal Therapy
7 The Ultimate Spin Doctor
8 Creative Chaos
9 Altered States
10 Consciousness and Beyond

Reference Foundation

- **Graph Middleware**:
  - Parallel Prog. Lib.
  - Power Optimization
  - GPU Optimization
- **Graph Analytics**:
  - Topological Analysis
  - Matching and Search
  - Path and Flow
- **Spatiotemporal Analytics:**
  - Spatiotemporal Mining
  - Spatiotemporal Indexing
- **Graph Database**:
  - Native Store
  - GBase
- **Graph Visualization:**
  - Multivariate Graph
  - Dynamic Graph
  - Big Graph
- **Machine Learning:**
  - Deep Learning Tools
  - Visual and Text Sentiment Tools
  - Anomaly Detection Tools
- **Mobile Cognition:**
  - iOS Cognition Tools
  - Robot Cognition Tools
- **Machine Reasoning:**
  - Bayesian Networks
  - Game Theory Tools
  - Multimodal Analysis Platform

4. Machine Reasoning Technologies

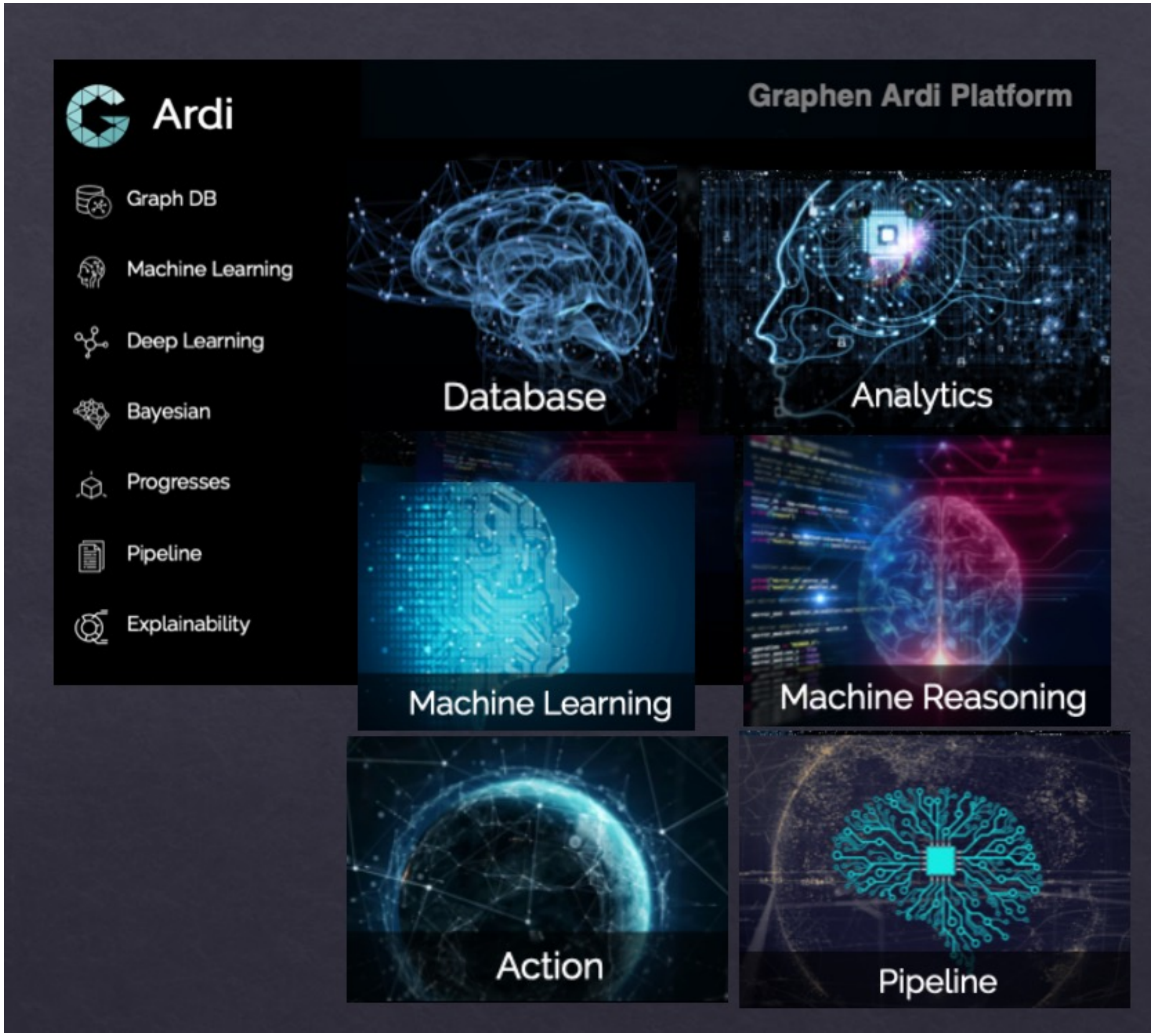
3. Machine Learning Technologies

2. Network Analytics Technologies

1. Graph Database Technologies

Reference Advanced AI + Big Data Platform

- Terabyte-sized native GraphDB, supports trillion of vertices and edges
- ACID-compliant and distributed Graph database and analytics
- Asynchronous job scheduling (both Autonomous ML and GraphDB)
- Scalable, distributed Analytics, modular and expandable through plugins
- Cluster, Replication and High-Availability with disaster recovery
- Error and event Logging, Monitoring, Backup and Recovery



Task Area 1: Full-Brain Machines

Area 1 ‘Cognitive Machine’ Tasks List:

A1: Deep Video Understanding (Visual + Knowledge) — Face Recognition, Feeling Recognition, and Interaction

A2: Deep Video Understanding (Language + Knowledge) — Speech Recognition, Gesture Recognition, and Feeling Recognition

A3: Deep Video Understanding — Event and Story Understanding

A4: Humanized Conversation — Personality-Based Conversations

A5: Autonomous Robot Learning of Physical Environment

A6: Autonomous Task Learning via Mimicking

A7: Digital Human for Fashion

A8: Digital Human for Tourism

A9: Digital Human for Retail

A10: Digital Human for Media and Marketing

A11: Feeling and Art Recognition

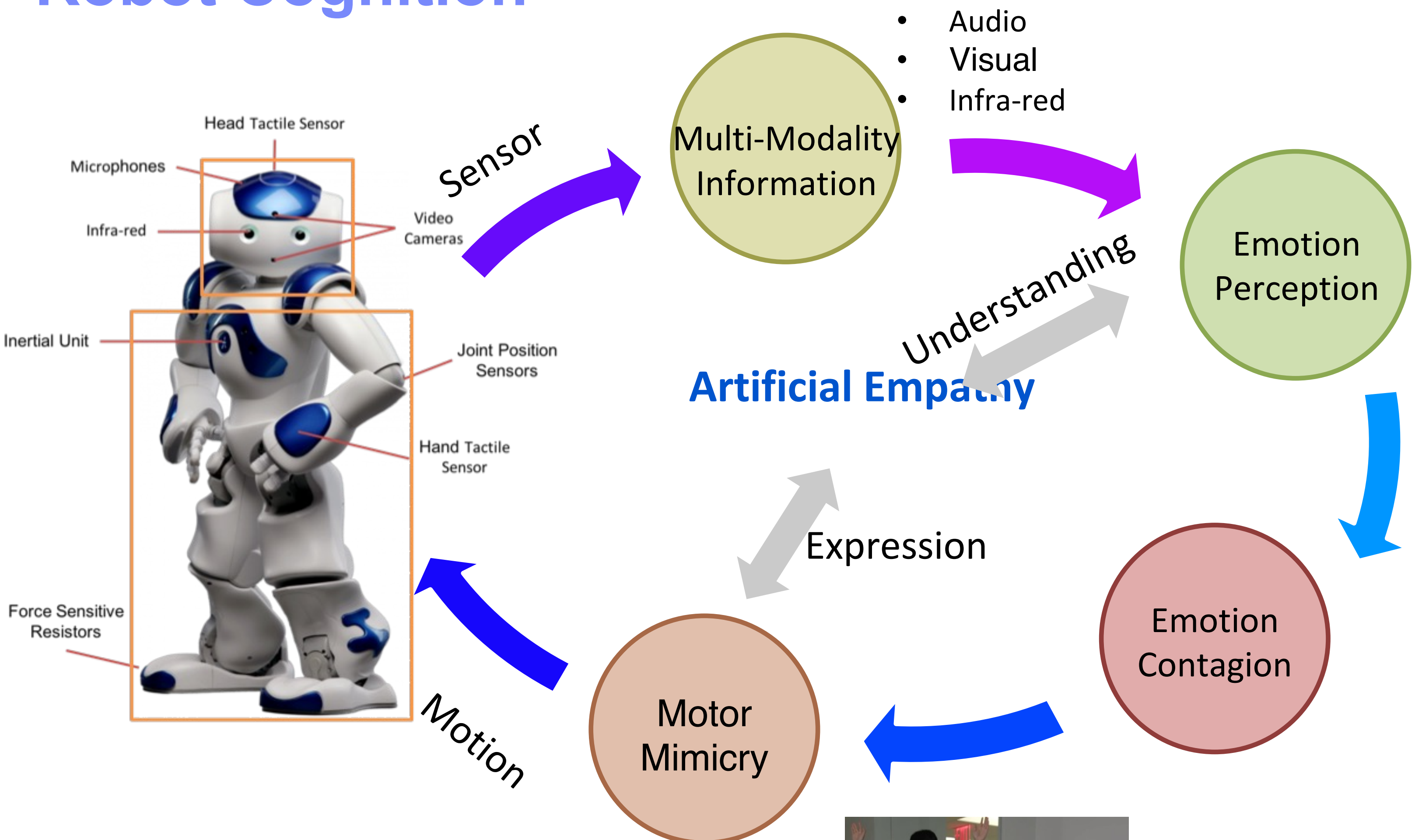
A12: Creative Writing & Story Telling

A13: Knowledge Learning & Construction

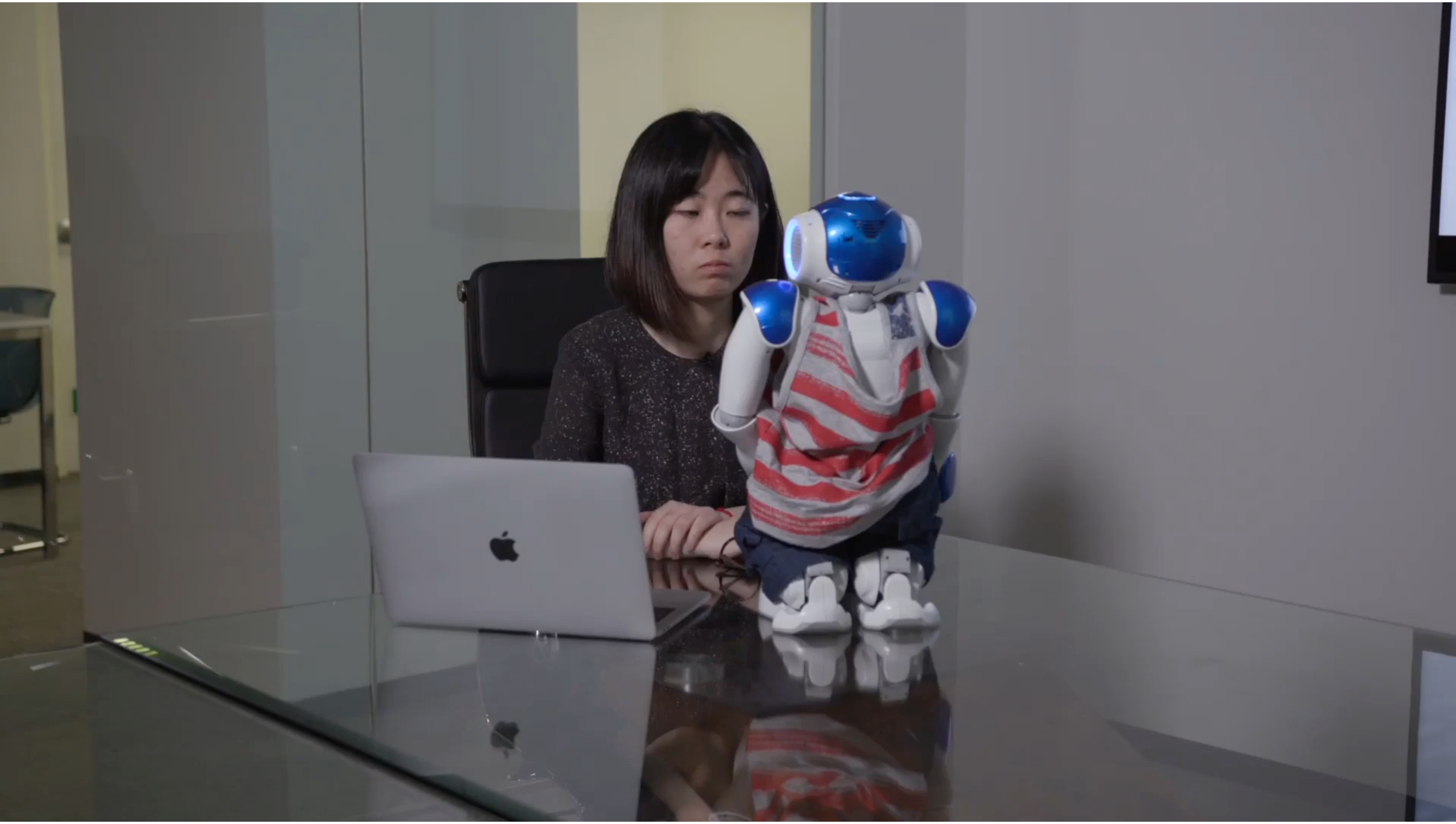
A14: Dreams — Simulating Brain functions while sleeping

A15: Self-Consciousness, Ethics, and Morality

Robot Cognition



Emotion and Cheers

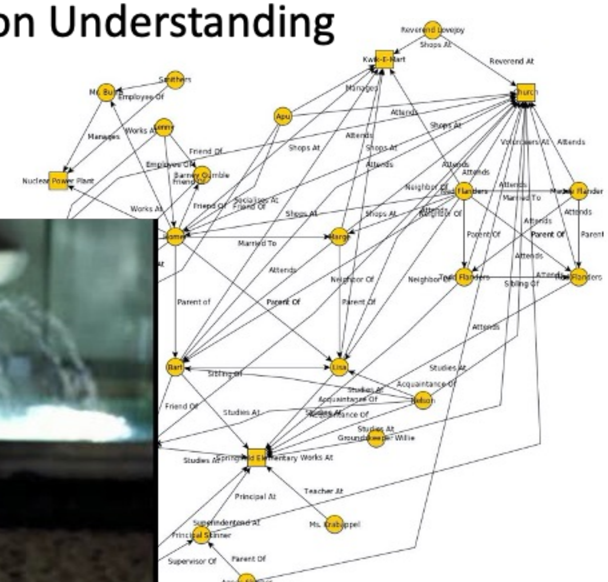


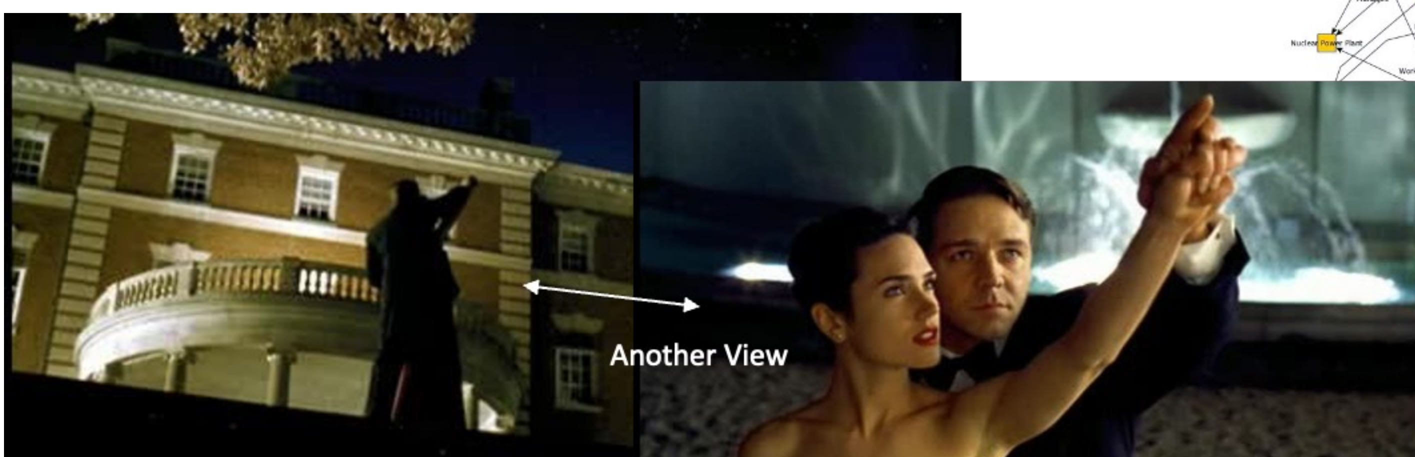
How Robot cheers you up

A1-A3. Deep Video Understanding

A complete system combining these for video understanding

- Visual Recognition
- Speech Recognition
- Knowledge Graph
- Face Recognition
- Emotion Recognition
- Speaker Identification
- Relationship Inference
- Event and Action Understanding





Potentially Target at NIST Deep Video Understanding 2022

A4. Humanized Conversation with Personality

Description:

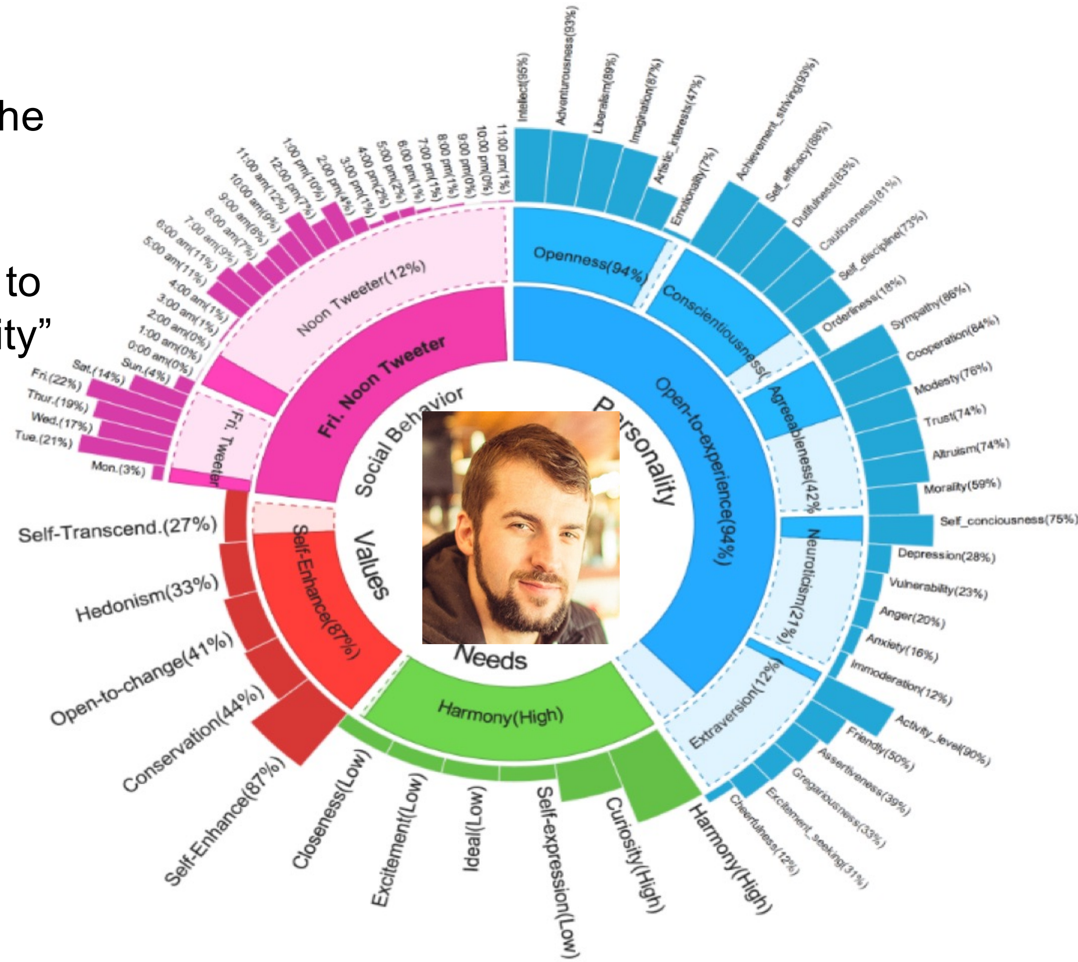
- Virtual Agents are progressing fast and entering people's life. However, the voice presented by the agents are mostly 'flat' — like machines.
- The first step to make virtual agents to be like human is to add the "personality" aspect in conversation.

Goal:

- Create Personality-based Speaking Model Text for Conversation

Advanced Goal:

- Modify the Speech Tones to reflect Personality



A5. Autonomous Learning of Physical Environments

Description:
- Simultaneous Localization and Mapping (SLAM) refers to the problem of incrementally building the map of a previously unseen environment while at the same time locating the robot on it.
- Active localization was proven that picking actions to minimize the localization's uncertainty would result in a better localization than using a passive approach.
- Active SLAM augments this approach to the SLAM problem, and it can be defined as the paradigm of controlling a robot which is performing SLAM to reduce the uncertainty of its localization and the map's representation

Goal:
- Robot Awareness of Physical Environments
- Robot Action with Environments

A6. Autonomous Learning of Tasks via Mimicking

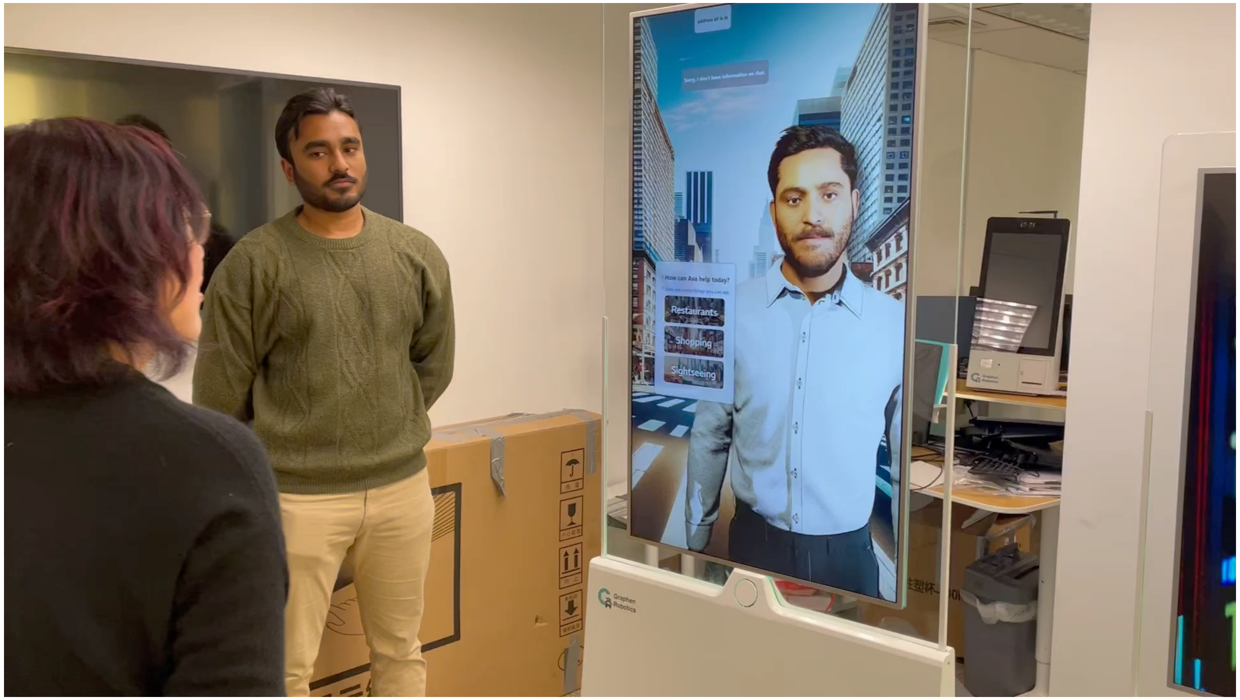
Description:

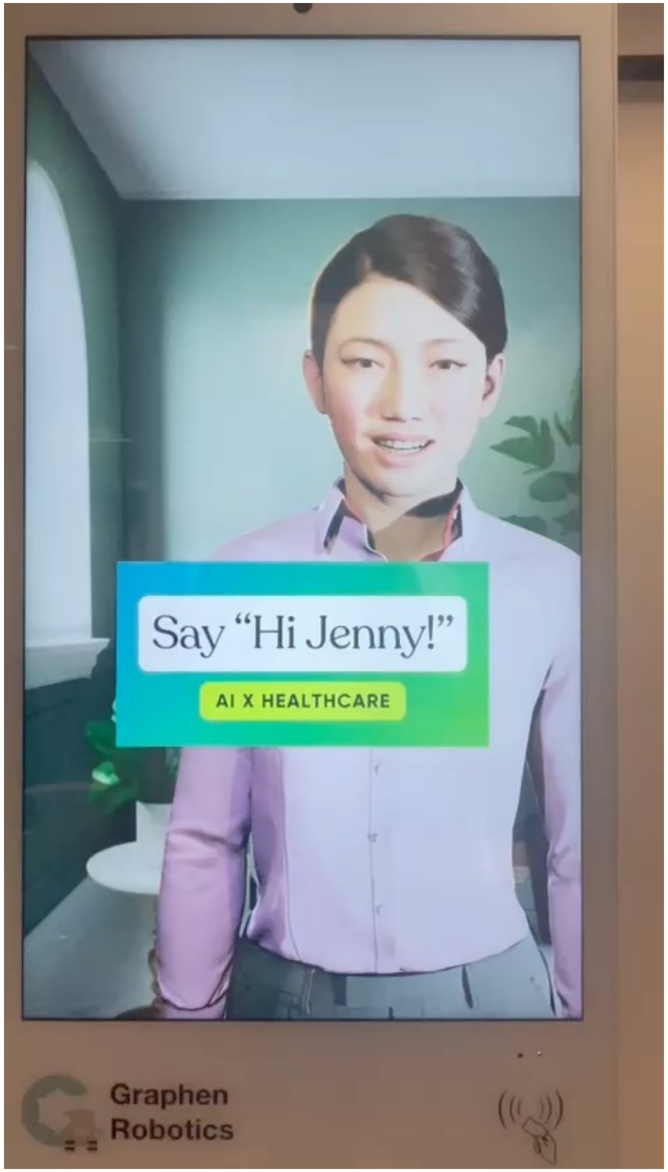
- Machine learning to act based on actions of human
- Watch how human activity in an environment and then learn how to behave by itself.

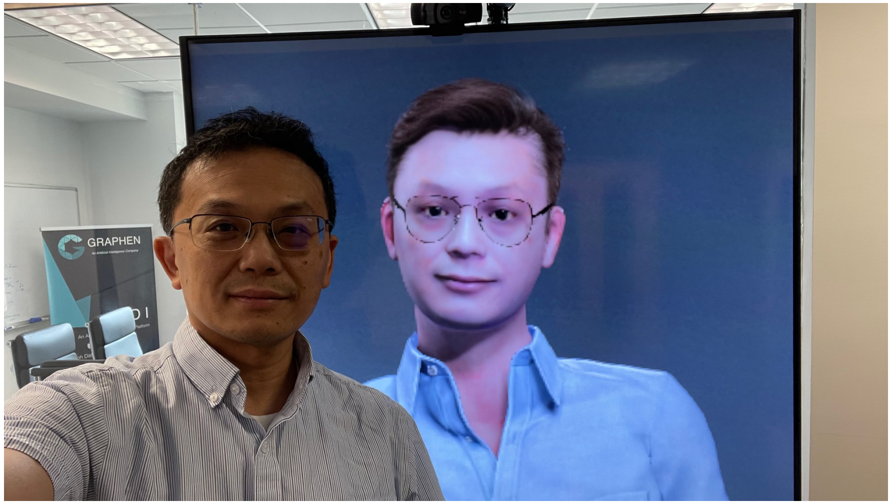
Goal:

- Observation and Action Extraction
- Reinforcement Learning to correct own actions

A7 - A10. Digital Human







A7. Digital Human for Fashion Industry

A8. Digital Human for Tourism Industry

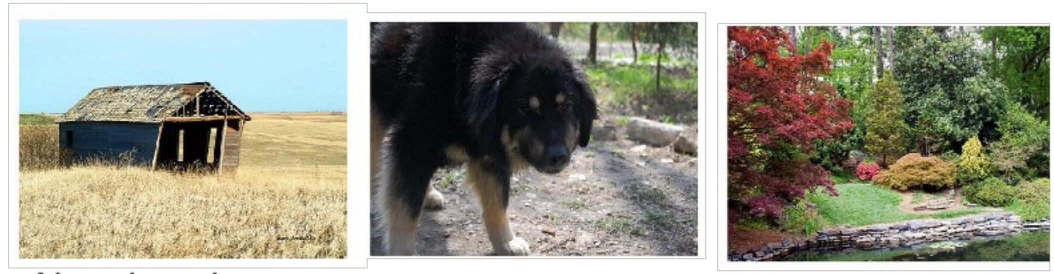
A9. Digital Human for Retail Industry

A10. Digital Human for Media and Marketing Industry

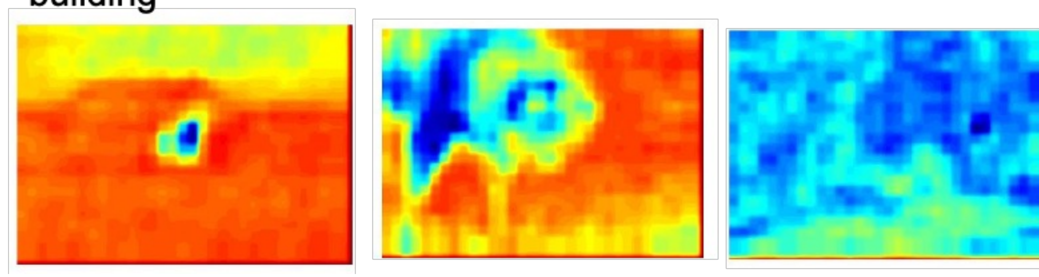
- Learning Industry Knowledges
- Local 'Brain'.
- Integrating with Mobile Apps.
- Multi-Languages
- Avatars with Personality & Emotion
- Reconstructing and Connecting with Real-World Objects
- (Optional) Utilizing with Physical Robotics

A11. Feeling and Art Recognition

- **Background**
  - Let machine to feel and appreciate arts like human
- **Project Goal**
  - A team will work on the subjective machine feeling of visual information
  - Allow machines to interpret arts.



Abandoned building | Scary dog | Colorful landscape



A12. Creative Writing and Story Telling

- **Background**
  - Overwhelming real-time information on media.
  - Automatic writing and telling a story based a set of news articles.

- **Project Goal**
  - A team will design and implement a platform that conducts data mining on various related media of a field.
  - Using NLP to summarize key text information.
  - Using visualization to create charts and graphs.
  - Automatically create descriptions

A13. Knowledge Learning and Construction

"Airplane"

"Grandma"

"Grandma is in Taiwan"

"Auntie is also in Taiwan"

"I like grandma"

"I like grandpa"

The boy said:

"I like grandma and grandpa"

Image Source: http://wonderforgood.com/category/visual-storytelling/

Memory with Knowledge Graphs

“see”

“Airplane”

“visit”

“Grandma”

“in”

“Taiwan”

“I”

"like”

“and”

“like”

“Grandpa”

“in”

“Auntie”

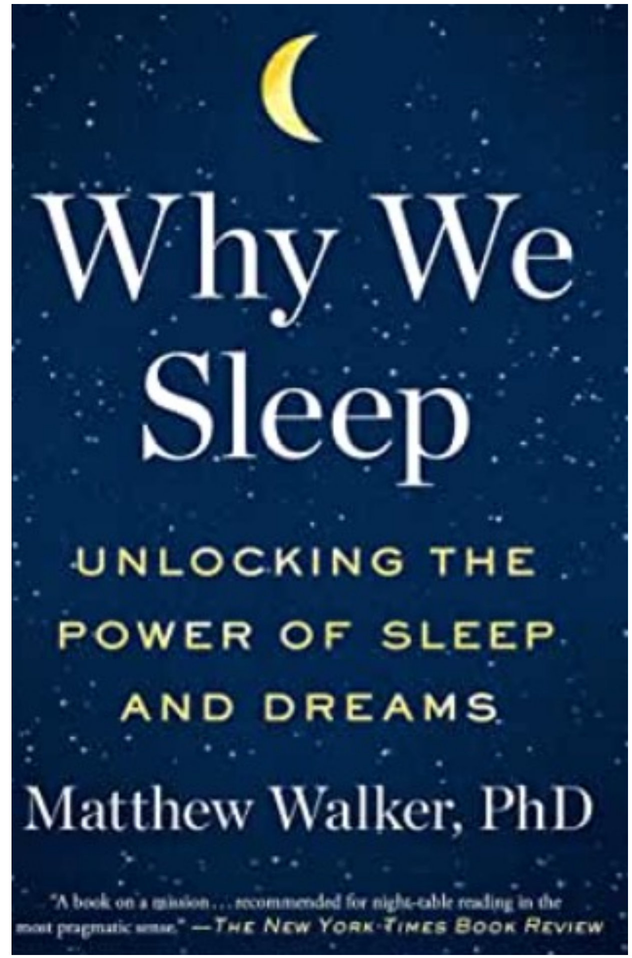
Image Source: http://wonderforgood.com/category/visual-storytelling/

Graph platform works like the human mind, connecting the dots when comprehending.

A14. Dreams Simulating Brain Functions while Sleeping

- **Background**
  - When human sleeps, our brain works on 'storing' the massive information we see , hear, and learn during the day time into 'storage'
  - Our brain would later on organize (in a bizarre way) to create dreams.

- **Project Goal**
  - A team will work on simulating how brain functions during sleep.
  - Create 'dreams',



A15. Self-consciousness, Ethics and Morality

- **Background**
  - Consciousness is how robots know its own existence
  - Can robot has self-identification?

- **Project Goal**
  - Simulates empathy
  - Simulates ethics and morality

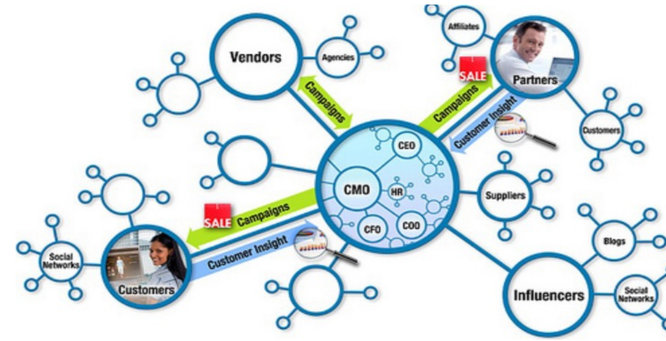
Task Area 2: Financial Advisors

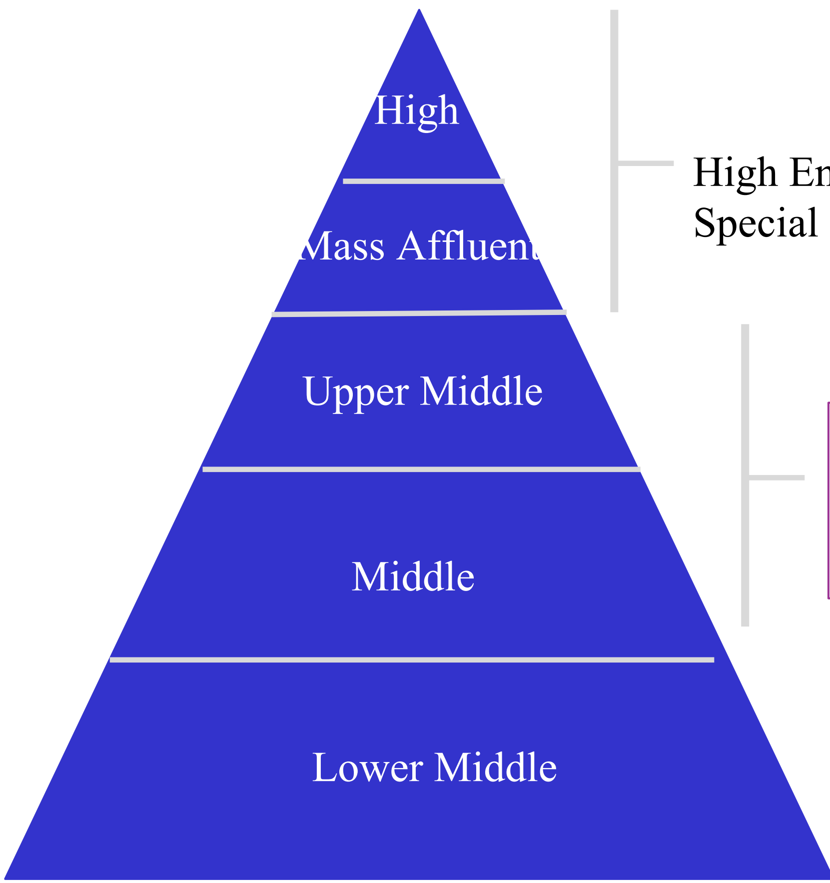
Market Data Analysis and Investment Targets

Advanced Dynamic 'Know Your Customer'

Optimized Personalized Investment Strategy

Bank-Customer Interaction Strategy





High End Customers （Private Bank / Special Investment Services）

Targeted Customers （Consumer Bank Services）: $15K - $1M
（Customer #: 30M～50M in China）

General Public （Consumer Bank Services）
（Customer # : > 1B in China）

Area 2 'Finance Advisor' Tasks List:

B1: Market Intelligence — Constructing Financial Knowledge Graphs
B2: Market Intelligence — Company Environmental, Societal, and Governance Performance
B3: Market Intelligence — Event Linkage and Impact Prediction
B4: Market Intelligence — Alpha Generation from Alternative Sources
B5: Advance KYC — Customer Profiling based on Personality, Needs, and Value
B6: Advanced KYC — Customer Behavior Prediction
B7: Investment Strategy — AI Trader (Foreign Exchange)
B8: Investment Strategy — AI Trader (Stock Markets)
B9: Investment Strategy — Automatic Dynamic Asset Allocation
B10: Customer Interaction — Customer Communication Strategies
B11: Customer Interaction — Insurance Product Sales & Marketing Strategy
B12: Automatic Story Telling for Marketing
B13: Automatic Market Competition Analysis
B14: Automatic Consumer Sales Leads Finding
B15: Human Capital Growth Recommendations

What is Robo-Advisor?

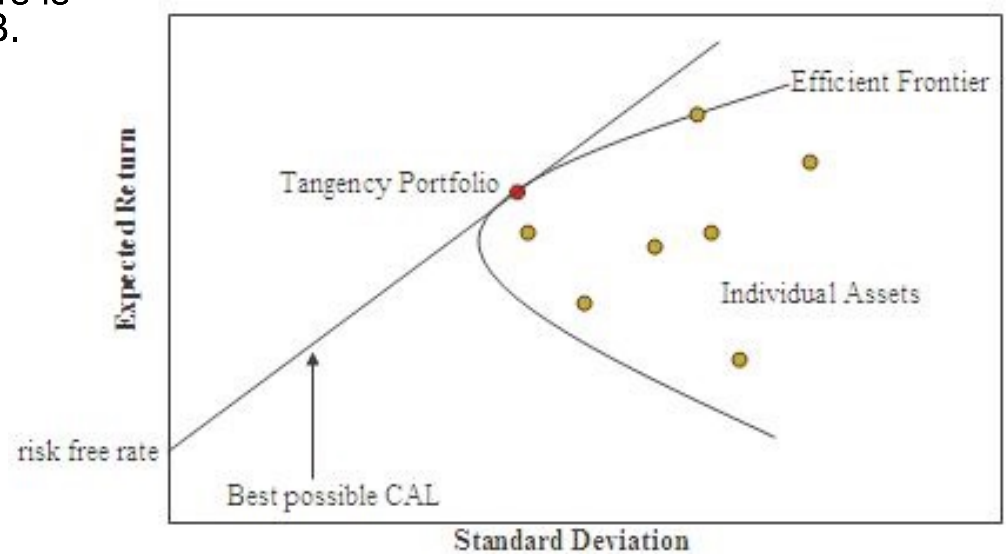
Robo-Advisor is a new type of wealth management service. Based on the risk level and investment goals provided by the investor, and it uses a series of 'smart algorithm' to calculate the optimal investment suggestions.

Robo-advisors directly managed about $19 billion as of December 2014. By 2020 the global assets under management of robo-advisers is forecast to grow to an estimated US$255B.

Features:

- **Strongly depend on technology, algorithm and financial theory**
- **Distributed investment, maximum long-term return**
- **Personalized portfolio allocation.**

- Non-biased
- Low investment threshold
- Low starting entry money
- Low agent fee



Example: Harry Markowitz Theory

Typical Steps of Robo-Advisory

Most of the robo-advisor platform is built based on the modern investment portfolio theory, using Exchange Trade Funds (ETFs) to build portfolio.

Customer Profiling	Construct Portfolio	Tracing Portfolio	Receiving Benefits	Rebalance
— design questionnaire; — Score Risk Capacity and Risk Willingness based on the answers of the questionnaire.	— portfolio strategy; — type analysis; — optimum allocation;	— Monte Carlo Simulation — Judge whether the goal is achieved — Suggest adjustments;	— Saving tax through the loss to compensate the gains; — outcome is highly related to the income; - Investment income tax (not applicable in China)	— set tolerance level to avoid over adjustment

Based on a survey of Wells Fargo, in US, there is only 16% of population in their 20s and 30s are willing to interact with investment consultants. The remaining people prefer to use these types of AI consultant.

Four Steps to use Big Data Cognitive Analysis for Robo-Advisor

Columbia University

Investment Market Analysis

- Analyze the market performance of various kinds of funds
- Analyze domestic and international financial and economic changes and how they may impact CPI, PPI, or GDP.
- Use Machine Learning and Deep Learning, based on historical economic numbers, find out how factors impact financial markets.

Data

- Product Data
- Market Data
- Historical Economic Data
- Industry-related Data

Dynamically Know Your Customer

- Customer Profiling, e.g, based on IPQ（Individual Profile Questionnaire), Feedback, Risk Capacity and Risk Willingness
- Understand what the customer really wants based on their past behaviors interacting with bank

Data

- Customer Data
- Behavior Data / Interaction Data

Optimized Personalized Investment Strategy

- Strategy computation and optimization based on personal history
- Demonstrate / Simulate 'what ifs' when the portfolio has different allocation.
- Explainability of 'what ifs' to customer to the customer.

Data

- Customer Data
- Market Data

Precise Bank-Customer Interaction

- Create and predict customer interaction strategy, including when, method, content to interact with customer – to achieve max customer and bank benefit.

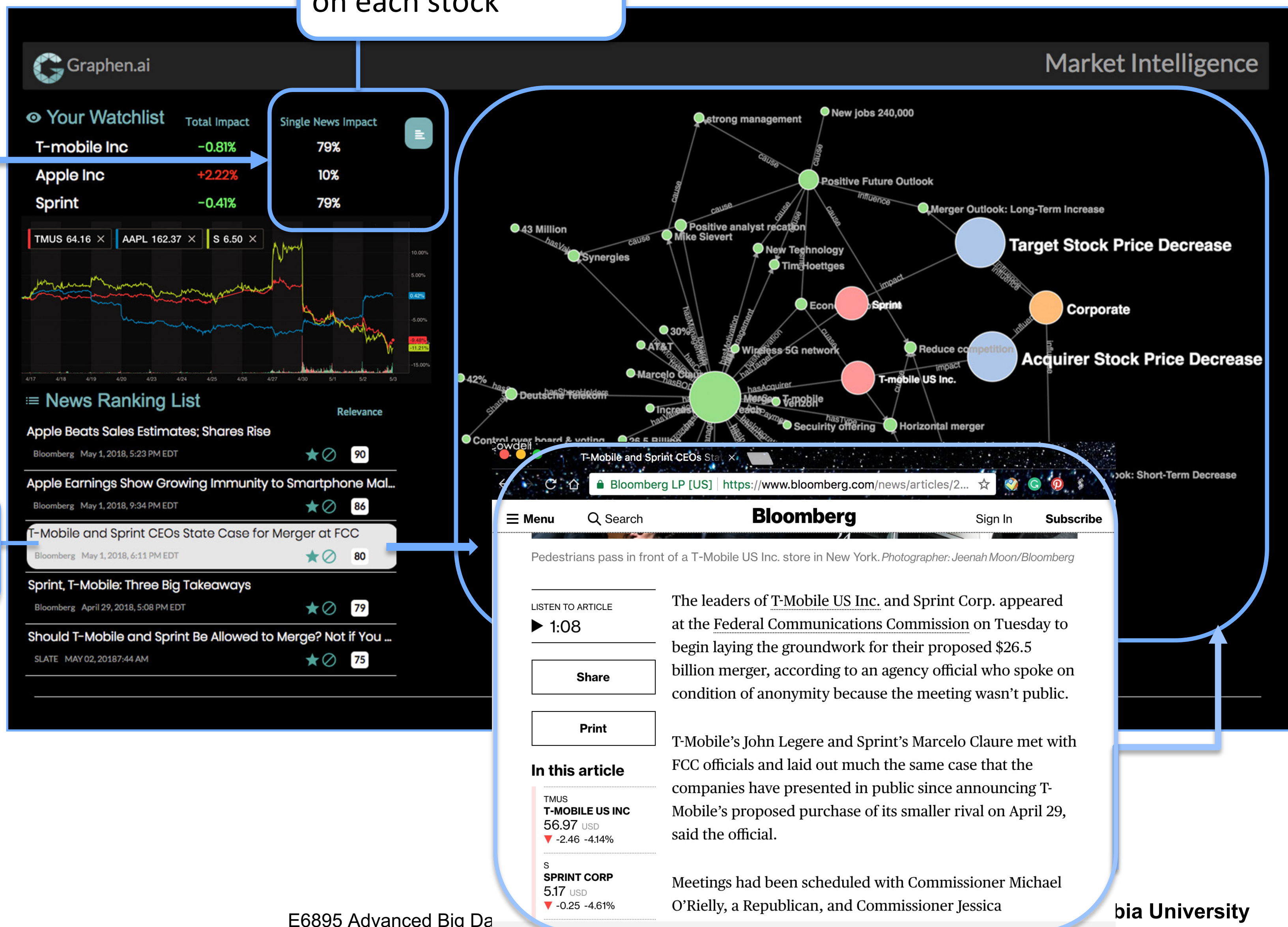
Data

- Customer Data
- Interaction Data

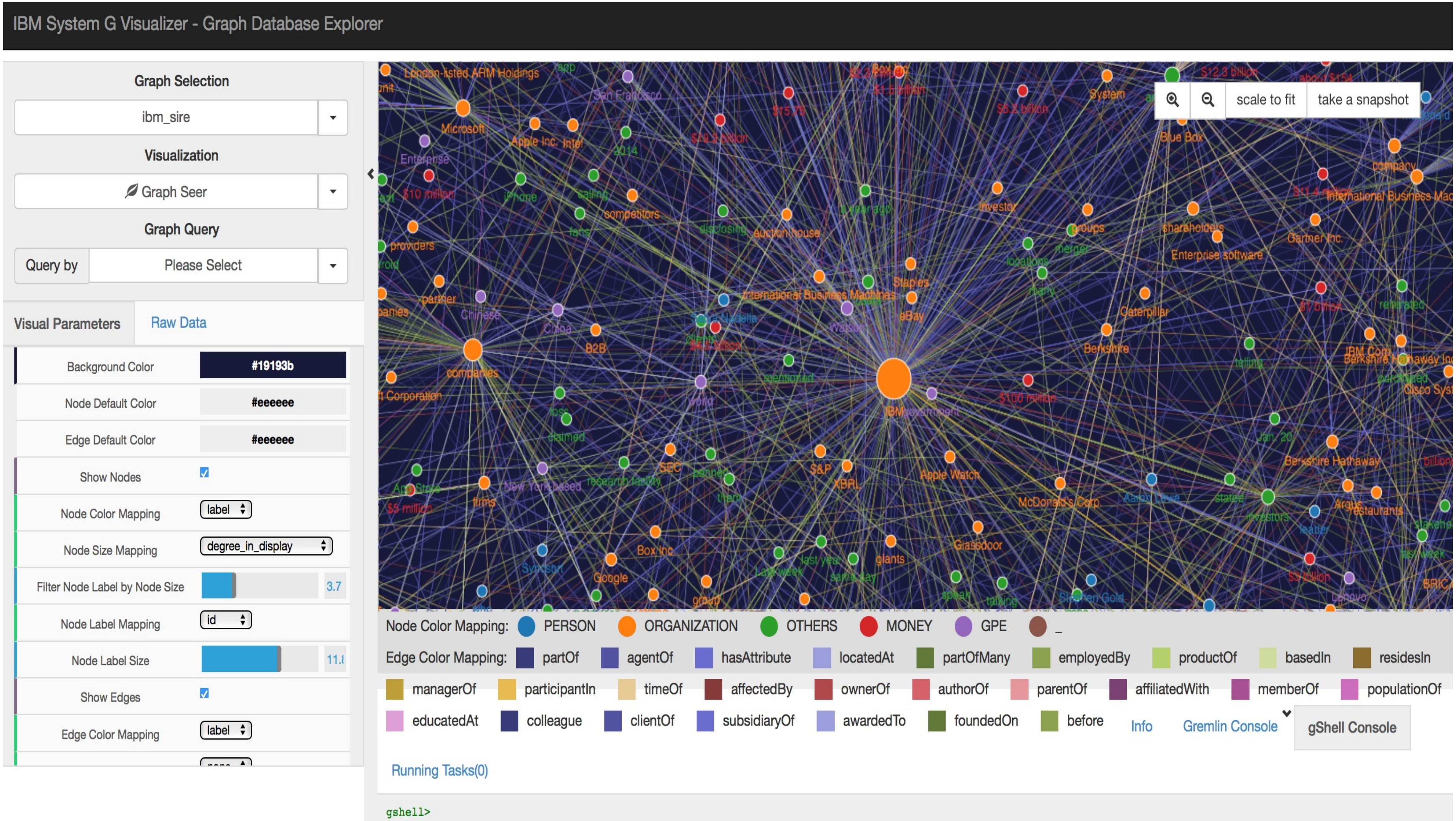
B1 - B4: Market Intelligence

Impact score of news on each stock

News to be analyzed

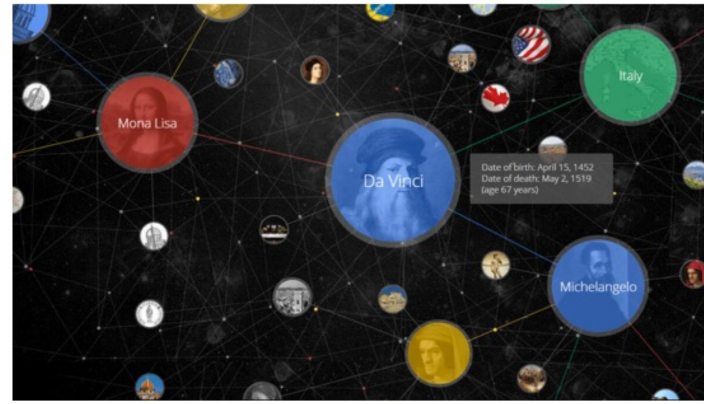


Knowledge Graphs

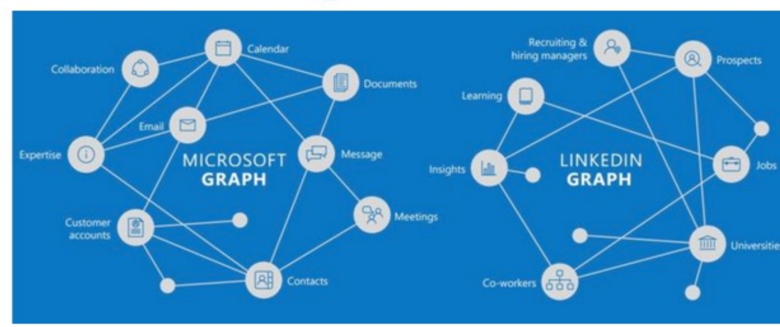


Example of Building and Utilizing Knowledge Graph

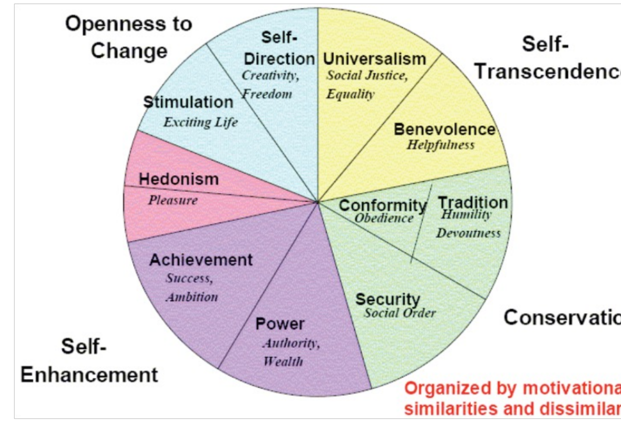
- **Background**
  - For artificial intelligence and better search, many search companies have created a knowledge graph.
  - However there are few knowledge graphs in the public domain.
- **Project Goal**
  - A team will create knowledge graphs in several application domains (e.g., Finance, Medical, etc) by crawling public web pages, news, Twitter, Wikipedia, etc.
  - A team will need to design the way of efficiently crawling data set, store them in a limited space, and quickly searching for required data set with the indexing functionality.

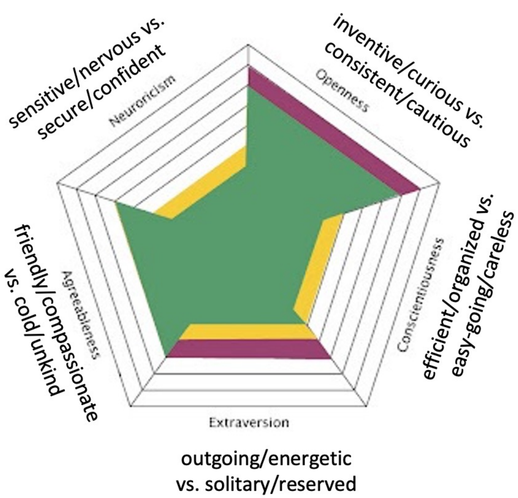


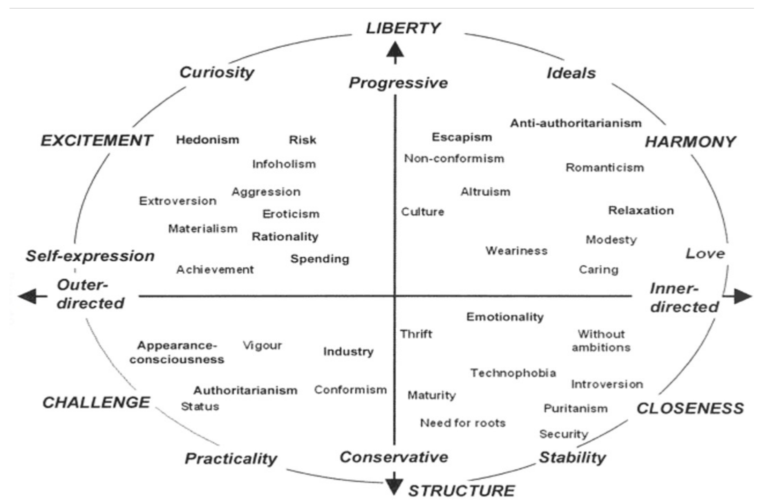
DBpedia



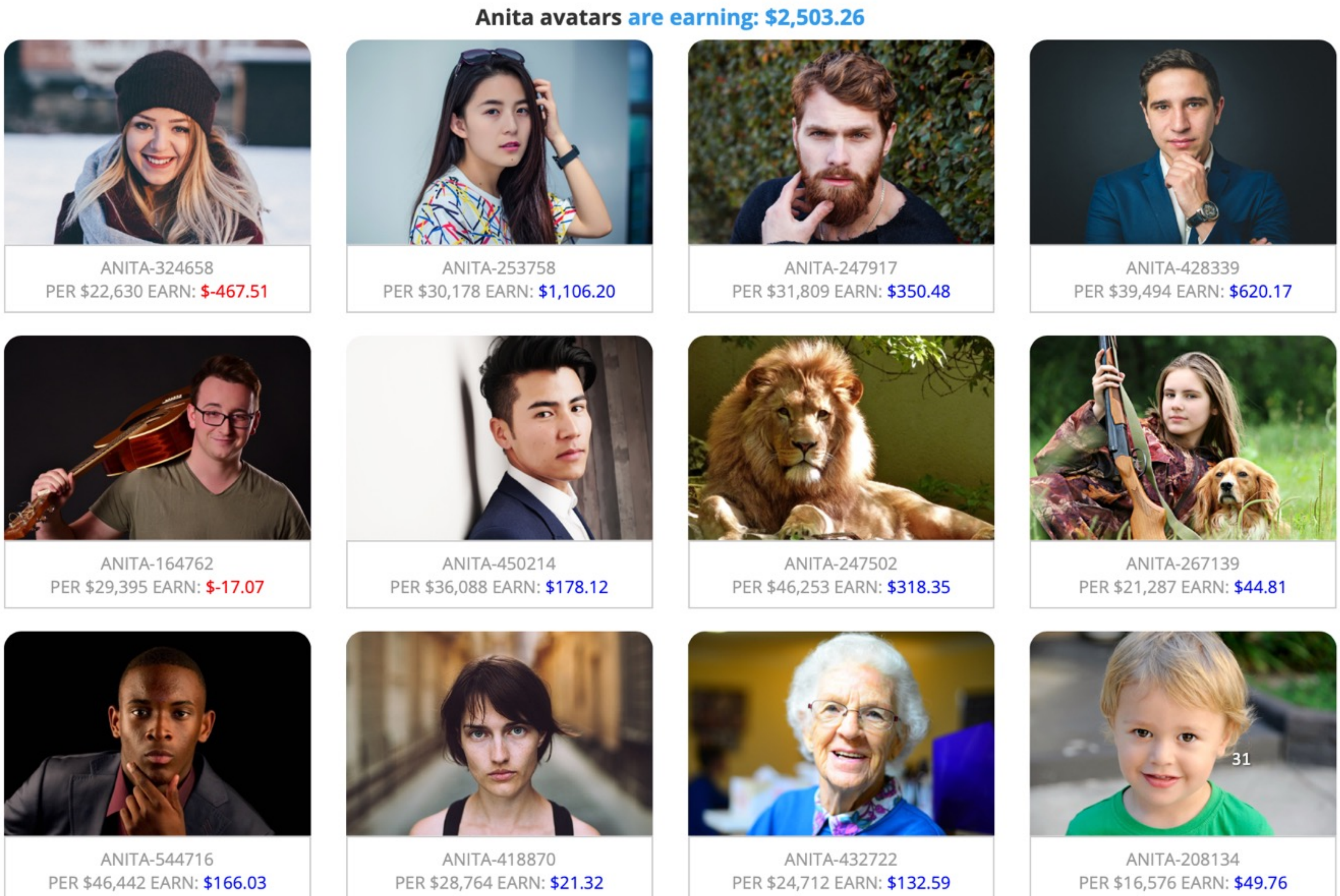
B5 - B6: Advanced KYC — Customer Profiling and Behavior Prediction







B7 - B9: AI Trader



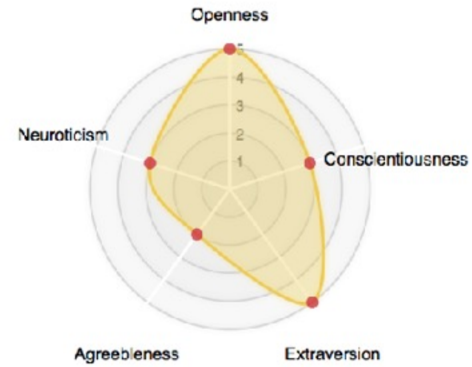
Anita 267139

-- an Adventurous AI Trader

***Specialized** at: EUR-USD*

***Knowledgable** of: Oil, Gold and Twitter*

***Strategy Learning Frequency** at: 2.0 hours*



Original: $1,000.00, Current: $1,404.50, Performance: *Gain $404.50*

Activities

Time	Action	Cash	Unit	Balance
2017-10-12 13:45:05	Sell 50,000	$1,404.50	0	$1,404.50
2017-10-12 12:57:25	Buy 100,000	$-57,792.00	50,000	$1,386.50
2017-10-12 11:19:10	Sell 100,000	$60,577.00	-50,000	$1,372.00
2017-10-12 11:11:55	Buy 100,000	$-57,822.00	50,000	$1,366.00
2017-10-12 09:08:05	Sell 100,000	$60,566.00	-50,000	$1,310.00
2017-10-12 08:34:40	Buy 100,000	$-57,935.00	50,000	$1,287.50

B10 - B11: Customer Interaction

Hundreds of products/campaigns Combinations with incompatibilities

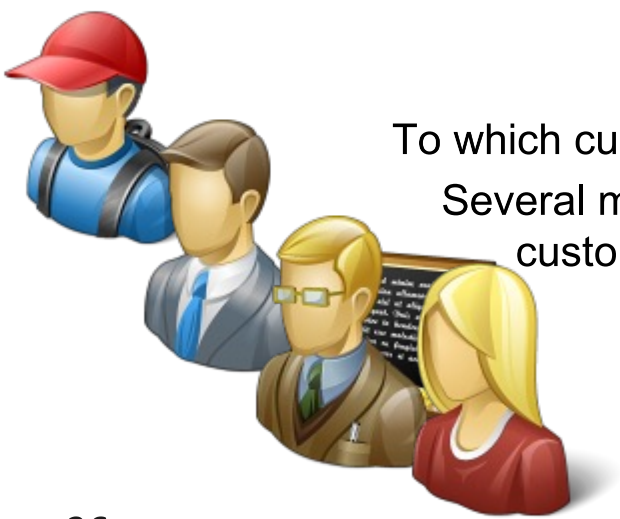
How much of each product/campaign ?

Telesales, Mail, email, Office, etc…

Done through which channel?

Nightly batch run, select over 1.2M

Experts doing what-if to improve process



To which customers?

Several millions of customers

When?
Select actions for the next days

B12. Automatic Story Telling for Marketing

- **Background**
  - Using Raw Materials in an organization to create marketing materials

- **Project Goal**
  - A team will design and implement platform that uses data in an organization
  - Automatically organize the information on a particular topic
  - Using visualization to create charts and graphs.
  - Manually or automatically create descriptions
  - Creates a video to tell story

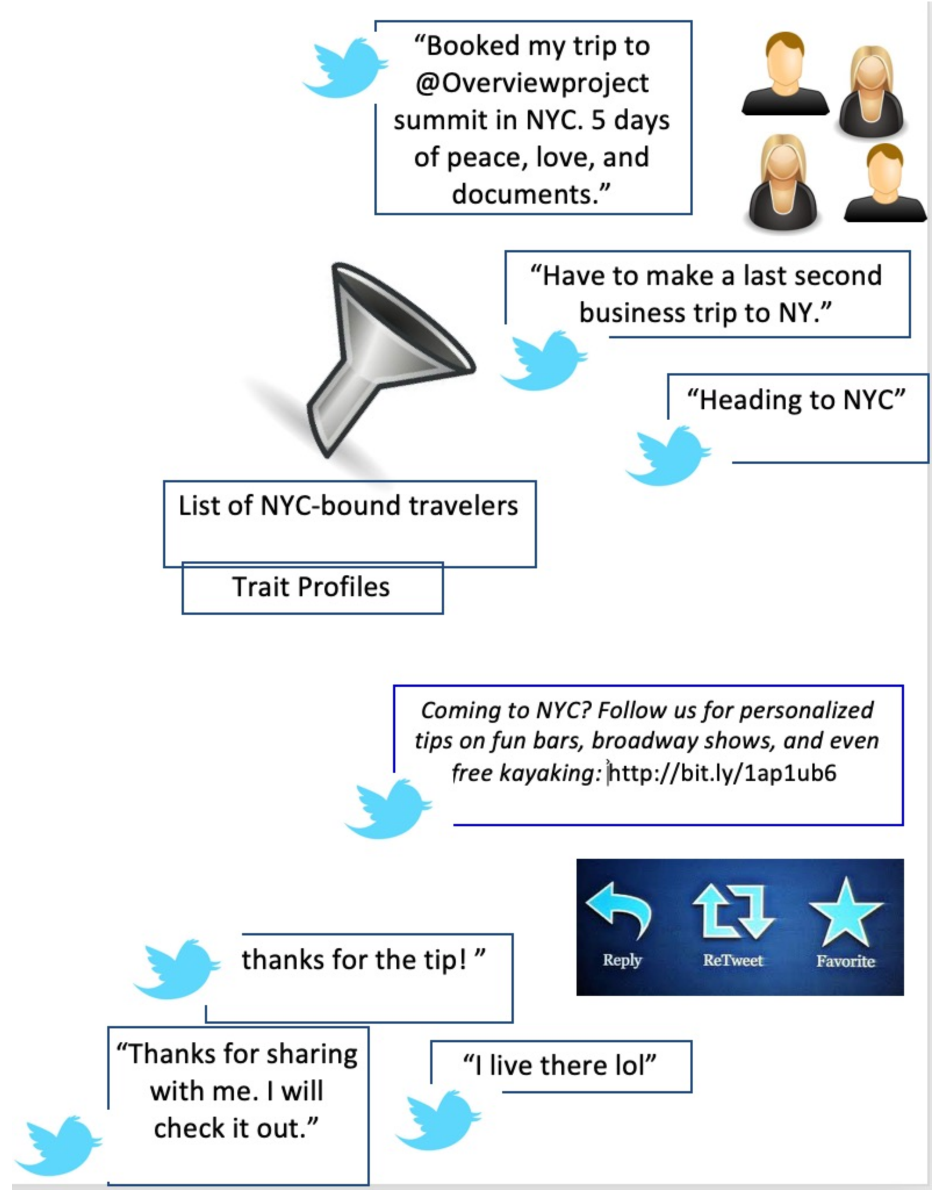
B13. Automatic Market Competition Analysis

- **Background**
  - Automatic searching internet to find competitor's information

- **Project Goal**
  - Automatic extraction of key information
  - Automatic compare key products and services of the company
  - Finding financial performance if those are available.

B14. Automatic Finding Sales Leads

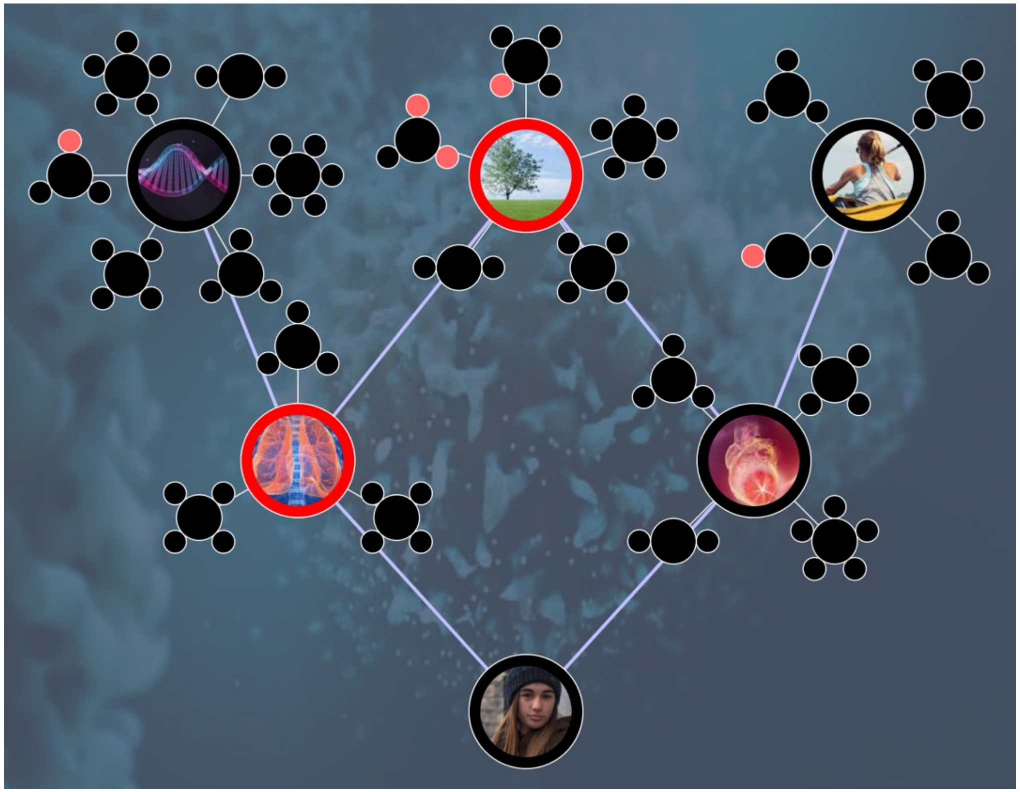
- **Project Goal**
  - Using Public Raw Materials on Social Media to find potential customers

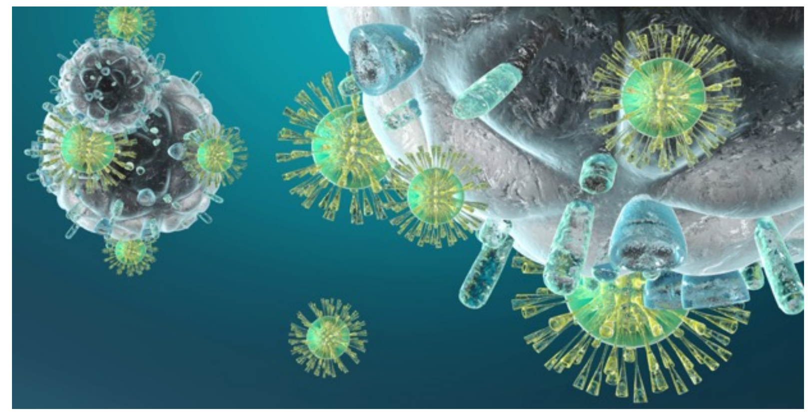


B15. Human Capital Growth Recommendation

- **Project Goal**
  - Automatic analyzing a person's personality and goals
  - Analyzing similar successful people from public datasources, e.g., LinkedIn.
  - Creating Knowledge Graphs that makes successful on goals
  - Suggesting what to learn to be competitive

Task Area 3: Healthy Life





Area 3 'Healthy Life' Tasks List:

C1: Precision Health — Gene and Protein Analysis of Network, Pathway, and Biomarkers

C2: Large-Scale System for Human Genome Analysis

C3: Genomic Mutations and Function Prediction

C4: Druggable Targets for Precision Medicine

C5: AI for Human Consciousness – EEG and AIoT

C6: AI for Human Consciousness – fMRI and Connectome

C7: Virtual Nurse -- Learning Medical Knowledge

C8: Virtual Doctor – Advanced Learning Medical Knowledge

C9: Virtual Doctor – Conversations

C10: Microbe and Disease Knowledge Graph

C11: Knowledge Graphs for Gene Interaction and Disease Relationships

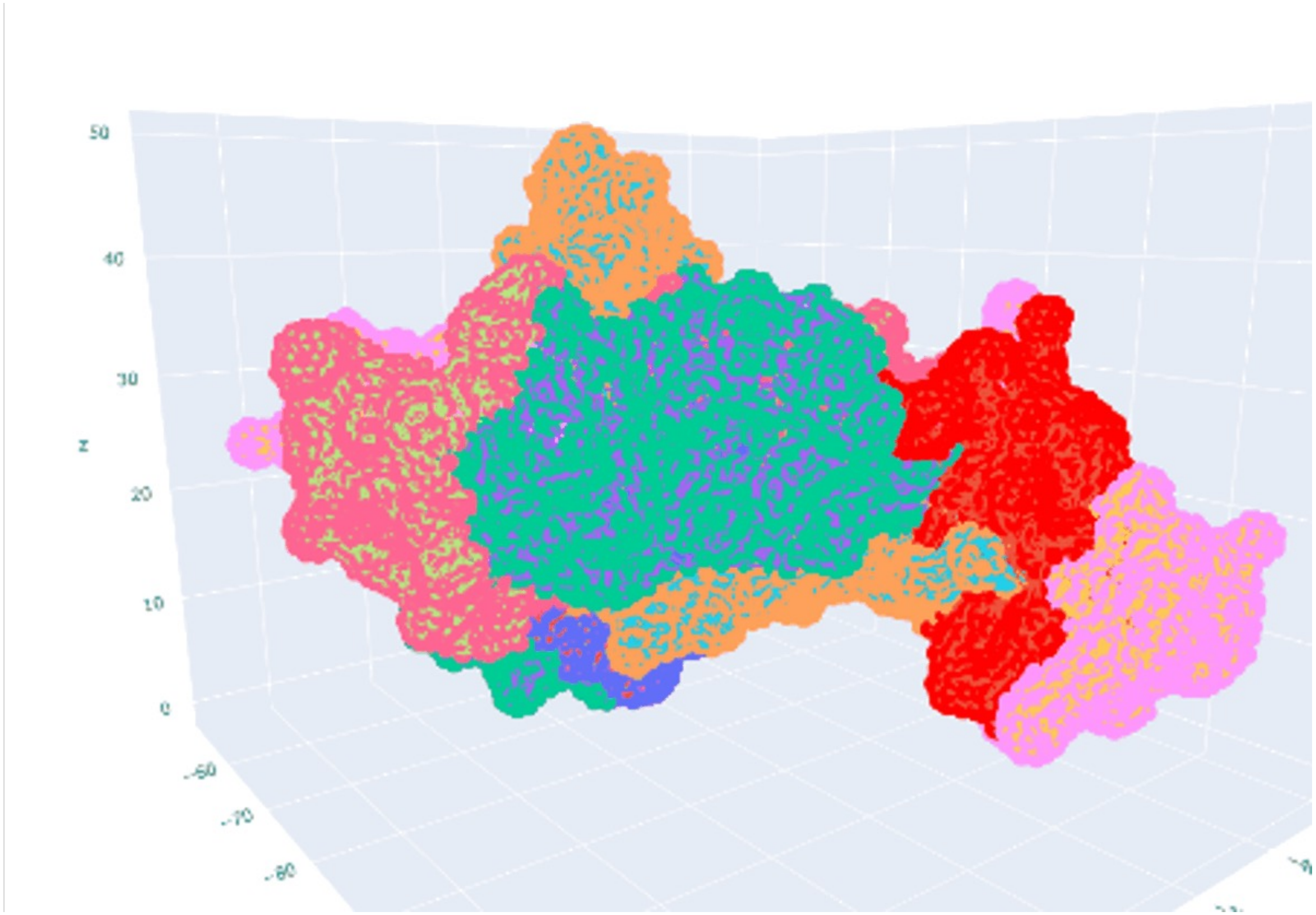
C12: Generating Gene or Immuno Therapy

C13: Molecular Drug Synthesis via Deep Learning

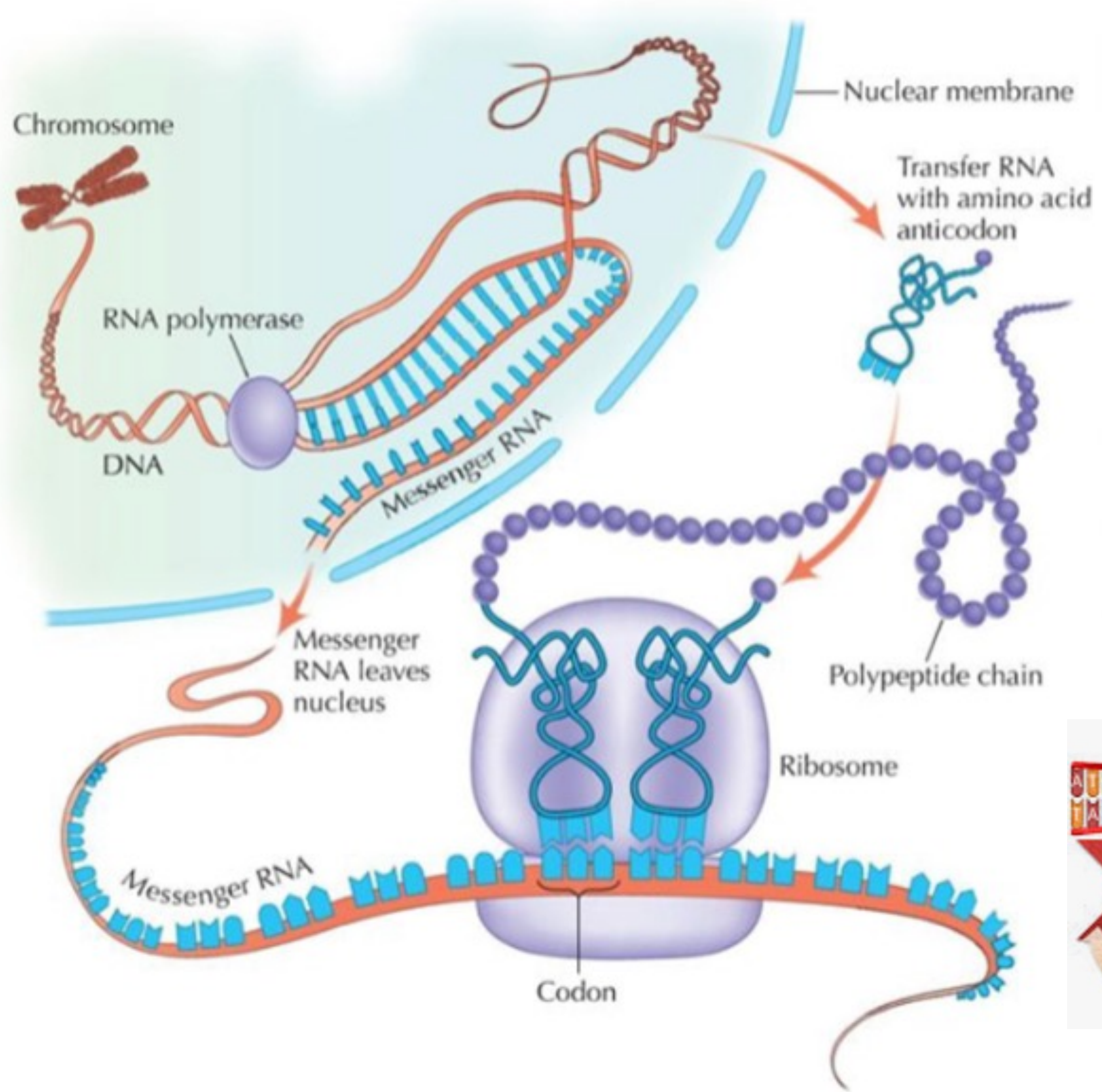
C14: Protein Interaction Predictor

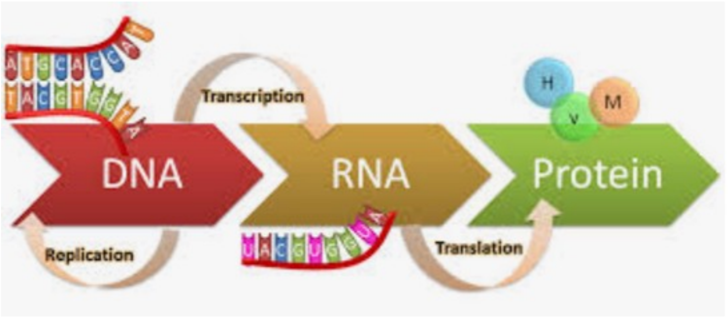
C15: AI Exploration and Understanding of Aging

Life is composed of graph of atoms

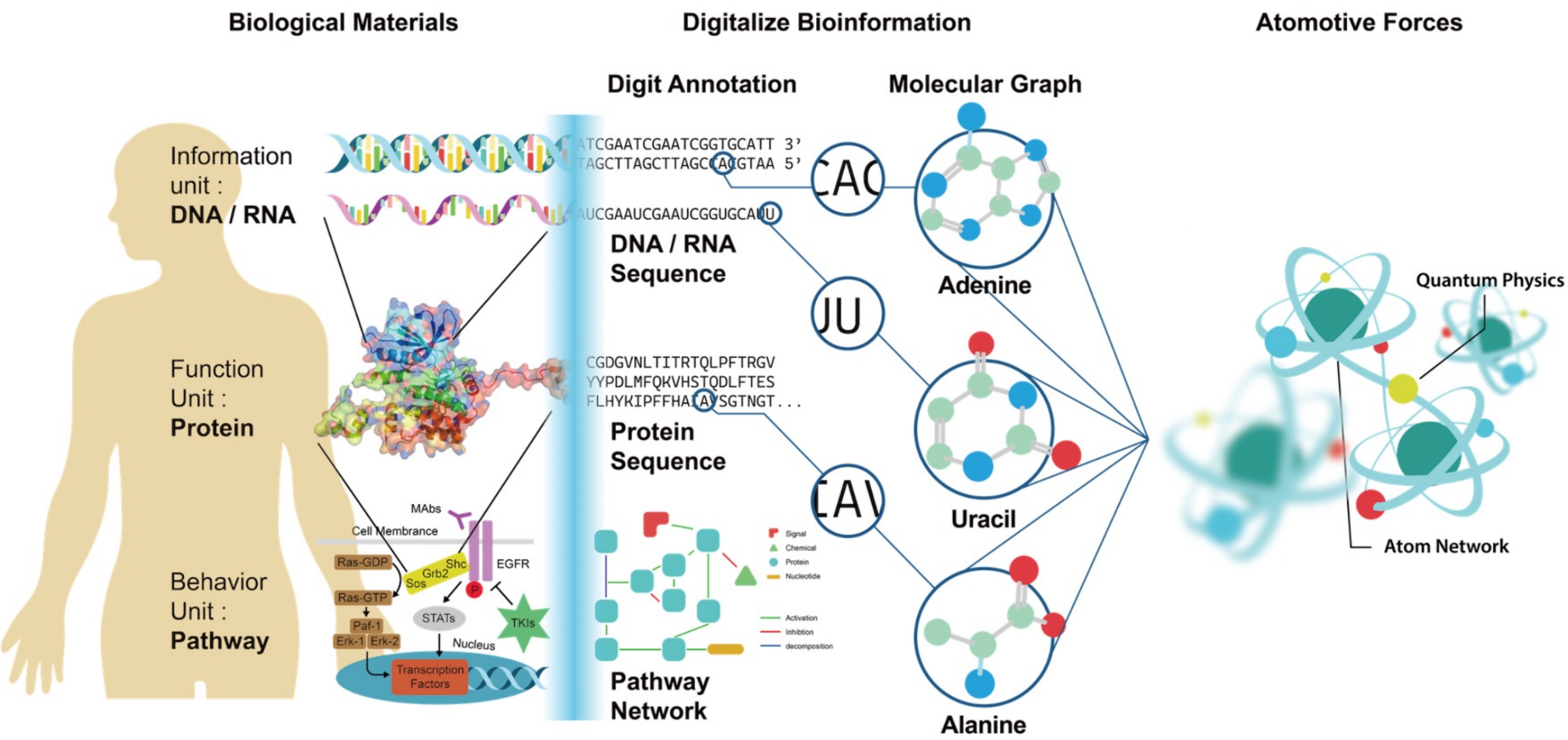


Central Dogma of Biology

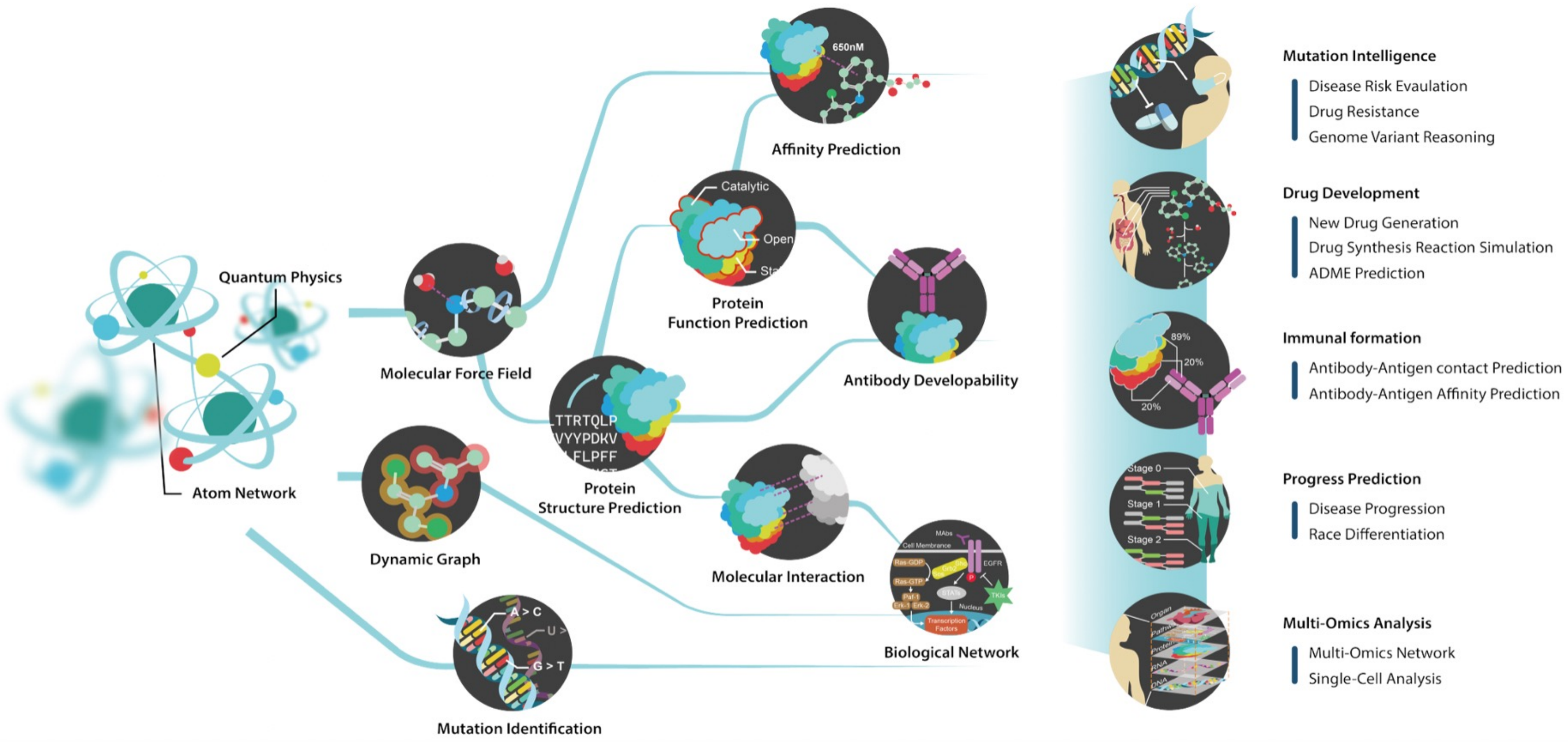




The Emergence of Digital Biology

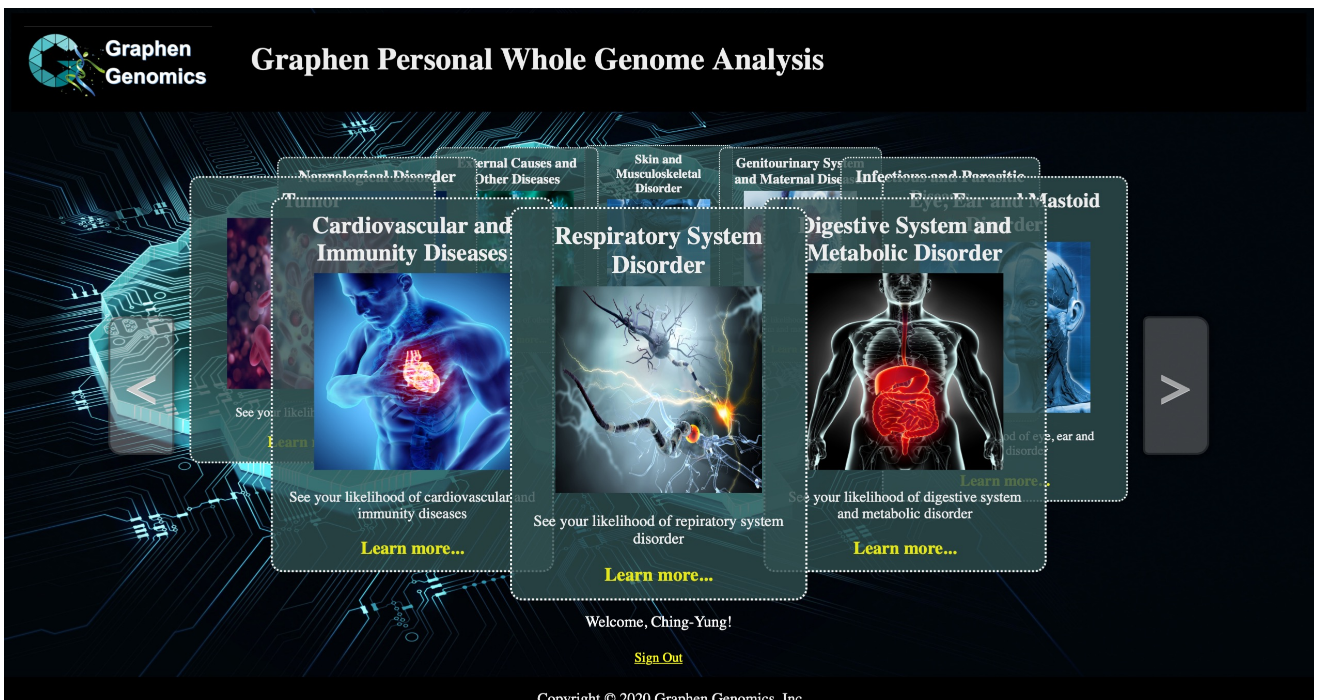


AI Tools power Digital Biology

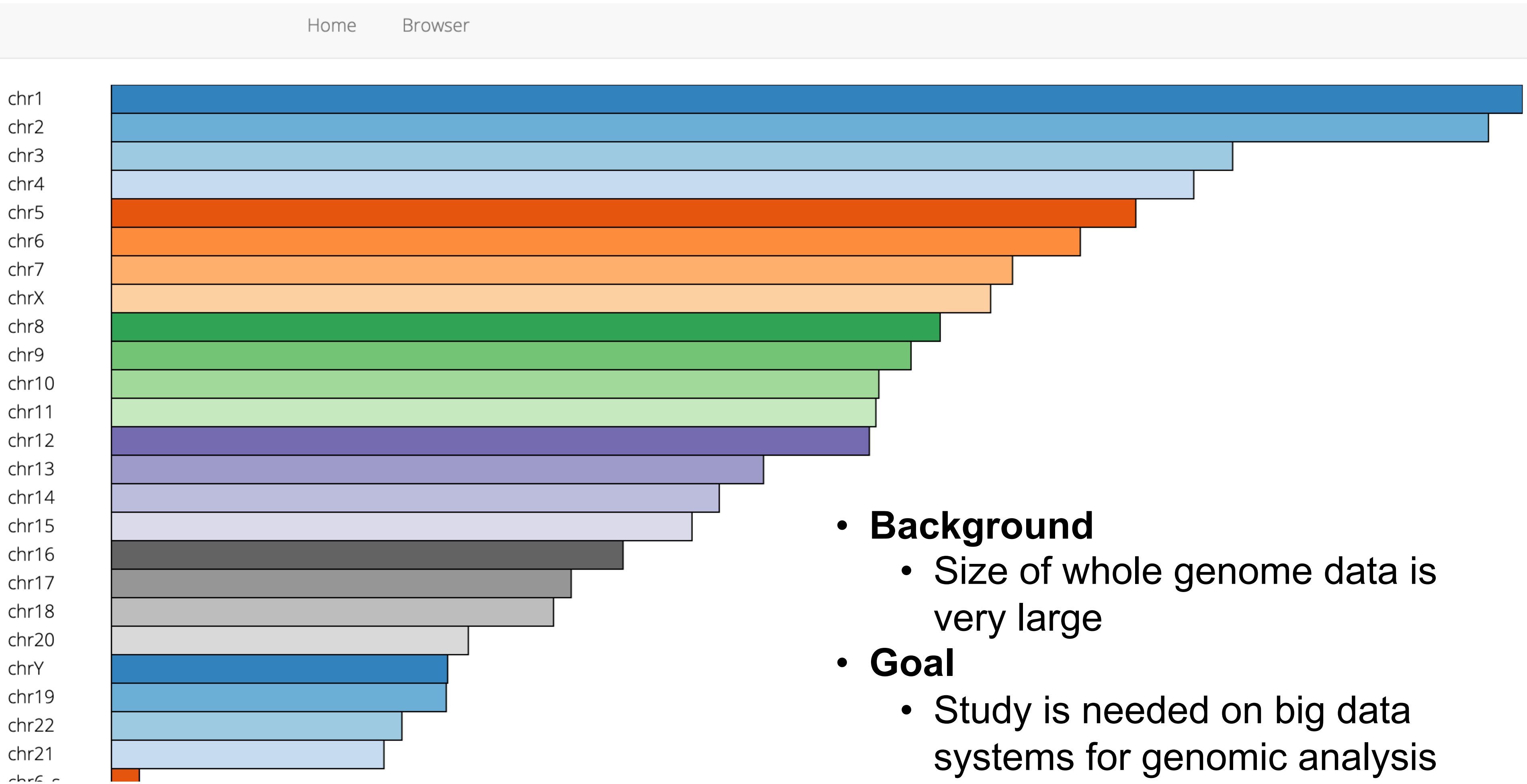


C1: Precision Health - Multiple Omics

- **Background**
  - Utilizing whole genome information can provide valuable information to patients
- **Goal**
  - Study open source whole genome data and explore their impact on disease prediction.



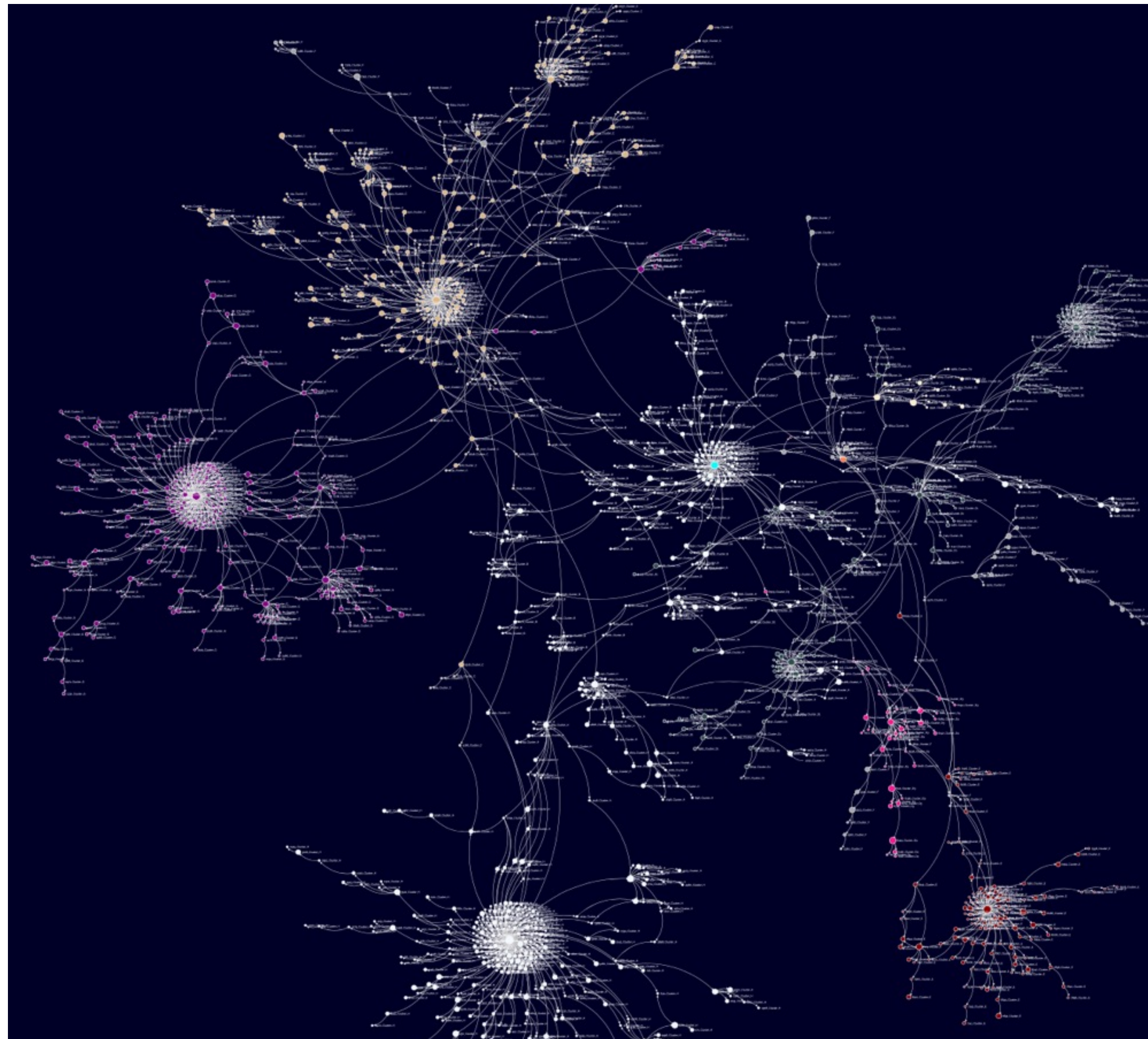
C2. Large-Scale System for Human Genome Analysis



- **Background**
  - Size of whole genome data is very large
- **Goal**
  - Study is needed on big data systems for genomic analysis and other comics analysis

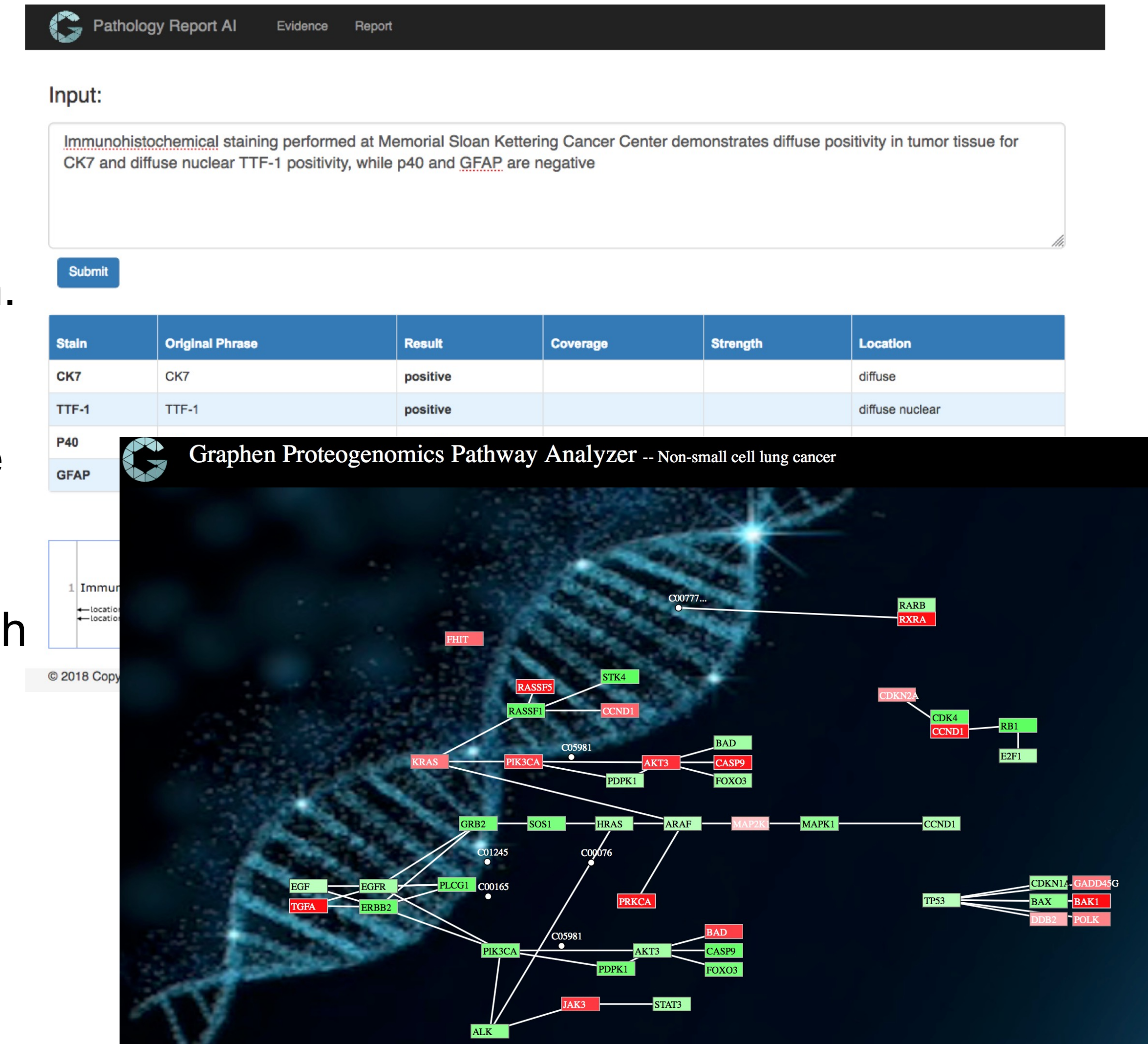
C3. Mutations and Function Prediction

- **Background**
  - We have been monitoring COVID-19 worldwide mutations since Feb 2020.
  - More than 12,000,000 virus strains have been sequenced
  - Continuous monitoring of large-scale data become more and more challenging.
- **Goal**
  - Keep exploring key algorithms for virus mutation classifications.
  - Use Protein function prediction tools to estimate the mutated virus impact.

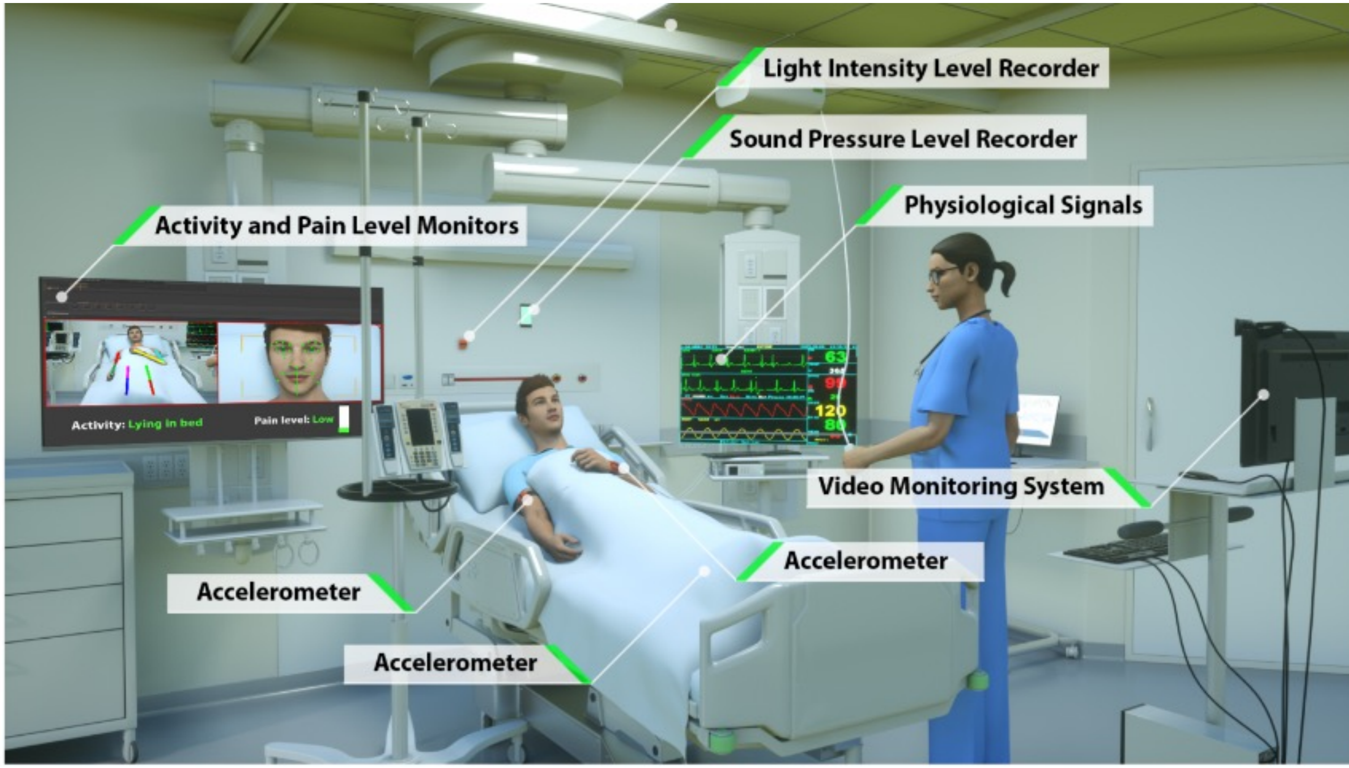


C4. Druggable Targets for Precision Medicine

- **Background**
  - Next-generation medicine will be based on personal genome data, proteome data, and pathway prediction.
  - It's a continuous challenging problem to explore the appropriate drugs for diseases
- **Goal**
  - Utilize Knowledge graph of disease and drugs.
  - Use the pathway analysis of patients to identify key variants
  - Analyze the potential drug targets



Building First Human Consciousness Monitoring and Prediction Open Platform



臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

GRAPHEN Building Next-Generation AI Solutions

NIH National Library of Medicine

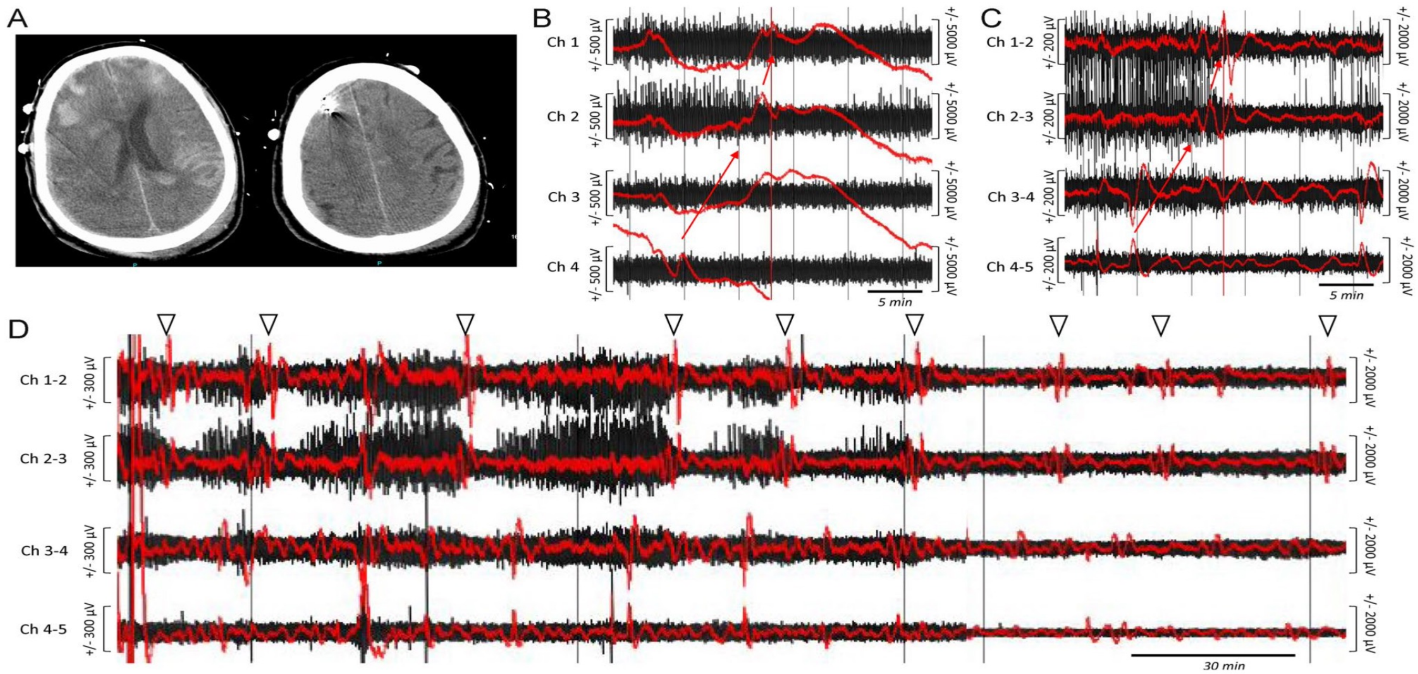
COLUMBIA UNIVERSITY

Massachusetts Institute of Technology

國立臺灣大學 National Taiwan University

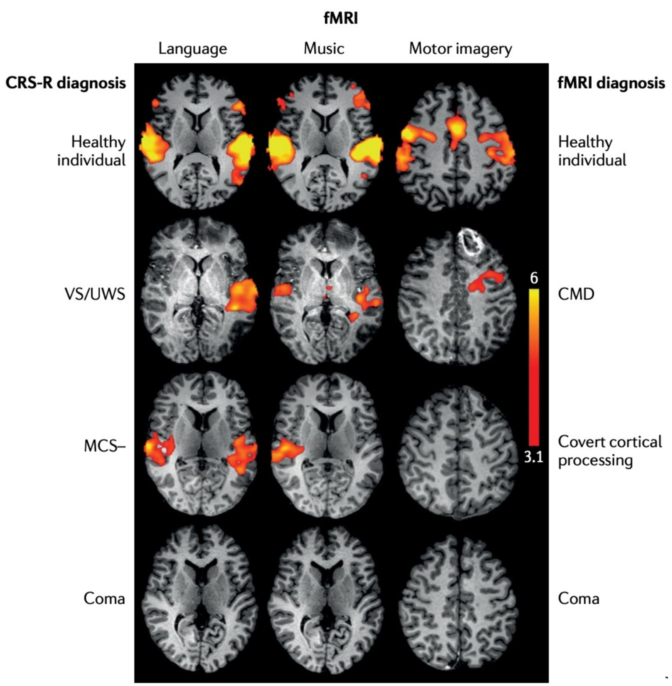
C5. AI for Human Consciousness -- EEG and AIoT

- **Background**
  - Human brain activities can be observed from sensing data
- **Goal**
  - Monitoring and Predicting Human Consciousness based on sensors, such as EEG sensors, biosensors, vital information, etc.

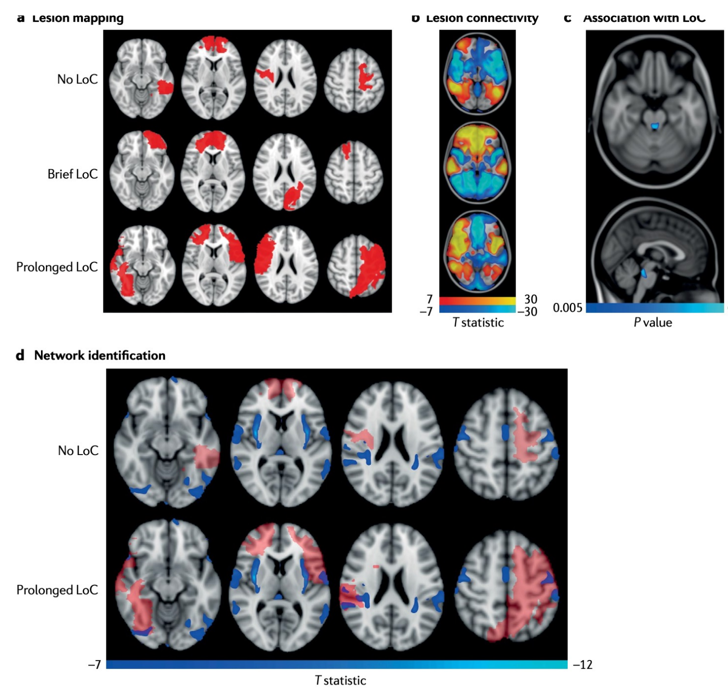


C6. AI for Human Consciousness – fMRI and Connectome

- **Background**
  - Human brain activities can be clearly observed from imaging data
- **Goal**
  - Monitoring and Predicting Human Consciousness based on medical images, such as CT, fMRI, Connectome, etc.



f-MRI

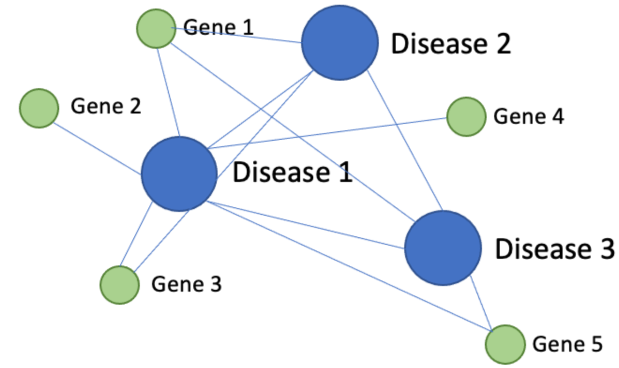


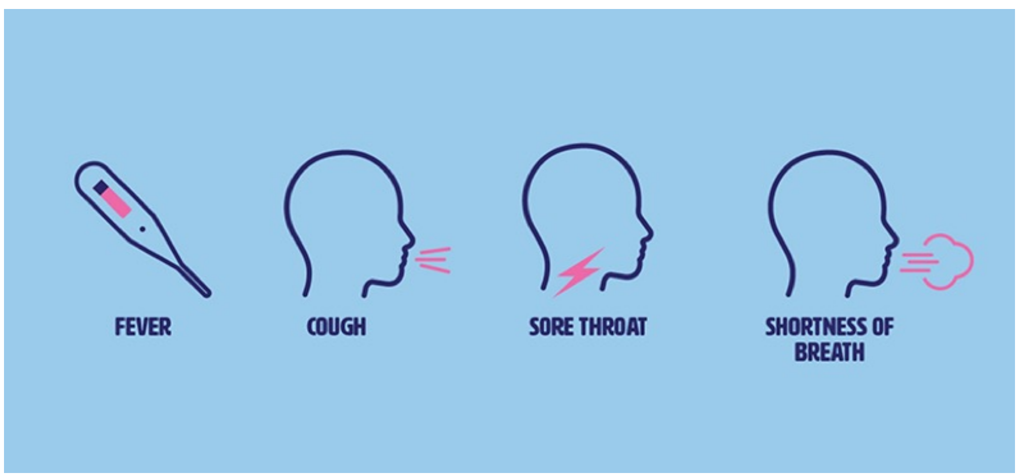
Connectome

Columbia University

C7. Virtual Nurse – Medical Knowledge Learning

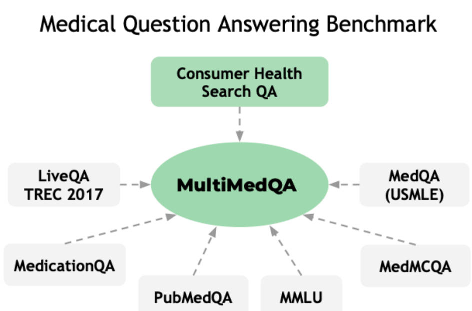
- **Background**
  - Big Data and AI technologies have significant progress lately. It becomes possible to learn knowledge from diverse sources.
- **Goal**
  - Establish AI system that can potentially pass the New York state nurse exam.

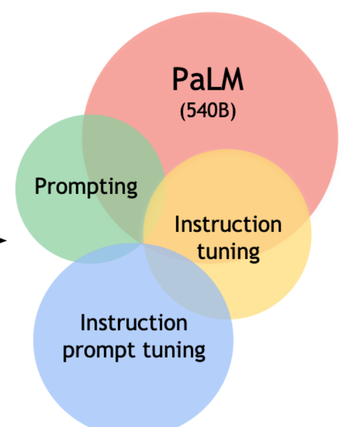


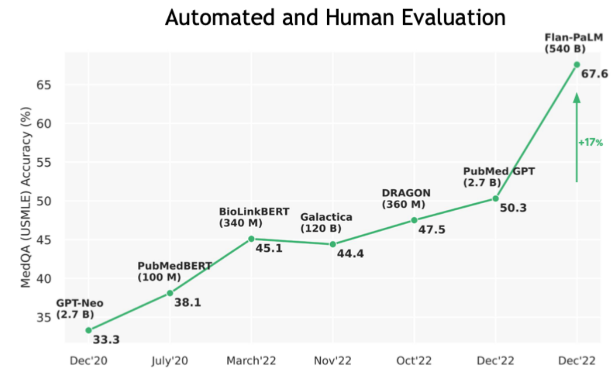


C8. Virtual Doctor – Advanced Medical Knowledge Learning

- **Background**
  - Big Data and Deep Learning technologies have been significantly progressed lately. It can probably pass the Doctor Qualification Exams
- **Goal**
  - Exploring Large Language Models and open Medical and Health datasets to learn medical knowledge







Q: How long does it take for newborn jaundice to go away?

Med-PaLM: Newborn jaundice is when a newborn baby's skin and eyes look yellow. It is caused by a buildup of a substance called bilirubin in the baby's blood. Bilirubin is a byproduct of the normal process of red blood cells breaking down. Newborn jaundice is common and typically harmless, but it can be a sign of a more serious condition. The jaundice typically goes away on its own within a few weeks. In some cases, it may last longer or require treatment. The time it takes for the jaundice to go away can vary depending on the cause and the severity of the jaundice. If the jaundice is severe or lasts longer than a few weeks, the doctor may recommend testing or treatment to determine the cause and prevent complications.

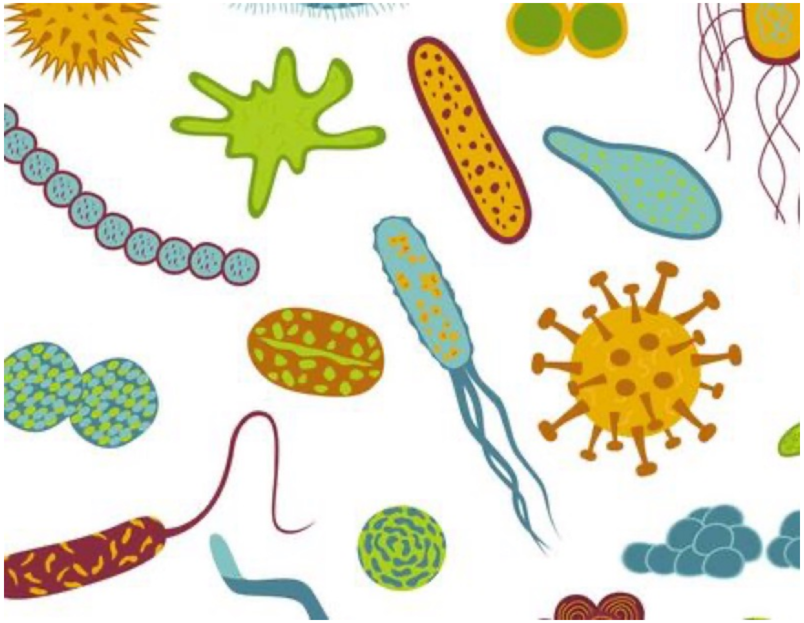
Med-PaLM performs encouragingly on consumer medical question answering

C9. Virtual Doctor – Conversations

- **Background**
  - With deep medical knowledge, it is becoming possible for building virtual doctors who can interact with patients
- **Goal**
  - Prototyping Virtual Doctors who can communicate with patients; observing from multi-modality information and QA from patients.

C10. Microbe and Disease Knowledge Graph

- **Background**
  - Microbes are tiny living things that are found all around us and are too small to be seen by the naked eye. They live in water, soil, air, and in human body, which is also called microorganisms. The most common types are bacteria, viruses and fungi.
  - Researches indicate that microbes and human health have strong correlations.
- **Goal**
  - Find the similarity of microbes and similarity of diseases.
  - Build the correlation network of microbes and diseases to help diagnose potential health conditions.

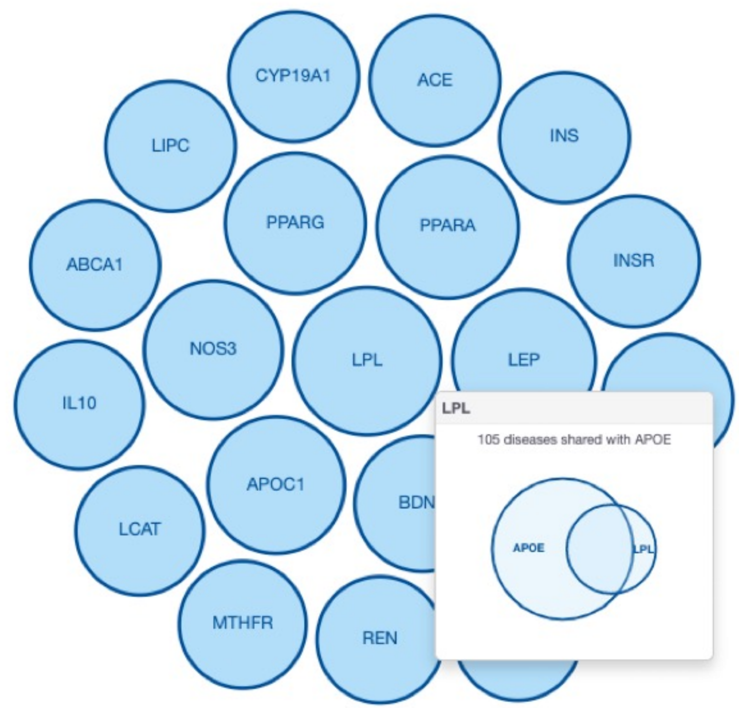


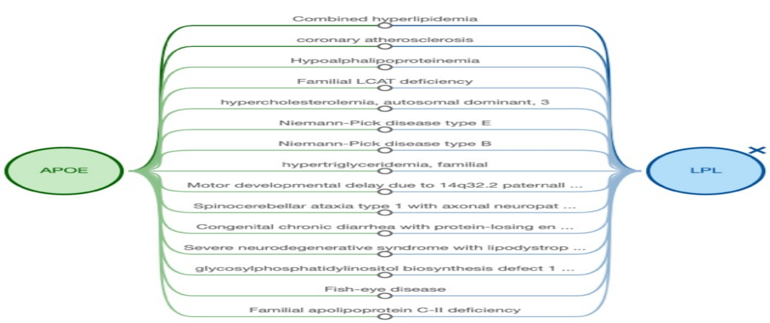
C11. Knowledge Graph of Gene Interaction and Disease Similarity

- **Background**
  - Understanding the genetic networks and their associations in diseases is one of the important objectives of biological researchers. The knowledge graph serves as a powerful tool to investigate this topic.
- **Goal**
  - Construct and visualize knowledge graphs demonstrating associations among genes based on disease similarity.

The top 20 targets related to APOE based on a computation of similar target-to-disease connections and overall association scores. Clicking on a related target bubble will show you a list of the diseases used to calculate the similarity between the two targets. For more information on this visualisation and how we compute similarity scores, please read our documentation on target and disease similarity.

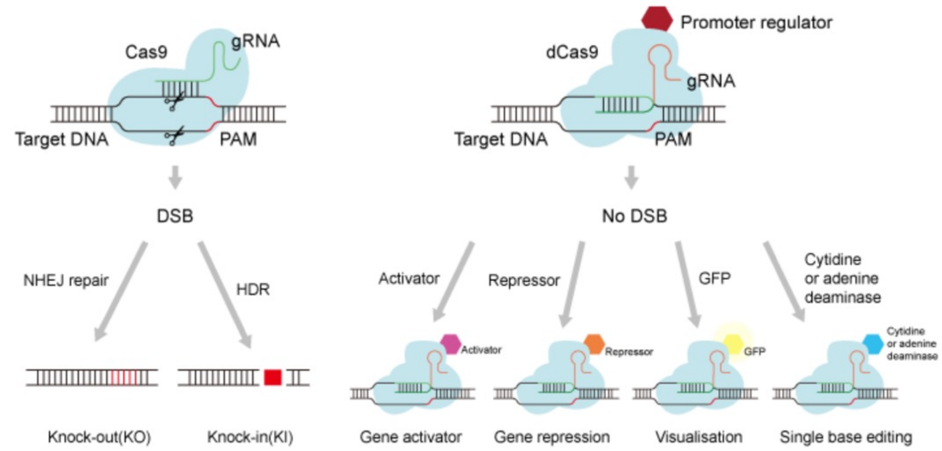
Source: Open Targets

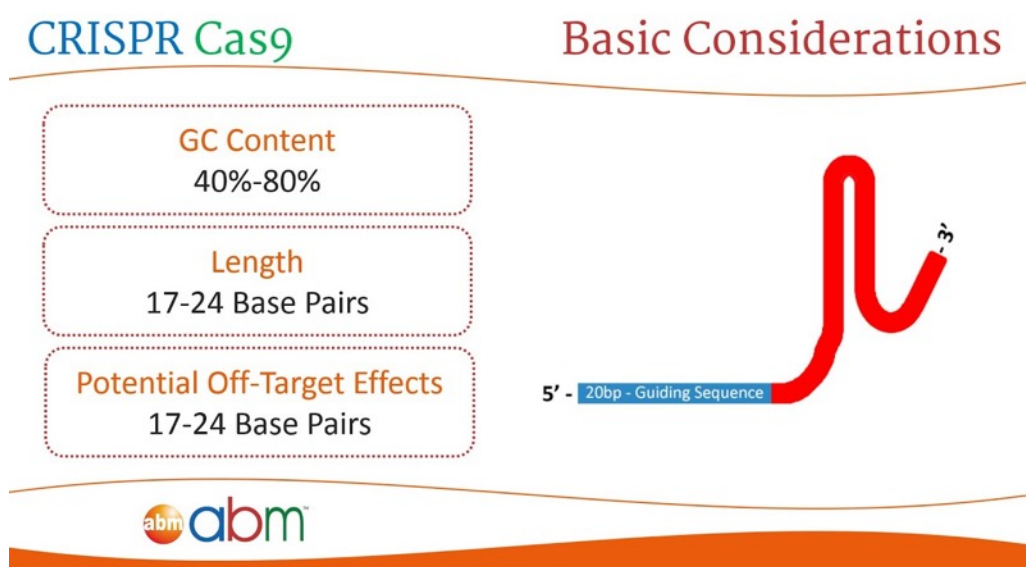




C12. Generating Gene or Immuno Therapy

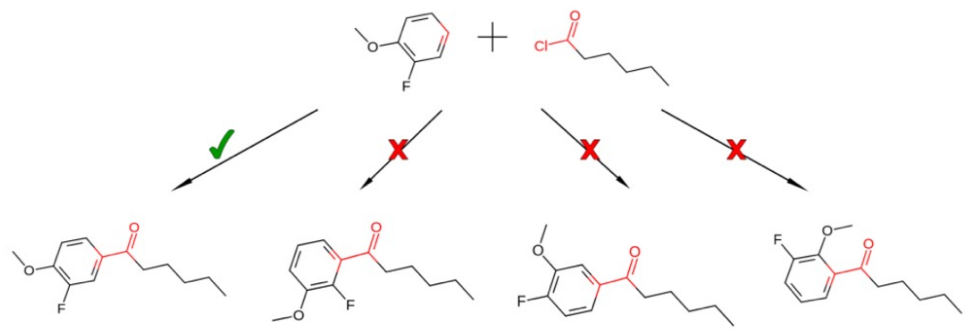
- **Background**
  - CRISPR, allowing to precisely edit the genome of cells by inducing double-stranded DNA (dsDNA) breaks at specific loci, is both an efficient and cost-effective technological tool.
  - But how to design perfect sgRNA for detect cell DNA and without any off-target is challenging
  - We need use outsourcing data to make a deep learning algorithm to solve this problem.
- **Goal**
  - Use TDC open source data to predict what kind of sgRNA have ability to edit or repair cell DNA. And design an auto encoder or GAN to generate template sequence.



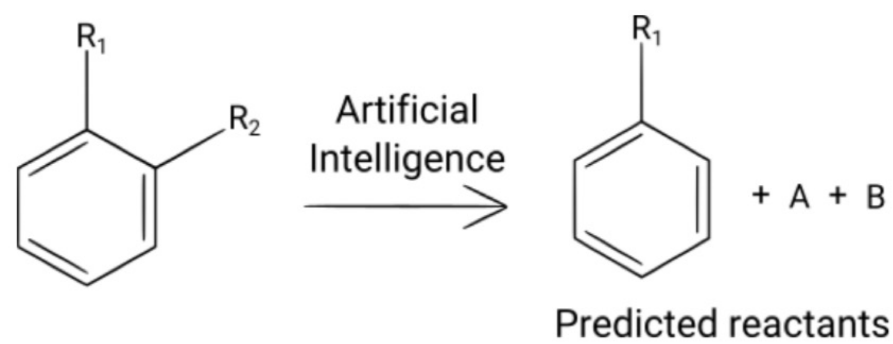


C13. Molecular Drug Synthesis via Deep Learning

- **Background**
  - Generate small molecular by deep learning will not be hard. In fact, these molecular are hard to be manufacture or synthesis. We need to design a algorithm to simulate chemical reaction and predict molecular synthesis feasibility.

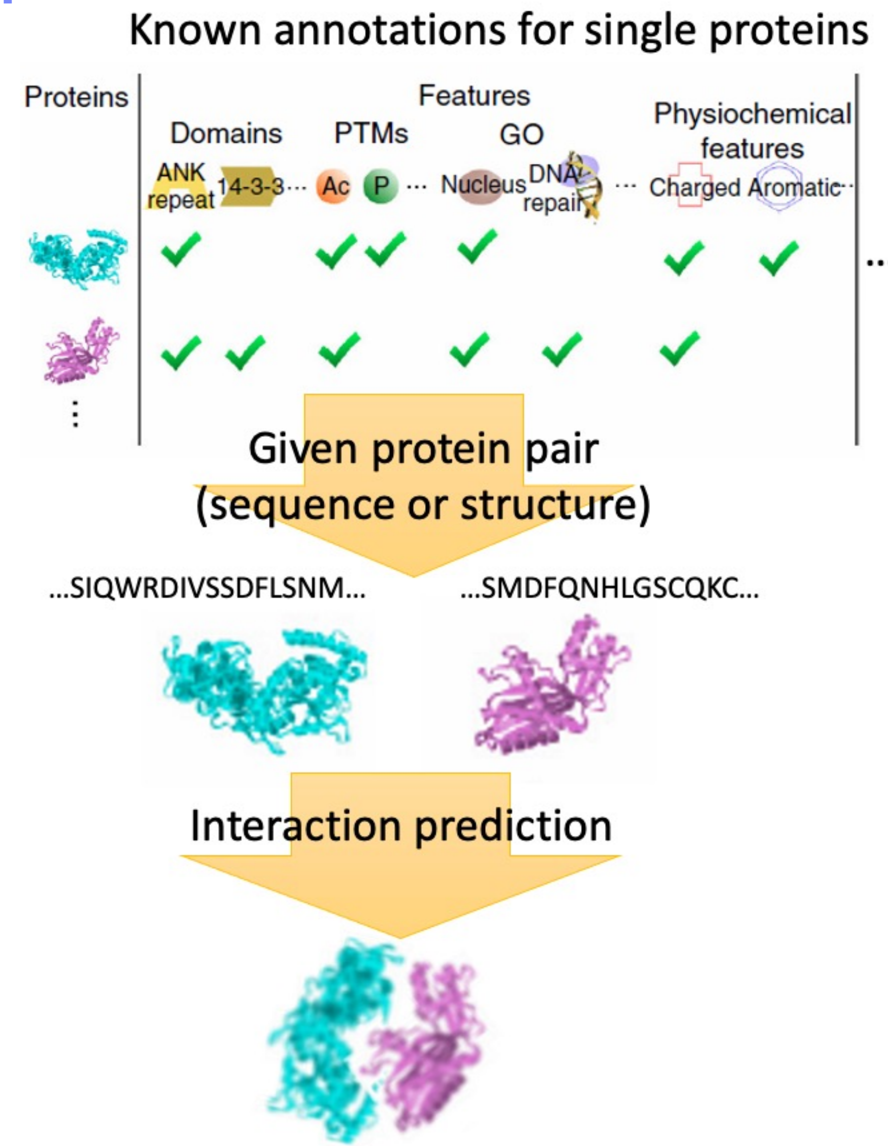


- **Goal**
  - Use open source data to predict the molecular can be produce or not, and try to simulate the synthesis processes by molecular properties.



C14. Protein Interaction Predictior

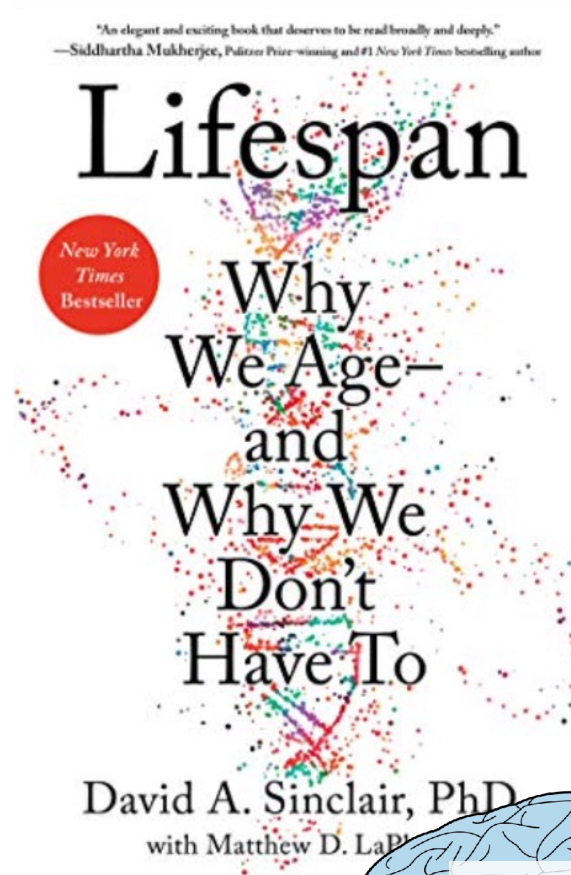
- **Background**
  - Protein-protein interactions (PPIs) are useful for understanding signaling cascades, predicting protein function, associating proteins with disease and fathoming drug mechanism of action.
  - Currently, only ~10% of human PPIs may be known, and about one-third of human proteins have no known interactions.

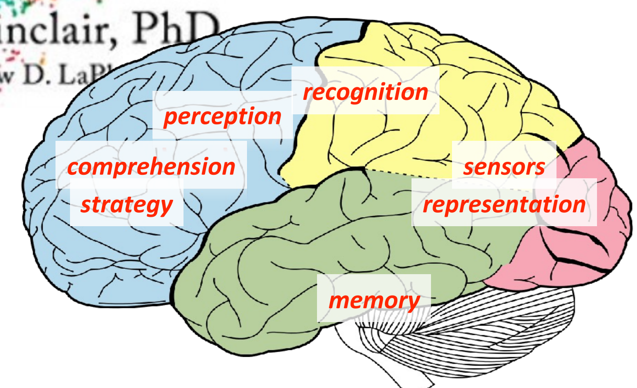


C15. AI Exploration and Understanding of Aging

- **Background**
  - Aging is a major impact on human. Recent studies have been giving more and more information on how agin functions and whether it’s possible to try to delay or even reverse some functions

- **Goal**
  - Study the mechanisms causing aging in the gene and protein level.
  - Use the protein structure prediction, protein-protein interaction, protein-drug biding tools to explore



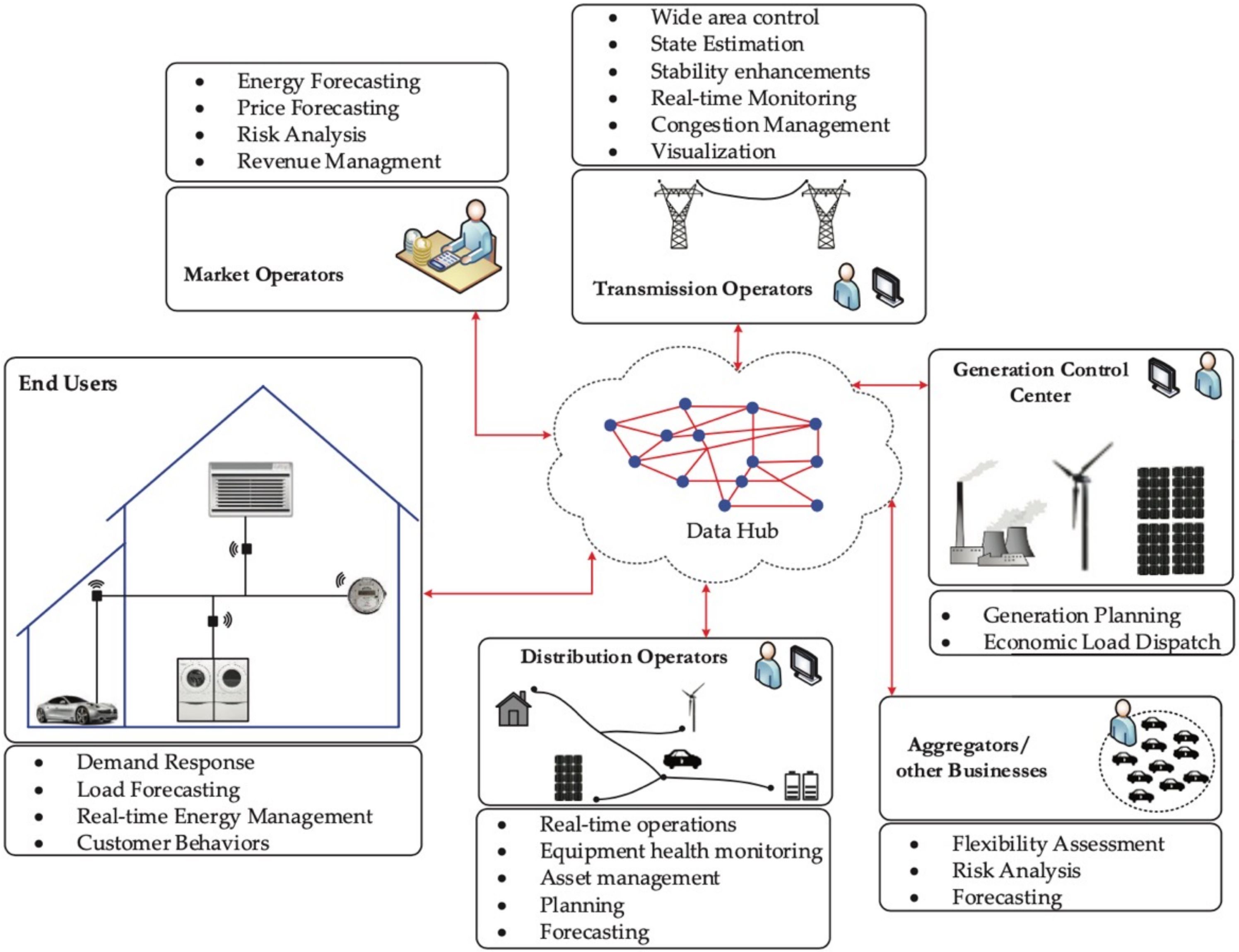


Task Area 4: Green Earth and Advanced Topics

Area 4 ‘Green Earth’ Tasks List:

D1: Distributed Solar Power Load Forecasting and Predictive Maintenance
D2: Distributed Wind Power Load Forecasting and Predictive Maintenance
D3: Power Flow Optimization
D4: Smart Grid Pricing Strategy
D5: AI for Novel Nuclear Fusion Power
D6: Stimulating Crop Growth
D7: Electronic Car Sensing and Predictive Maintenance
D8: Autonomous Driving
D9: Smart Cabin of Electrical Vehicles
D10: Social Policy Monitoring
D11: International Relationships and Policy Monitoring
D12: AI Chips – AI System on Chip
D13: AI Chips – Neural Processing Units
D14: Exploration in Immersive Environment
D15: Computer Vision Enhanced Immersive Environment

AI + Big Data Makes Smart Grid Possible

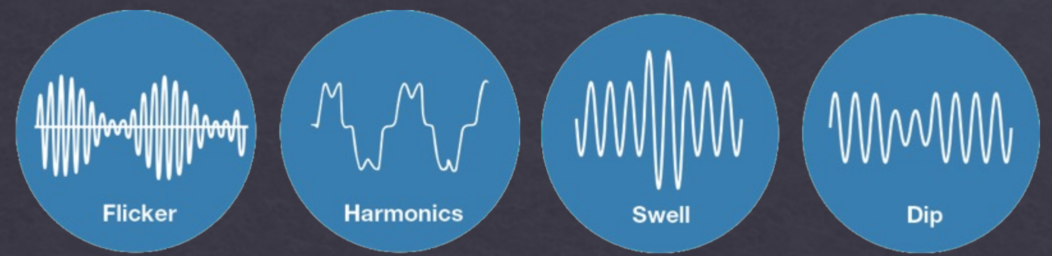


Key factors contribute to the complexity of the electric grid

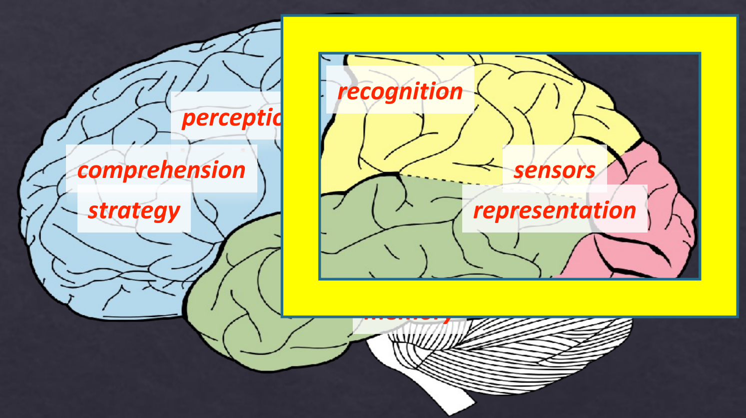
- ➢ The variability and intermittency of renewable generation.
- ➢ Decreased frequency response capability and decreasing system inertia.
- ➢ Changing load patterns and unpredictability.
- ➢ The need to manage vastly increasing number of endpoints.
- ➢ Growing cyber attack risks to the electric grid.

Department of Energy. Smart Grid System Report, Nov 2018

Power Quality Analysis



- ➢ Power system modeling
- ➢ Power system classifying
- ➢ Point of event occurrence into one unified frame
- ➢ Equipment sensitivity to the event disturbance
- ➢ Power quality events detection and characterizing
    - ➢ E.g. sag, swell, outage, harmonic, notch, flicker, impulse, etc.



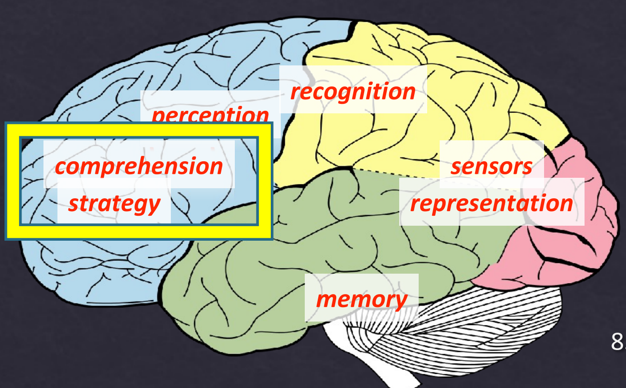
Voltage Waveform → Fourier and Wavelet-transform Based Feature Extraction → Fuzzy Neural Network Decision Making → Detection and Classification of a Disturbance

Detection and Classification → Characterization → Equipment Sensitivity Study → Event Location Prediction

Predictive Maintenance

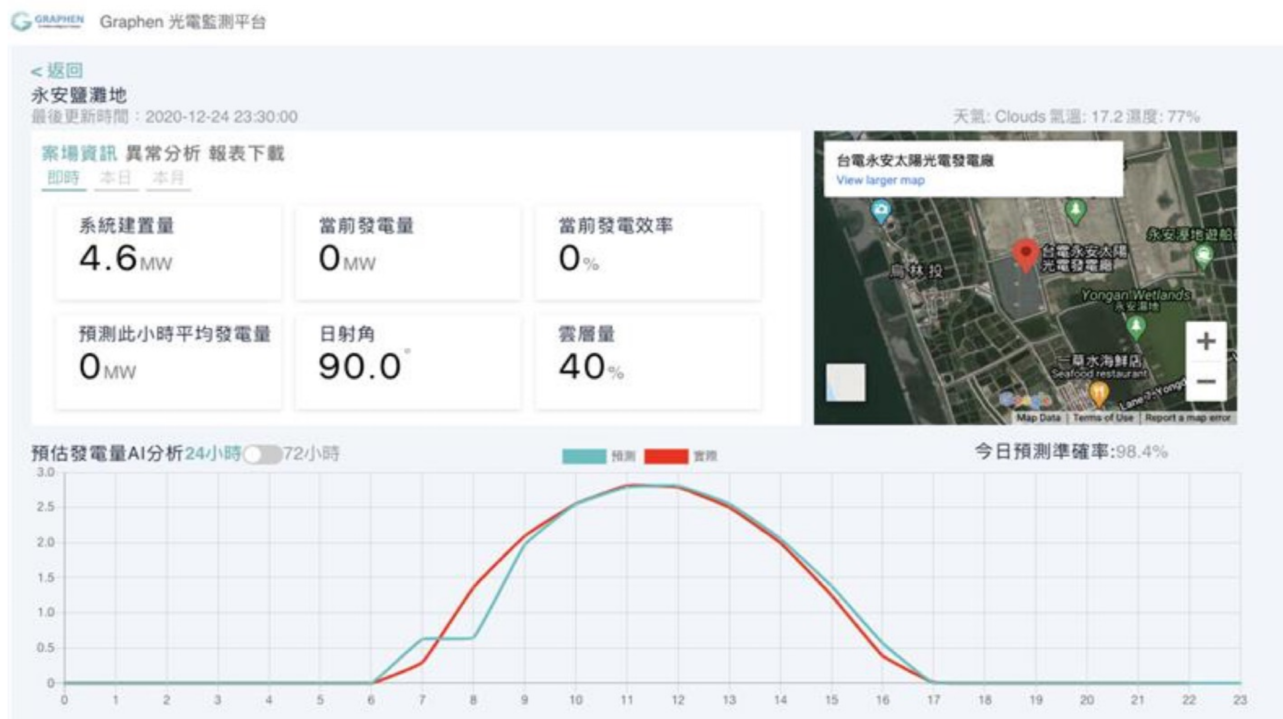
- Apply machine learning to historical power system data to reduce operating costs and failure risk
- Avoid or minimize the downtimes and reduce associated costs
- Optimize the periodic maintenance operations.
- Health indicator by machine learning
  - Classification – health indicator predicts what is the probability of failure in the future.
  - Regression approach – health indicator predicts how much time is left before the next failure.

Historical power system data
↓
Feature extraction
↓
Machine learning algorithms
↓
Failure Prediction
↓
Health indicator



D1. Distributed Solar Power Load Forecasting and Maintenance

- **Background**
  - Situation of Solar Power Plants varies and are time dependent. Power companies need good prediction on conditions and exclude anomalies in short time.
- **Task Goal**
  - Predict Solar Power Generations based on weather data
  - Anomaly Detection of Solar Power Plants
  - Predictive Maintenance of solar power plants.



D2. Distributed Wind Power Load Forecasting and Maintenance

- **Background**
  - Situation of Wind Power Plants varies and are time dependent. Power companies need good prediction on conditions and exclude anomalies in short time.

- **Task Goal**
  - Predict WindPower Generations based on weather data
  - Anomaly Detection of Wind Power Plants
  - Predictive Maintenance of solar power plants.

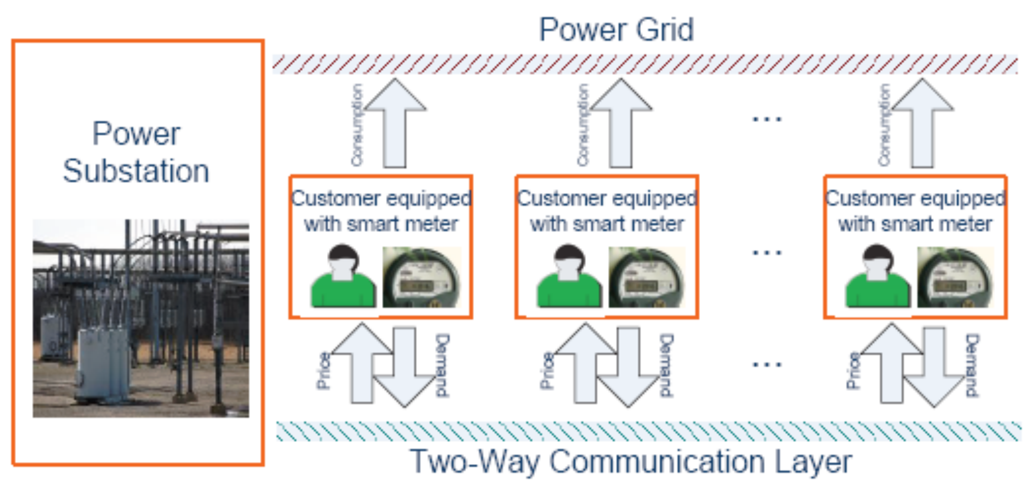
D3. Power Flow Optimization

- **Background**
  - Transmission is key to a Low-Cost Decarbonized US Grid
- **Task Goal**
  - Study the optimal straggles for power flow
  - Simulate various scenarios

https://www.greentechmedia.com/articles/read/study-transmission-is-the-key-to-a-low-cost-decarbonized-u.s-grid

D4. Smart Grid Pricing Strategy

- **Background**
  - Pricing strategy can be a way to optimize consumer behavior
  - After more and more cars and IoT devices rely on power, it's critical to influence customer behavior to optimize use of power grid
- **Task Goal**
  - Implement methodologies that can help change customer behavior
  - Game theory is a possible solution.
  - Other solutions should be also considered.



Chen et al. A cheat-proof game theoretic demand response scheme for smart grids, IEEE ICC 2012

D5. AI for Novel Nuclear Fusion Power

- **Background**
  - Desktop-size Nuclear Fusion is becoming a reality

- **Task Goal**
  - Studying and applying AI technology to advance novel desktop-size Nuclear Fusion Power

https://alpharing.com

Energy of the Future

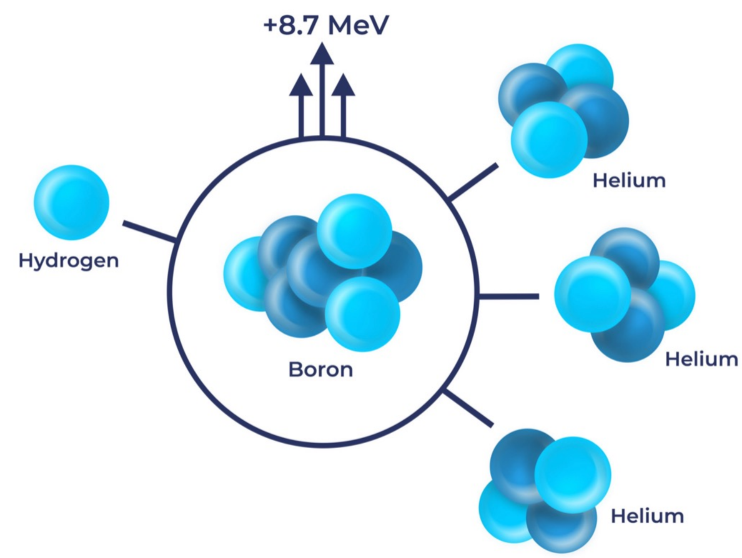
Nuclear fusion is the process where two light nuclei merge to form a single heavier nucleus.

It is the most efficient and clean way to produce energy known to humanity. It is also the process the stars, including our Sun, use to create energy. It is ten times more fuel-efficient than nuclear fission, promising to solve our energy needs once and for all.

We are working on making aneutronic Proton-Boron fusion a reality here, today.

Our approach

Get in touch

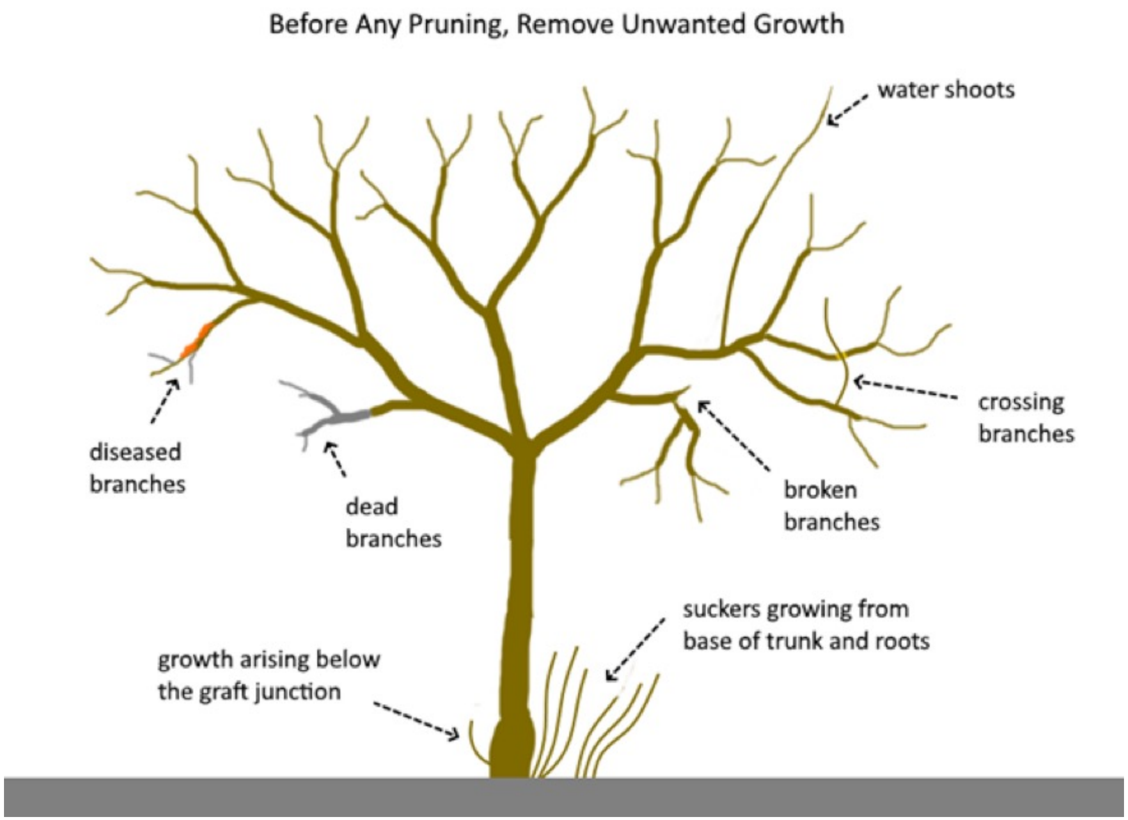


Fusion: two light nuclei merge to form a single, heavier nucleus.

D6. Stimulating Crop Growth

- **Background**
  - Machine learning in image recognition crop growth status and crop management strategy

- **Task Goal**
  - Establish the ideal growth model of crops
  - By using the image recognition find the crop growth status. Mark the unwanted growth. Give suggestion of the location to do the pruning.
  - Using climate and soil data to give suggestion for irrigation and fertilization

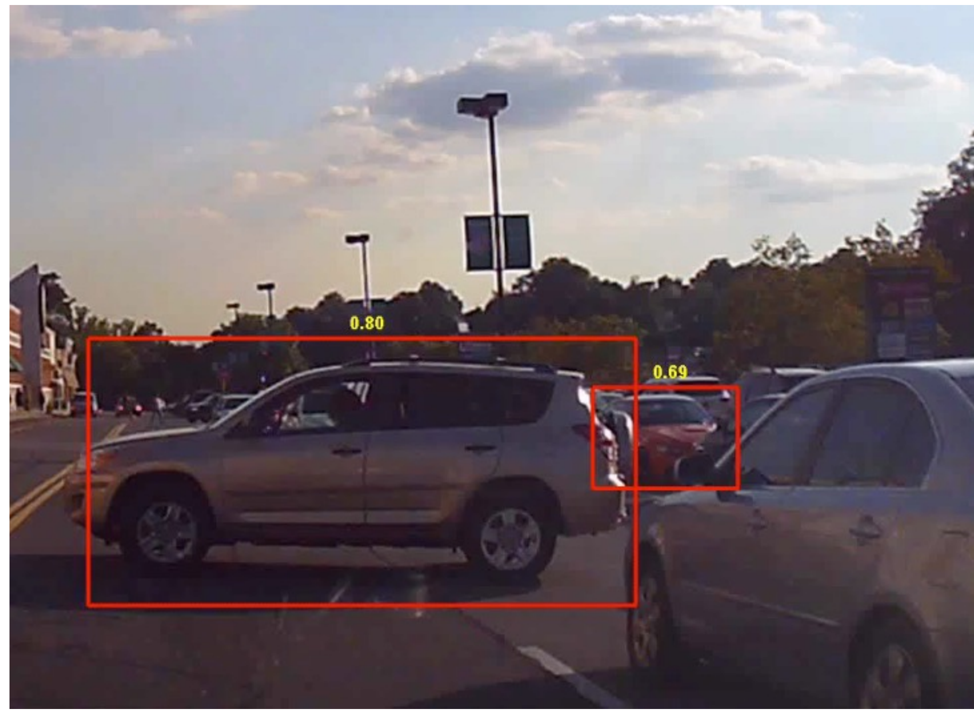


D7. Electronic Car Predictive Maintenance

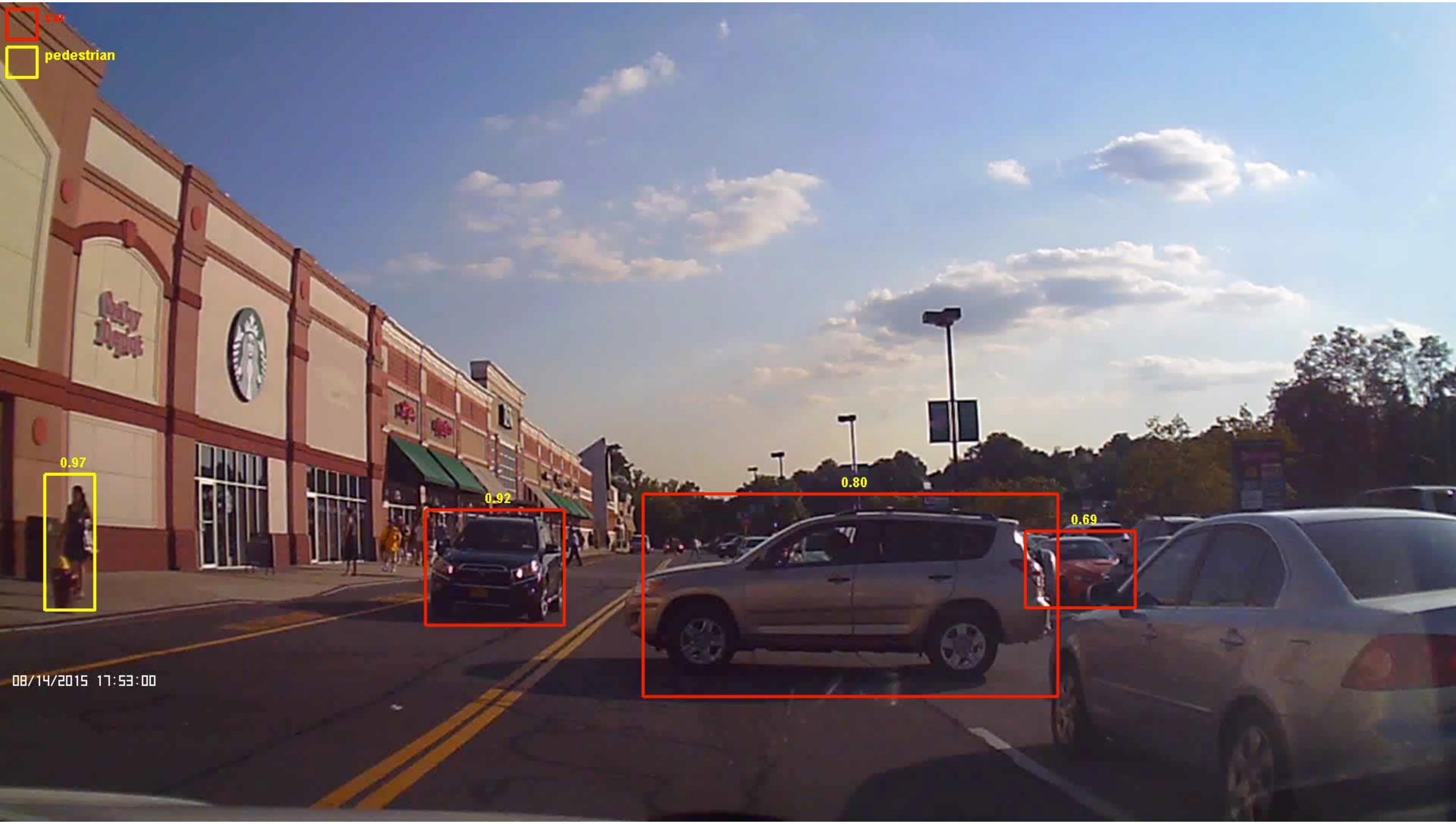
- **Background**
  - Car Fixing and Predictive Maintenance are important issues in the automobile industry
  - Pure electronic car is relative new
- **Task Goal**
  - Model Knowledge Graphs of the functioning of subsystems in an electronic car
  - Studying the sensors available in novel cars
  - Detection Car Problems from sensors
  - Prediction of maintenance requirements based on sensor signals
  - Incorporating other information such as environmental and demographical patterns into consideration.

D8. Autonomous Driving

- **Background**
  - Autonomous Driving is becoming mature
  - Autonomous Driving has to consider the complex situations in the road.
- **Task Goal**
  - Explore and experiment on autonomous driving technologies
  - Utilizing sensors to come up with optimal strategies to drive the car.
  - Build a Game Theory and Bayesian Network model to consider the complex behaviors on the road.



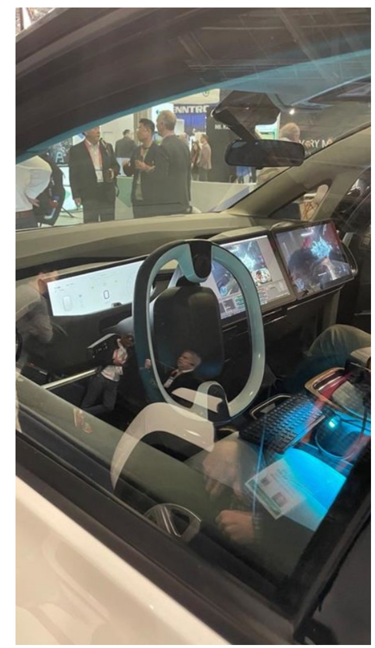
Example: Mobile Cognition in complex scenario



D9. Smart Cabin of Electrical Vehicles

- **Background**
  - Cars are being connected with all kinds of systems in a city
  - Novel applications in car based on digital human platform.

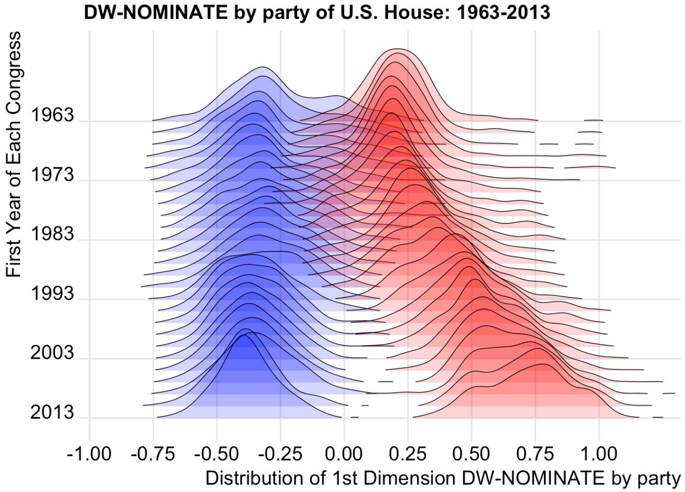
- **Task Goal**
  - Exploring novel car driving experience via Digital Human

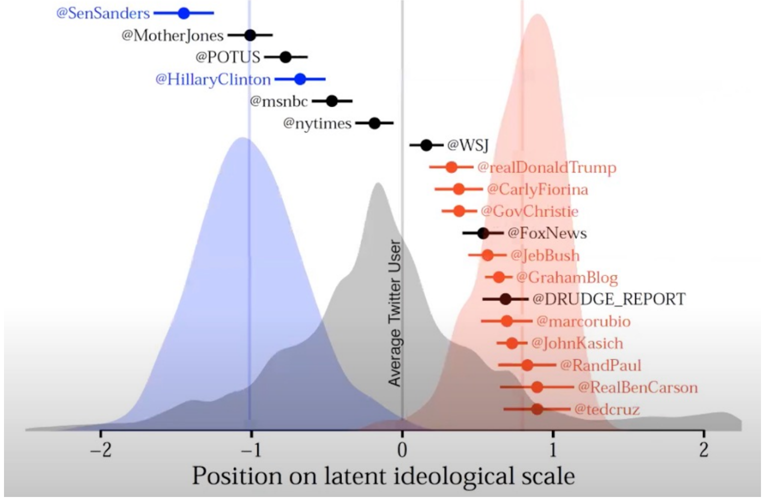


D10. Social Policy Monitoring

- **Background**
  - Social Issue and Policies have been impacting people's life

- **Task Goal**
  - Information Mining from Social Media to analyze the impact of social policy.
  - Analyze the effectiveness of policy making

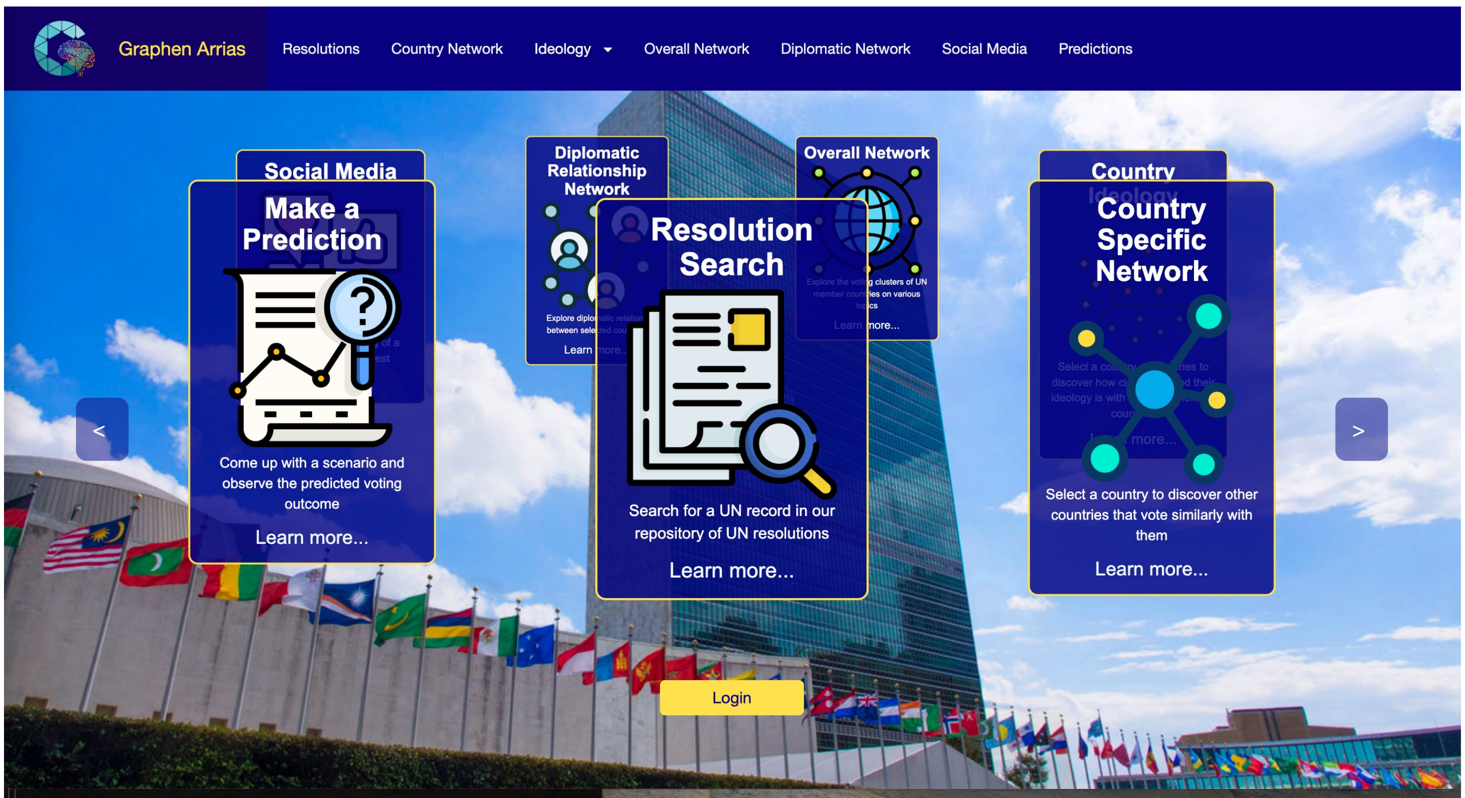


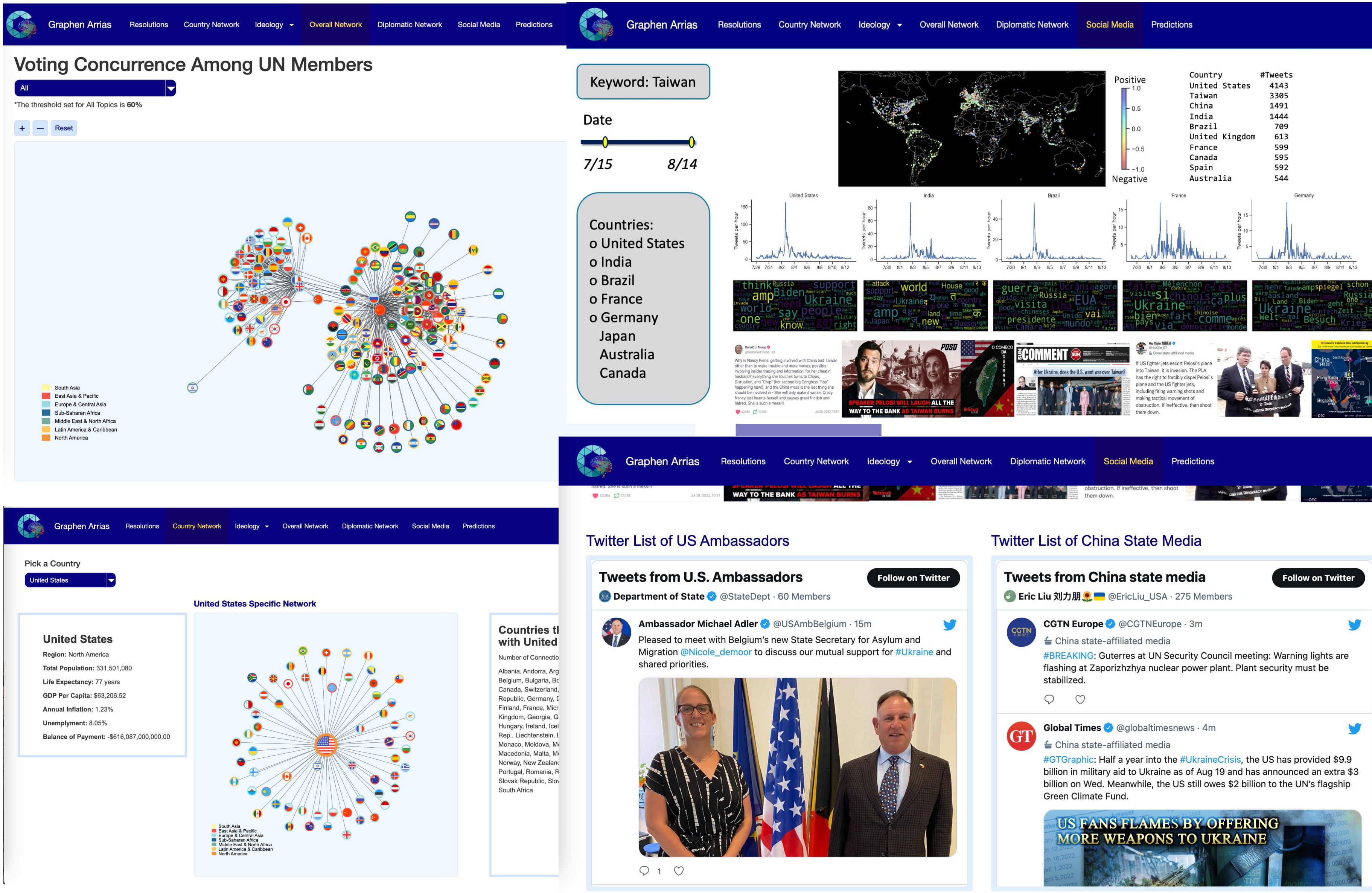


D11. International Relations and Policy Monitoring

- **Background**
  - Relationships between countries have been a major issue toward world policy changes

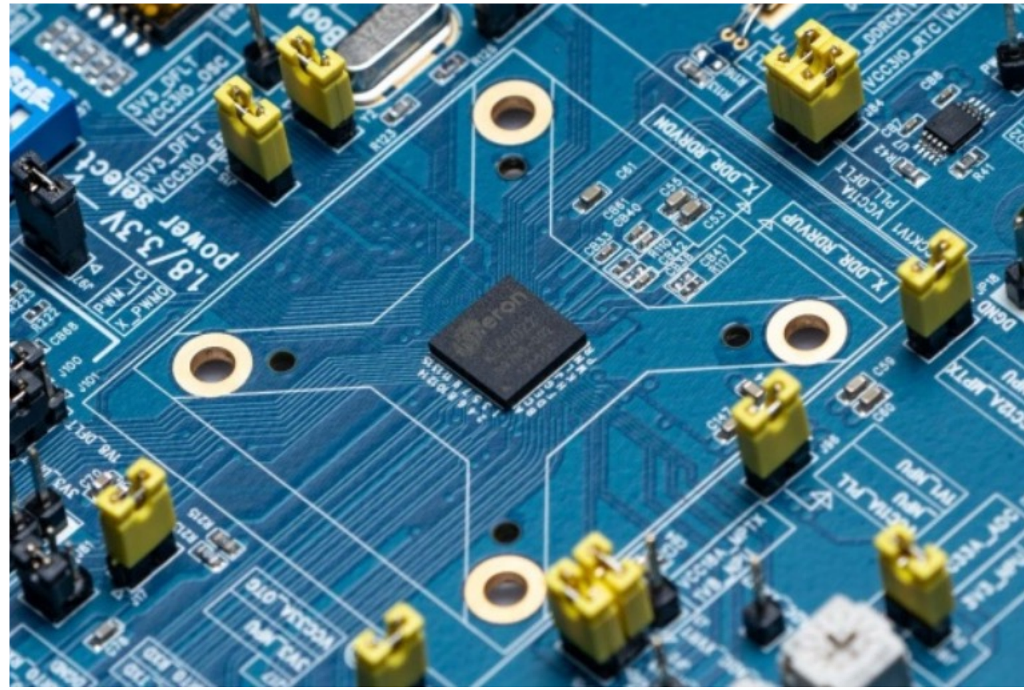
- **Task Goal**
  - Large-Scale Data Mining of international relationship evolutions
  - Visualize and create early alert of relationship changes




Graphen Arrias
Resolutions
Country Network
Ideology
Overall Network
Diplomatic Network
Social Media
Predictions
Voting Concurrence Among UN Members
All
*The threshold set for All Topics is 60%
Reset
South Asia
East Asia & Pacific
Europe & Central Asia
Sub-Saharan Africa
Middle East & North Africa
Latin America & Caribbean
North America
Keyword: Taiwan
Date
7/15
8/14
Countries:
o United States
o India
o Brazil
o France
o Germany
Japan
Australia
Canada
Positive
Negative
Country #Tweets
United States 4143
Taiwan 3305
China 1491
India 1444
Brazil 709
United Kingdom 613
France 599
Canada 595
Spain 592
Australia 544
United States
India
Brazil
France
Germany
Tweets per hour
Pick a Country
United States
United States Specific Network
United States
Region: North America
Total Population: 331,501,080
Life Expectancy: 77 years
GDP Per Capita: $63,206.52
Annual Inflation: 1.23%
Unemployment: 8.05%
Balance of Payment: -$616,087,000,000.00
Twitter List of US Ambassadors
Tweets from U.S. Ambassadors
Follow on Twitter
Department of State @StateDept · 60 Members
Ambassador Michael Adler @USAmbBelgium · 15m
Pleased to meet with Belgium's new State Secretary for Asylum and Migration @Nicole_demoor to discuss our mutual support for #Ukraine and shared priorities.
Twitter List of China State Media
Tweets from China state media
Follow on Twitter
Eric Liu 刘力朋 @EricLiu_USA · 275 Members
CGTN Europe @CGTNEurope · 3m
China state-affiliated media
#BREAKING: Guterres at UN Security Council meeting: Warning lights are flashing at Zaporizhzhya nuclear power plant. Plant security must be stabilized.
Global Times @globaltimesnews · 4m
China state-affiliated media
#GTGraphic: Half a year into the #UkraineCrisis, the US has provided $9.9 billion in military aid to Ukraine as of Aug 19 and has announced an extra $3 billion on Wed. Meanwhile, the US still owes $2 billion to the UN's flagship Green Climate Fund.
US FANS FLAMES BY OFFERING MORE WEAPONS TO UKRAINE

D12. AI Chips – AI System on Chip

- **Background**
  - Hardware AI Chips design is getting more and more popular
- **Task Goal**
  - Explore the functions and roadmaps of Edge AI Chips



D13. AI Chips -- Neural Processing Units

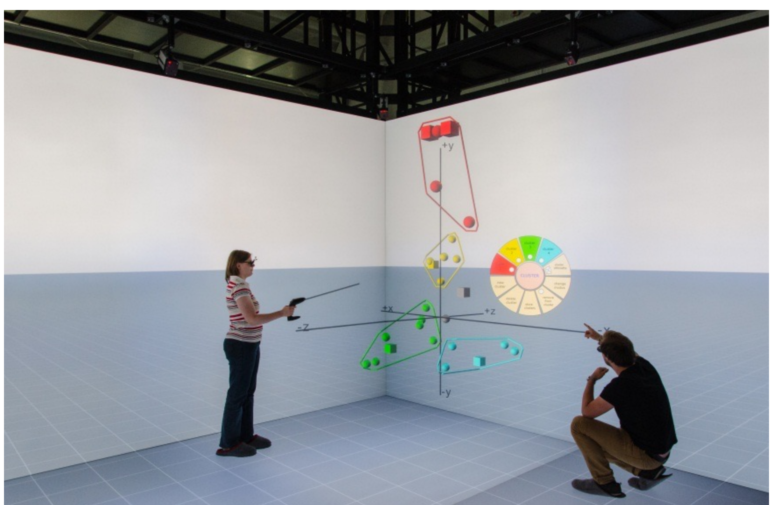
- **Background**
  - Hardware AI Chips design is getting more and more popular

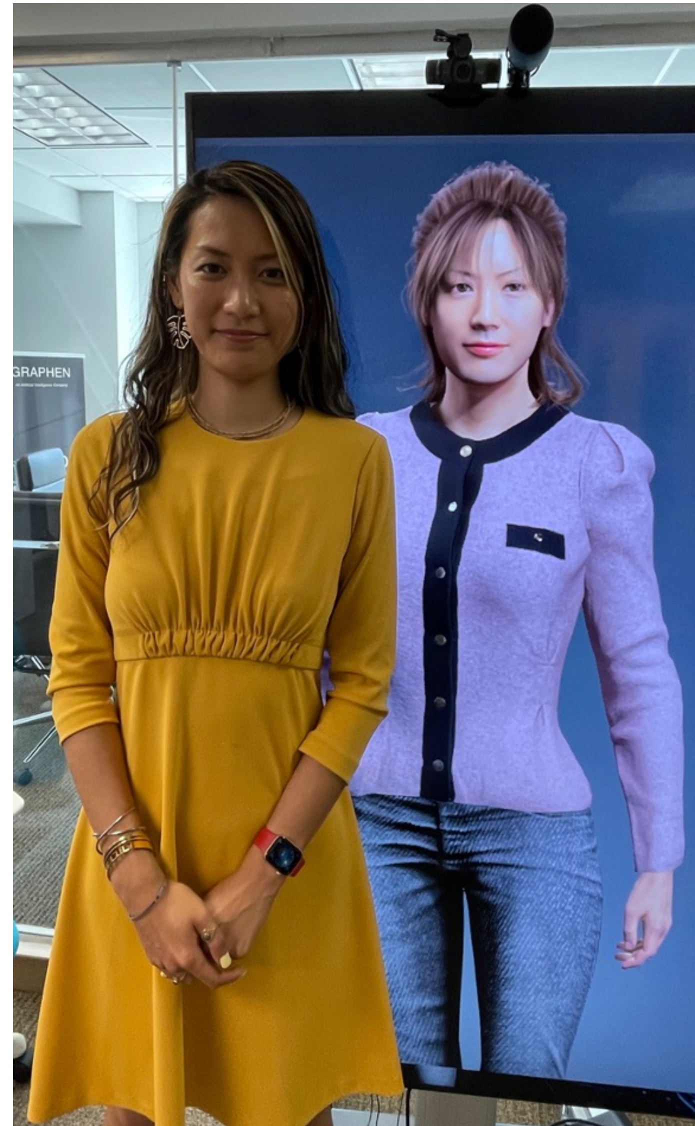
- **Task Goal**
  - Explore the functions and roadmaps of AI Chips based on Neural Processing Units

D14. Exploration in Immersive Environment

- **Background**
  - Augment Reality is now becoming more popular and more and more devices have been available in the market.
  - So far, less research and few systems are available for exploring networks in such condition.

- **Task Goal**
  - A team will design and implement augment reality applications based on Google Lens or Graphen Space.





D15. Computer Vision Enhanced Immersive Environment

- **Background**
  - Augment Reality is now becoming more popular and more and more devices have been available in the market.
  - Computer Vision techniques, such as objection recognition, can further enhance the intelligence and improve the capability of what can be achieved.
- **Project Goal**
  - A team will design and implement an augment reality application based on HoloLens or GoogleLens.
  - Some computer vision techniques will be implemented, such as object recognition and OCR.
  - The team encouraged to bring out any interesting usage scenarios on how these techniques can seamlessly enhance user experience of HoloLens or Google Lens.