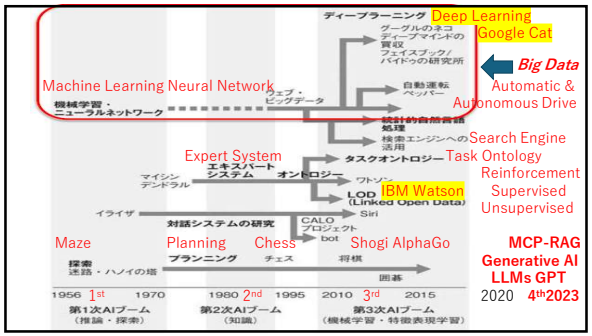
The evolution of AI and its application to predictive analytics and big data analysis for business

Emeritus Prof. Saga University
Kohei Arai

Content

- Definition of AI
- AI Evolutional History
- Current Status
- AI Applications
- Predictive Analytics
- Big Data Analysis for Business
- Conclusion

Reinforcement learning: Swing width→Reward



The lineage of AI evolution

- [First Generation: Rule-Based AI] (1950s-1980s)Symbolic processing and coding of specialized knowledge (Expert Systems)"Humans teach all the rules"
- [Second Generation: Statistical and Machine Learning AI] (1990s-2000s)Pattern recognition from data (SVM, early neural networks)"Humans define features, and AI learns statistically"
- [Third Generation: Deep Learning and Recognition AI] (2010s)Advanced image and sound recognition (CNN, RNN)"**Deep learning**: AI discovers features on its own and achieves recognition accuracy beyond that of humans"

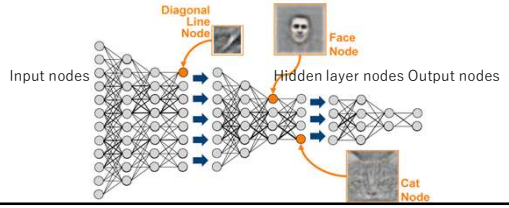
- [Fourth Generation: **Generative AI and Large-Scale Language Models (LLM)**] (Early 2020s)Understanding meaning and generating text and images (Transformer, multilingual and multimodal)Genetic Engineering: Learning vast amounts of knowledge on the web and thinking and outputting like a human **Vibe Coding, Agentic Engineering**
- [Fifth Generation: **Physical AI (Physical Experiential Learning)** (Present to Future)Acquisition of Embodiment and Interaction with the Physical World (Robotics, Autonomous Driving, Spatial Intelligence)"A digital brain possesses a physical 'body' and perceives and operates in the real world."
- AGI⇒ASI

Vibe Coding, Vibe Design, Multi-Agent Pipeline, Agentic Engineering

- **Vibe Coding Tools**
  - **Cursor**
  - **Windsurf**
  - Lovable.dev
  - Bolt.new
  - Replit + AI Agent
  - Tempo Labs
  - **Cline**
  - Continue
  - Devin
  - **Claude Code**
  - **CoPilot**
  - **Codex**
- Multi-Agent Tools
  - **CrewAI**: strong for role-based multi-agent collaboration and easy-to-read workflows.
  - **AutoGen**: good for conversational agent coordination and flexible agent-to-agent patterns.
  - **LangGraph**: best when you want explicit control over stateful agent graphs and routing.
  - **n8n**: useful if you want agent-style automation integrated with business workflows and many connectors.
  - **Flowise**: helpful for visual building of LLM apps and agent flows with low-code setup.

Deep Learning: Google cat

- Apply deep learning to 10 million images extracted from YouTube: 3 days of learning on 1000 servers → Neural network reacts to cat faces, same goes for human faces



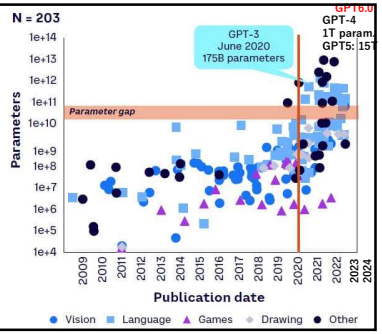
AI Fourth Era

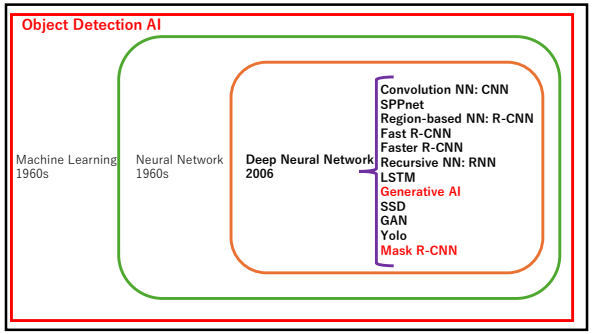
- 2022~: The era of large-scale models
- Introducing ChatGPT (3.0): **175 billion parameters** (number that refers to data input from outside) → By applying parameters to calculation formulas and learning them, it is now possible to derive **more than 40 trillion pieces of knowledge → This is 1000 times more knowledge than a person can gain in a lifetime.**
- **ChatGPT (4.0) expands to over 1 trillion parameters (6000 times) GPT-5: 15 trillion = 90,000 times**

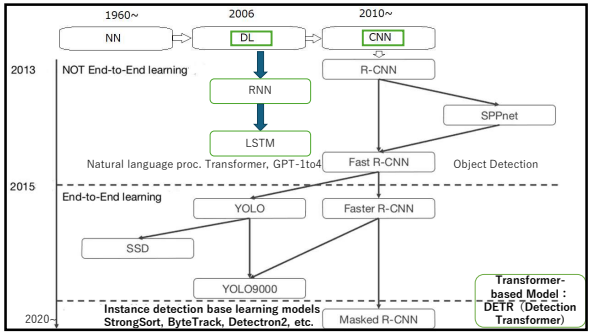
Year	2018	2019	2020	2022	2023	2025	2026
GPT ver.	1	2	3	3.5	4	5	5.5⇒6.0
No.Parameter	117M	1.5B	175B	355B	1.5T	15T	?
Learning Data	4.5GB	40GB	570GB	?TB	>5TB	>50TB	?

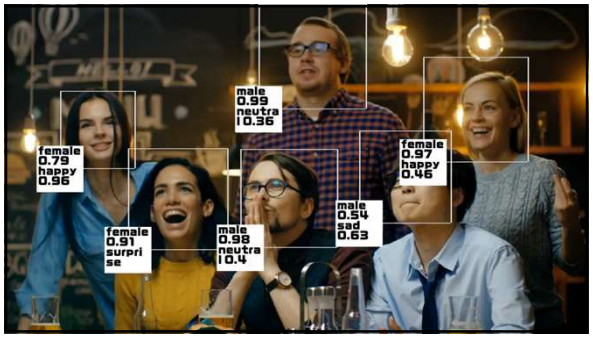
GPT-4 and 5

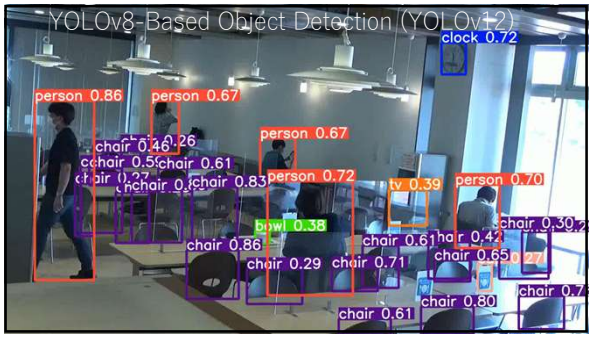
- Growth in parameters in ML models, 2009-2022
- GPT-4 has 6 times many parameters than GPT-3
- Memory capacity of human brain knowledge is 1.75zB
- GPT-4 has 6000 times many knowledge than human
- GPT-5: 90,000 times
- 5.5->6.0





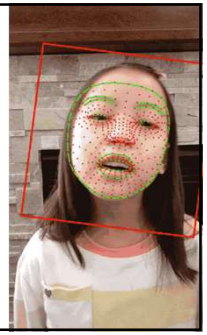






https://nemutas.github.io/app-mediapipe-facemesh-demo/

- 468 of landmarks are selected for **face authentication**
- Using image similarity is calculated then the authentication using MediaPipe Face Mesh



Visualize emotion

Startup company Nuro (Waymo under Alphabet) launches AI self-driving system

- Started Kroger product delivery in 2018
- Delivery of Domino's Pizza begins on public roads in select areas of Huston

Walmart stores using AI

- Opened Intelligent Retail Lab: IRL with inventory management using AI from April 2019
- Unlike Amazon Go, AI specializing in **inventory management: Manage product inventory in grams**

McDonald's uses AI for drive-thru

- Acquisition of Dynamic Yield (provision of product information tailored to individual customer preferences, Apprente (**AI voice recognition**))
- Provide product and customer experience recommendations based on **time of day, weather, current restaurant traffic and trending menu items**

Moët Hennessy Louis Vuitton Starts Up

- 3DLook (scanning the human body with a smartphone), Entrupy (checking the authenticity of luxury handbags with a smartphone)
- Learn the surface micro image database of millions of materials and perform material analysis
- It can distinguish between real and fake

Introduction to Big Data

- Why do we tend to click on ads on search pages and SNS?
- Or will there be advertisements for things I want?
- **Online shops use customer purchase data, member information, and website information.**
- We store various data such as click information within the website as big data. Google Analytics
- **Using that big data, artificial intelligence can help customers or agents**
- **Decide which products are likely to be purchased and display them as recommended products to customers. This will lead to sales.**

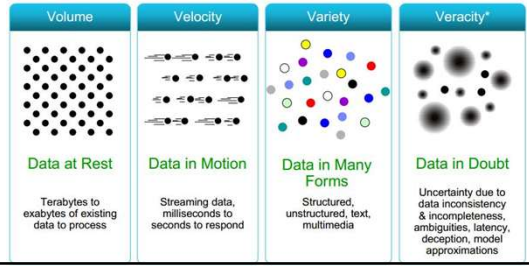
Just an example of Big Data

- Search history; Scraping
  - looking at car
  - looked for test values
  - looked for a campsite

Big Data

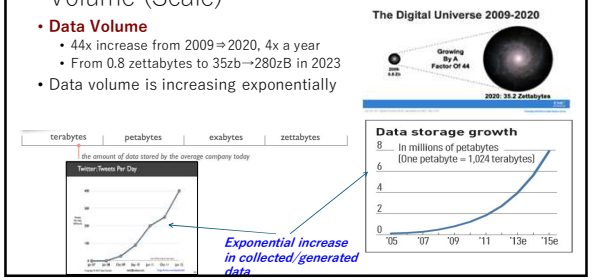
- As a result of the identification prediction, **you will be directed to the desired site and click on it by Search Engins or agents**

Some Make it 4V's ⇒ Definition of Big Data



Volume (Scale)

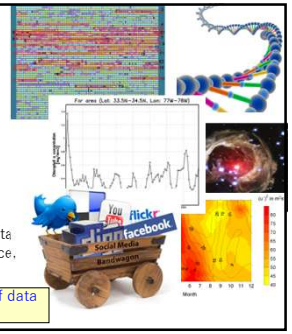
- **Data Volume**
  - 44x increase from 2009⇒2020, 4x a year
  - From 0.8 zettabytes to 35zb→280zB in 2023
- Data volume is increasing exponentially



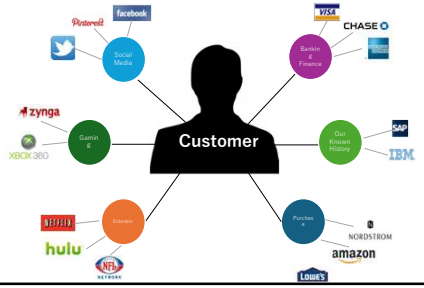
Variety (Complexity)

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
  - Social Network, Semantic Web (RDF), …
- Streaming Data
  - You can only scan the data once
- A single application can be generating/collecting many types of data
- Big Public Data (online, weather, finance, etc.)

To extract knowledge➔ all these types of data need to linked together



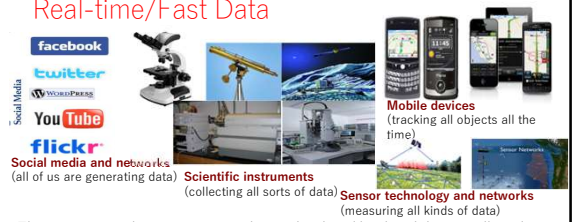
A Single View to the Customer



Velocity (Speed)-Real time analysis(Decision making)

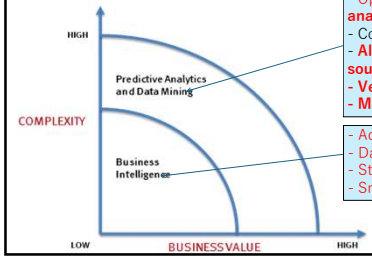
- Data is begin generated fast and need to be processed fast
- **Online Data Analytics**
- **Late decisions ➔ missing opportunities**
- **Examples**
  - **E-Promotions:** Based on your current location, your purchase history, what you like ➔ send promotions right now for store next to you
  - **Healthcare monitoring:** sensors monitoring your activities and body ➔ any abnormal measurements require immediate reaction
  - **Traffic flow monitoring,** etc.

Real-time/Fast Data



- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in **a timely manner and in a scalable fashion**

What's driving Big Data



- Optimizations and **predictive analytics**
- Complex statistical analysis
- **All types of data, and many sources**
- **Very large datasets**
- **More of a real-time**

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

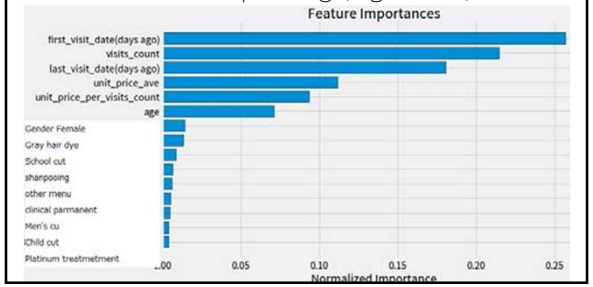
Generative AI

- University of Chicago compares GPT-4's accuracy in financial analysis
- Content: A research team at the University of Chicago Booth School of Business found that OpenAI's GPT-4 model outperformed human financial analysts in the task of analyzing corporate financial statements and predicting future changes in performance.
- Results: **GPT-4 predicted performance changes with approximately 60% accuracy, while human analysts were only 53-57% accurate**. In particular, it was emphasized that AI demonstrated superiority in cases where analysts were weak.

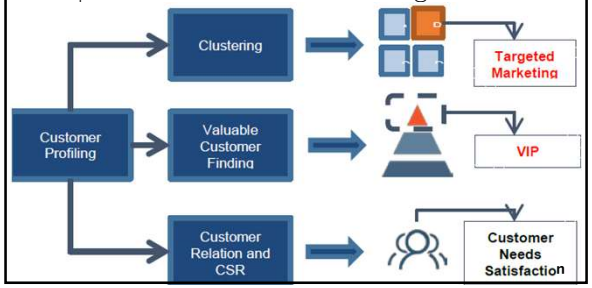
Some rules say customer profiling and churn prediction are important

- The 2:8 rule states that 80% of sales come from the top 20% of customers (fans and heavy users).
- The Pareto law states that "80% of sales are made by the top 20% of fans and heavy users"
- The 5:25 rule states that "preventing just 5% of customer defections will greatly increase profits by 25%"
- "Depending on the industry, if you can reduce customer churn by 5%, profits will increase by 25-85%."

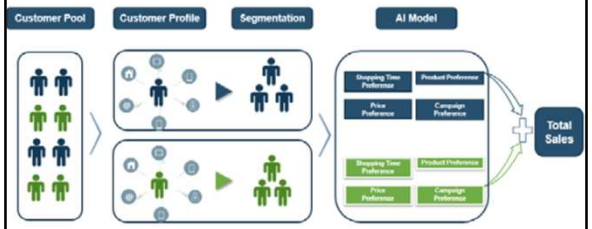
Churn customer profiling (LightGBM)



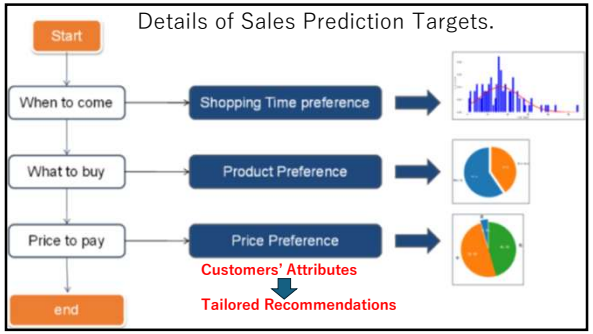
Importance of Customer Profiling.

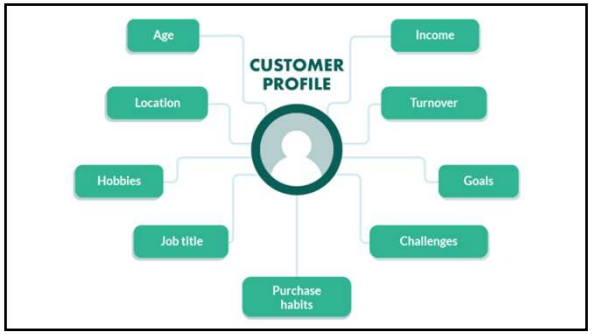


Process Flow of Sales Prediction.

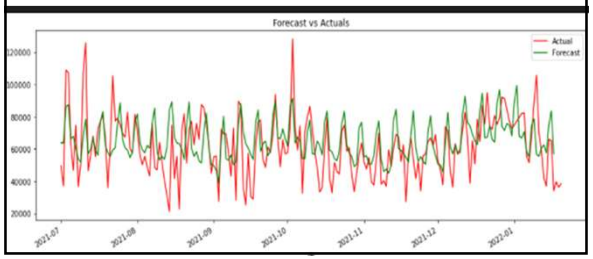


Details of Sales Prediction Targets.





Prediction of the sales in a target year with the past 10 years data based on Prophet+Optuna



Risks of AI

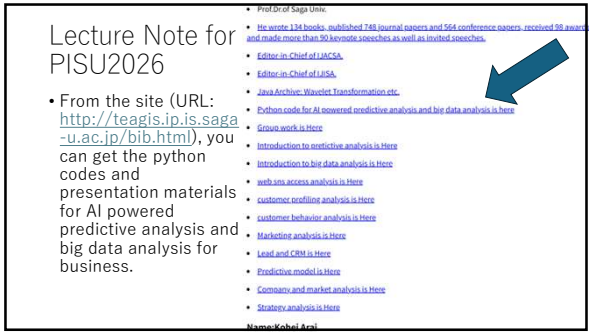
- Vulnerability: Garbage in, Garbage out.←Importance of big data analysis in particular quality of the training datasets
  - Open AI has developed a method that can easily upset AI image recognition by adding noise data such as mosaics to the original image data.
  - "Dolphin Attack" that allows you to operate the voice assistant such as Siri, Google at frequencies that are inaudible to humans
- Blackbox: Unknown logics (How can we rely on AI?) →accountability problem
- Bias Problem: Human prejudice is learned by AI

OECD 8 principles for using AI

- Transparency
- User support
- Controllability
- Secure security
- Security
- Privacy protection
- Ethics
- Accountability

Lecture Note for PISU2026

- From the site (URL: http://teagis.ip.is.saga-u.ac.jp/bib.html), you can get the python codes and presentation materials for AI powered predictive analysis and big data analysis for business.



Summary

- So many applications of AI: Small examples (respond to users' requirements) are presented here
- Current AI methods and systems are Week AI (Only specific job can be done)→Strong AI (AGI, ASI: Has human-like self-awareness and intelligence and makes advanced judgments) is coming soon (Singularity⇒Kurtwile: 2045, Hinton: 2030)
- Risks of the current AI: Vulnerability to noise, Blackbox and Bias problem, ELSI issues, Hallucination, etc.
- The OECD (38 member countries) reported 27% of the region's workforce is in jobs at high risk of being replaced by AI.
- In a **human-centered society**, humans will let AI do the work that AI can do, and AI will not take away jobs from humans.

AI for business: Sales, customer numbers, average customer spending, labor productivity improvement, production cost reduction, operational efficiency improvement

- Hydrant, a Wellness brand based in the US, has successfully leveraged Predictive Analytics to identify churn indicators and predict churn propensity with **83% accuracy**, while increasing conversion rates and average customer spend by **2.7x and 3.1x** respectively, when compared to control groups.
- The predictive model creates detailed forecasts for each **customer's possibility of churn**. With these accurate individual forecasts, Hydrant dynamically segments customers to receive **tailored marketing messages and discounts** that match their future buying power.