

# Will AI surpass human intelligence?

## -- Beyond Deep Learning

The University of Tokyo  
Yutaka Matsuo

# Yutaka Matsuo



- 1997      Graduated from The University of Tokyo
- 2002      Obtained Ph.D. degree (Ishizuka Lab) from The University of Tokyo  
Researcher at National Institute of Advanced Industrial Science  
and Technology (AIST)
- 2005年    Visiting Scholar at Stanford University  
Center for the Study of Language and Information (CSLI)  
Prof. Stanley Peters
- 2007年    Associate Professor at Graduate School of Engineering,  
The University of Tokyo
- 2014年    Director, Chair of Global Consumer Intelligence,  
The University of Tokyo

Majors in artificial intelligence, web engineering, and deep learning.

Served as Editor-in-chief of the Japan Society of Artificial Intelligence (JSAI) from 2012 to 2014, and as Director of ELSI committee in JSAI from 2014.

Program committee at WWW conference from 2007, and track chair of web mining at WWW conference in 2014 and 2016.

Visiting Associate Professor at National University of Singapore (NUS). Founder of Ohma Inc. and READYFOR Inc. Technical advisor of Money Forward, Industrial Growth Platform, Inc. (IGPI), and Plug AI Asset Management (PAAM).

# History of AI

- First AI boom (1956–1960s): Era of search and inference
  - Dartmouth Conference (1956)
    - The term “Artificial Intelligence” is coined
    - Shortly after the first computer, ENIAC (1946)
  - Theorem proving, chess playing
- ... winter years
- Second AI boom (1980s): Era of knowledge
  - Expert systems
  - Medical diagnosis and identification of organic compounds
  - Fifth-generation computer project: 50 billion yen funded by METI
- ... winter years
- Third AI boom (2013--): Era of machine learning and deep learning
  - Development of the web and big data
  - Development of computational power

Intelligence that can think fast

Knowledgeable intelligence

Intelligence that can learn

## 将棋電王戦



## IBM Watson



## Deep learning (2007-)

Historical win at ILSVRC (2012)  
Google cat (2012)  
Acquisition of DeepMind (2013)  
FB/Baidu AI institute (2013)

## Machine learning

Web and big data

Applications to  
Cars and robots Self-driving car  
Pepper

Statistical NLP  
(machine translation)  
Search engines

MYCIN (medical diagnosis)

DENDRAL

Expert  
systems

ontology

Task ontology

Watson (2011)

Linked Open Data (LOD)

Eliza

Dialogue systems

Calo project  
bot

Siri (2012)

search  
games

planning  
STRIPS

Chess (1997)  
Deep Blue

Shogi (2012-)  
Denou-sen

go

1956

1970

1980

1995

2010

2015

First AI boom  
(inference and search)

Second AI boom  
(knowledge representation)

Third AI boom  
(machine learning and deep learning)

# Bottleneck of past AIs $\doteq$ Bottleneck of feature extraction

- Hard problem 1: Feature engineering on machine learning
  - It was difficult to have good features when we use machine learning.
  - Humans must observe the target phenomenon and devise new features.
- Hard problem 2: Frame problem
  - We describe knowledge for robot planning.
  - There are many different exceptions. Common sense can not be described easily.
- Hard problem 3: Symbol grounding problem
  - Computers cannot “understand” that a zebra is a horse with stripes.
  - Symbols are not grounded and are therefore not manipulated (or added) correctly.

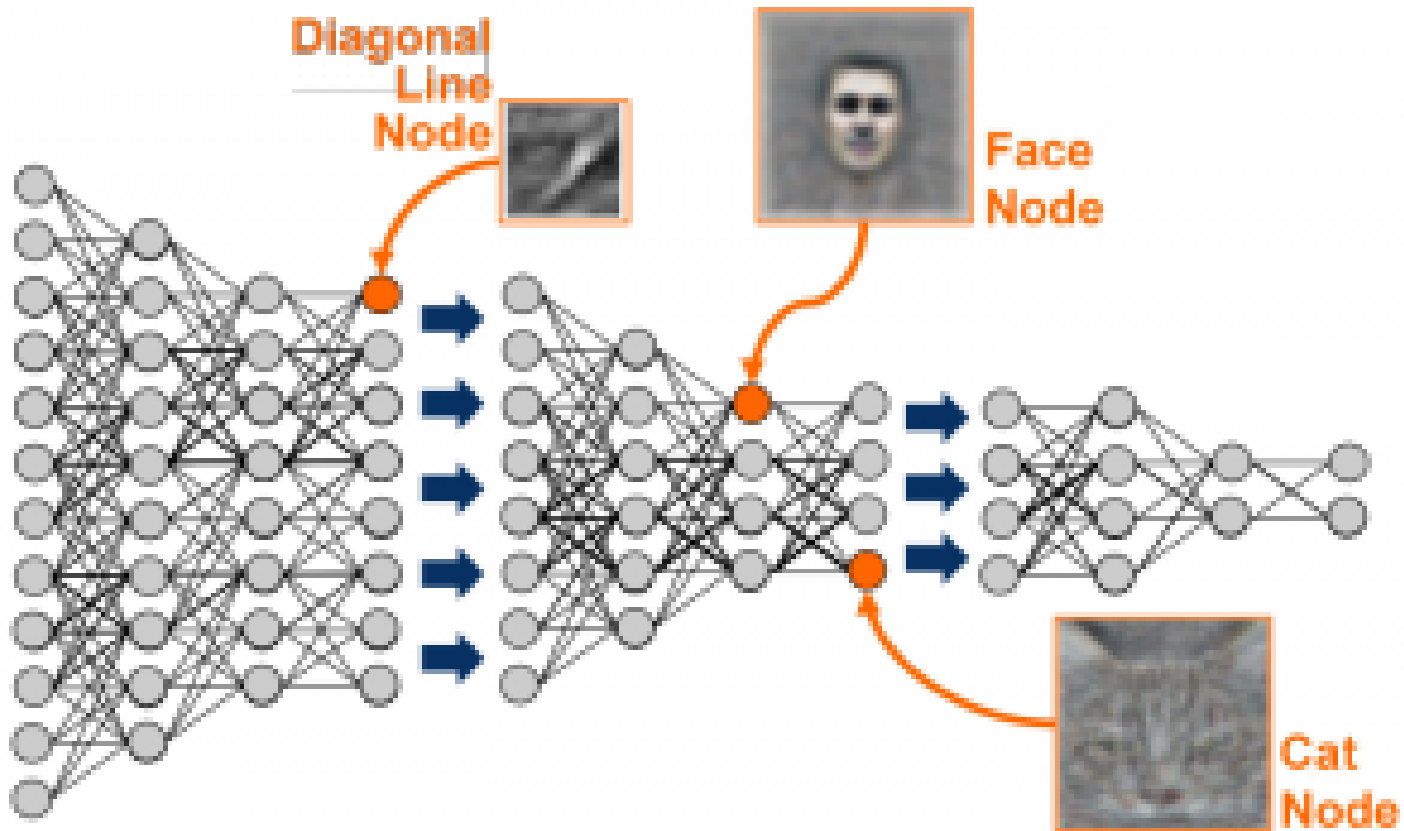


In past AI systems,

**humans must extract features by observing phenomena, and then build models to simulate the phenomena.**

After building the models, computation processes were automated. However, the very first step of modeling depends heavily on human efforts.

# Google cat (2012)



- By inputting numerous images obtained from YouTube, simple features are obtained in lower-layer neurons. More complex features are obtained in higher-layer neurons.
- It closely resembles human visual area visual cortex.

# Performance of deep learning (2012)

- Performance suddenly improved. Many researchers were shocked.
- ILSVRC2012: Large Scale Visual Recognition Challenge 2012

	Team name	Error	Description
Deep learning	SuperVision	15.315%	Using extra training data from ImageNet Fall 2011 release
	SuperVision	16.422%	Using only supplied training data
Conventional Feature engineering	ISI	26.602%	<b>Huge jump</b> of scores from classifiers using each FC
	ISI	26.646%	Naïve sum of scores from classifiers using each FV
	ISI	26.952%	Naïve sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV and CSIFT+FV, respectively
	OXFORD_VGG	26.979%	Mixed selection from High-Level SVM scores and Baseline Scores, decision is performed by looking at the validation performance.
	...	...	...

# Improvement of performance: from 2012 to present

		Error
Before Deep learning	Imagenet 2011 winner (not CNN)	25.7%
	Imagenet 2012 winner	16.4% (Krizhevsky et al.)
After Deep learning	Imagenet 2013 winner	11.7% (Zeiler/Clarifai)
	Imagenet 2014 winner	6.7% (GoogLeNet)
	Baidu Arxiv paper: 2015/1/3	6.0%
	<b>Human:</b> Andrej Karpathy	5.1%
	MS Research Arxiv paper: 2015/2/6	4.9%
	Google Arxiv paper: 2015/3/2	4.8%

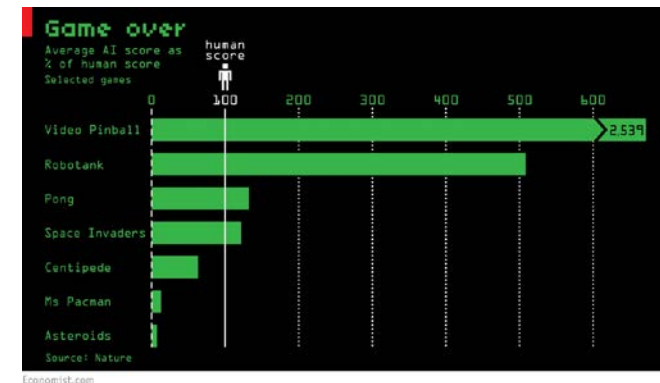
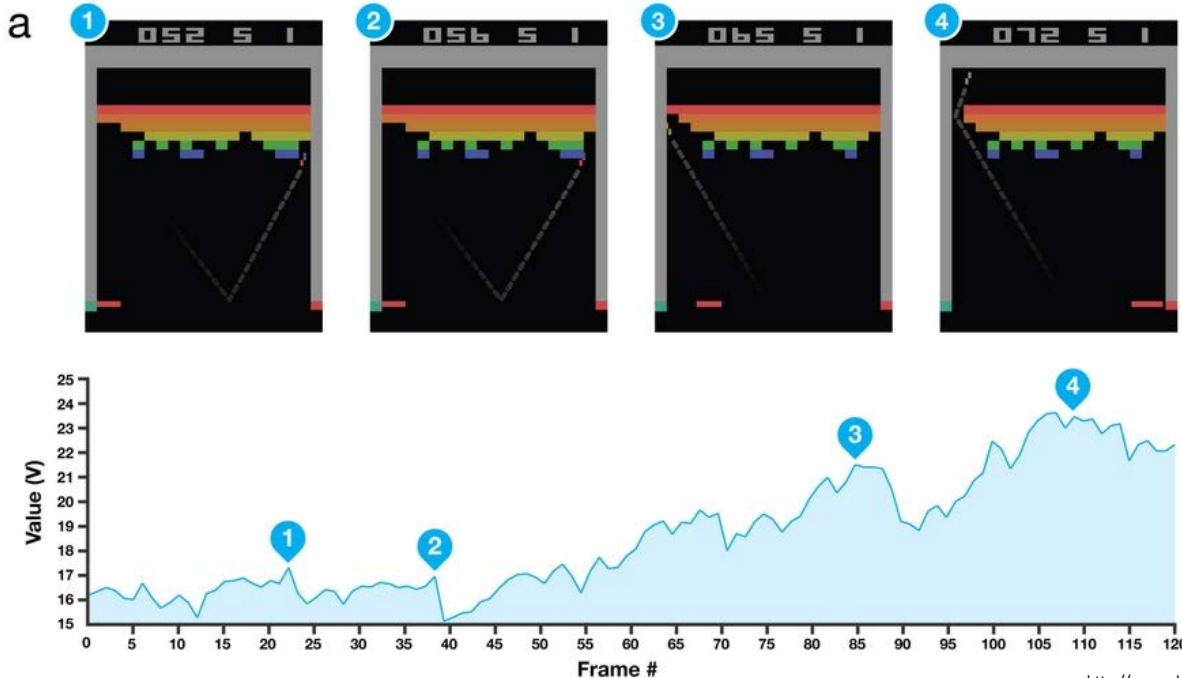
Surpassed human performance in Feb. 2015

No one imagined this only several years ago.



# Deep learning + reinforcement learning (2013)

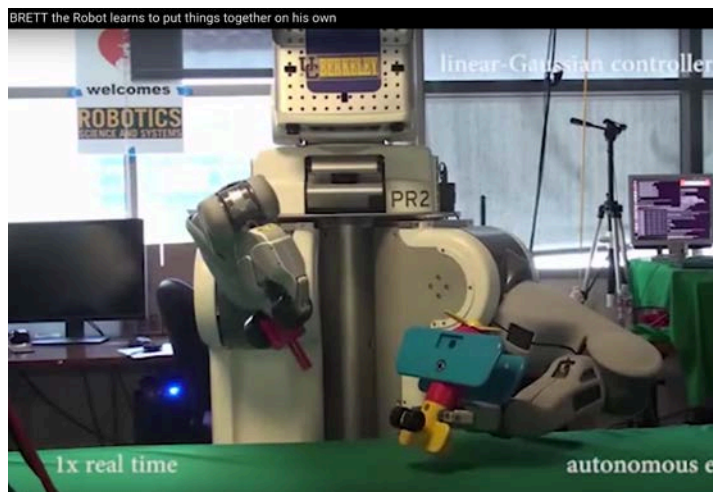
- Reinforcement learning enables computers to learn actions.
  - Rewards are obtained with time delay. The algorithm maximizes the expected rewards.
  - In Q-learning, a Q-value is assigned to the set of state and action.
  - Conventionally, the state is defined with variables set by humans.
- DeepMind researchers (D. Hassabis et al.) used features obtained by deep learning for state representation.
  - They applied the algorithm to play games in 2013: The score is set to be the reward. Google acquired them in 2014.
- Actions become proficient by trial and errors
  - Initially they make a bad move. Gradually they play better.
  - Finally they find clever ways to earn high scores.
  - The same algorithm can be applied to different games.
  - In more than half of ATARI games, computers score higher than the best human score.



# Deep learning + reinforcement learning: Real world (2015--)

- Application to real world
  - May 2015: Deep learning robots that learn actions (assembling a toy plane) by trial and error (UC Berkeley)
  - May 2015: Deep learning cars that learn driving by trial and error (PFI, Japan)
  - Other projects in Maryland U. and EU.
- It is not surprising because
  - Dogs and cats can do it. It does not require high intelligence such as language ability. Only (visual) feature extraction matters.
  - Historically many AI researchers advocate intelligence without representation, embodiment, and cognitive developmental robotics. Intelligence is derived from interaction between an environment and an agent.

Robots that learn actions by trial and error (UC Berkeley)



Mini-cars that learn driving by trial and error (PFI社、日本)



# Future development of deep learning

## (1) Image

Extraction of features from images

Image recognition

## (2) Multi-modal

Extraction of features from video, sensor, and multi-modal information

Video recognition, video prediction, anomaly detection

## (3) Robotics

Extraction of features from sensory and motoric information

Make an action plan by symbol manipulation

Planning, inference

## (4) Interaction

Extraction of features of things in the world by interaction

Ontology, situation recognition

## (5) Symbol grounding

Connecting symbols to high-order features

Language understanding,  
Machine translation

## (6) Knowledge acquisition through language

Extraction of even higher-order features by abstraction on grounded language data

Resolution of knowledge acquisition bottleneck

recognition



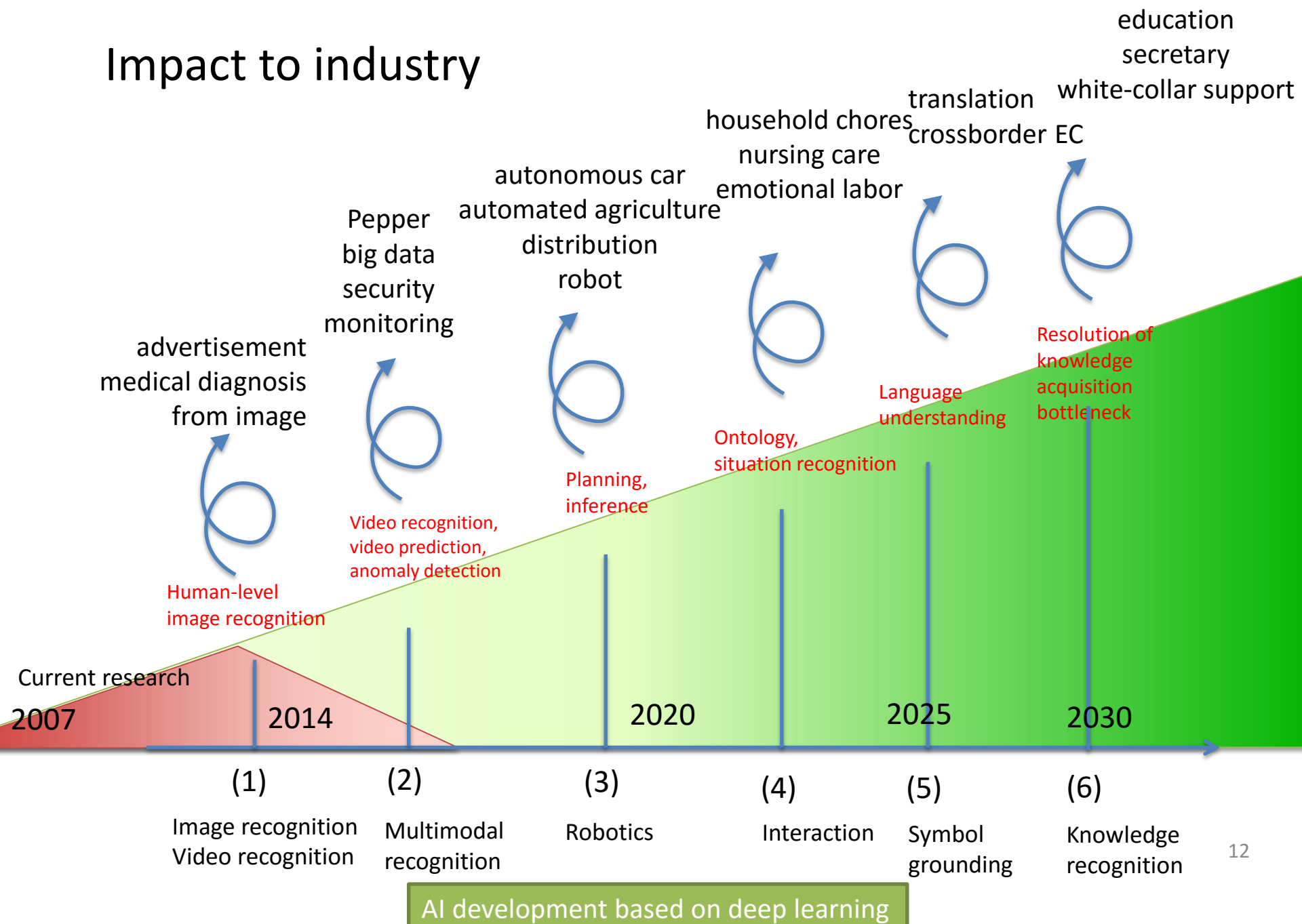
action



language

Deep learning is great,  
but the research field beyond deep learning is huge.

# Impact to industry



# “Child AI” and “Adult AI”

- **Adult AI: Sustaining innovation from big data to AI**
  - Many applications of big data and IOT, commercial projects such as Watson, Siri, Pepper...
  - Expert tasks (which can only be done by adults or experts) are automated. However, human developers in the back end extract features and build models by observing a target phenomenon.
  - A fair amount of expert knowledge and background knowledge are integrated.
  - Several areas such as marketing and advertisement have been good application domains. Other areas such as medical/educational/financial areas are expected to be promising domains, where new data become available and conventional AI methods might be applicable.
- **Child AI: Disruptive innovation from deep learning**
  - Development based on deep learning
  - (Seemingly) easy tasks that can be done even by children became possible.
  - Requires little or no background knowledge
  - Technologies will progress similarly to human child growth: recognition, action, and language.

In short, child AI requires no feature engineering by humans, but adult AI does.

# Difference between Child AI and Adult AI

- Not only in a laboratory environment.
- Not only for demonstration.
- Examples:
  - Image recognition was nearly impossible.
  - Voice recognition was good only for the demonstration time.
  - Machine translation was terrible.
  - Dialogue systems were terrible.
  - Robots can not walk outside the laboratory.
  - Robots can not even recognize things and grab things.
- Driverless cars as an example
  - Current system is based on “Adult” AI: precise measurement of current car location and very precise maps.
  - It can not run without the map.
  - Can it run safely downtown? Can it run suburbs? How about when an earthquake occurs or when any disastrous event occurs?
  - We all know that “Adult” AI is good most of the time, but not good for environmental changes or exceptions.

# Development of current industries

Agriculture

Harvest  
recognition

High applicability of  
tractors and combines

Automatic harvesting  
Automatic cultivation

Construction

measurement

Earthwork,  
Exterior work,  
Interior work,...

Automation of  
overall construction

Food processing

sorting

Cutting, paring,  
meat processing

Automation of  
various food processing

⋮

A: Image recognition

B: Proficient actions

C: Planning

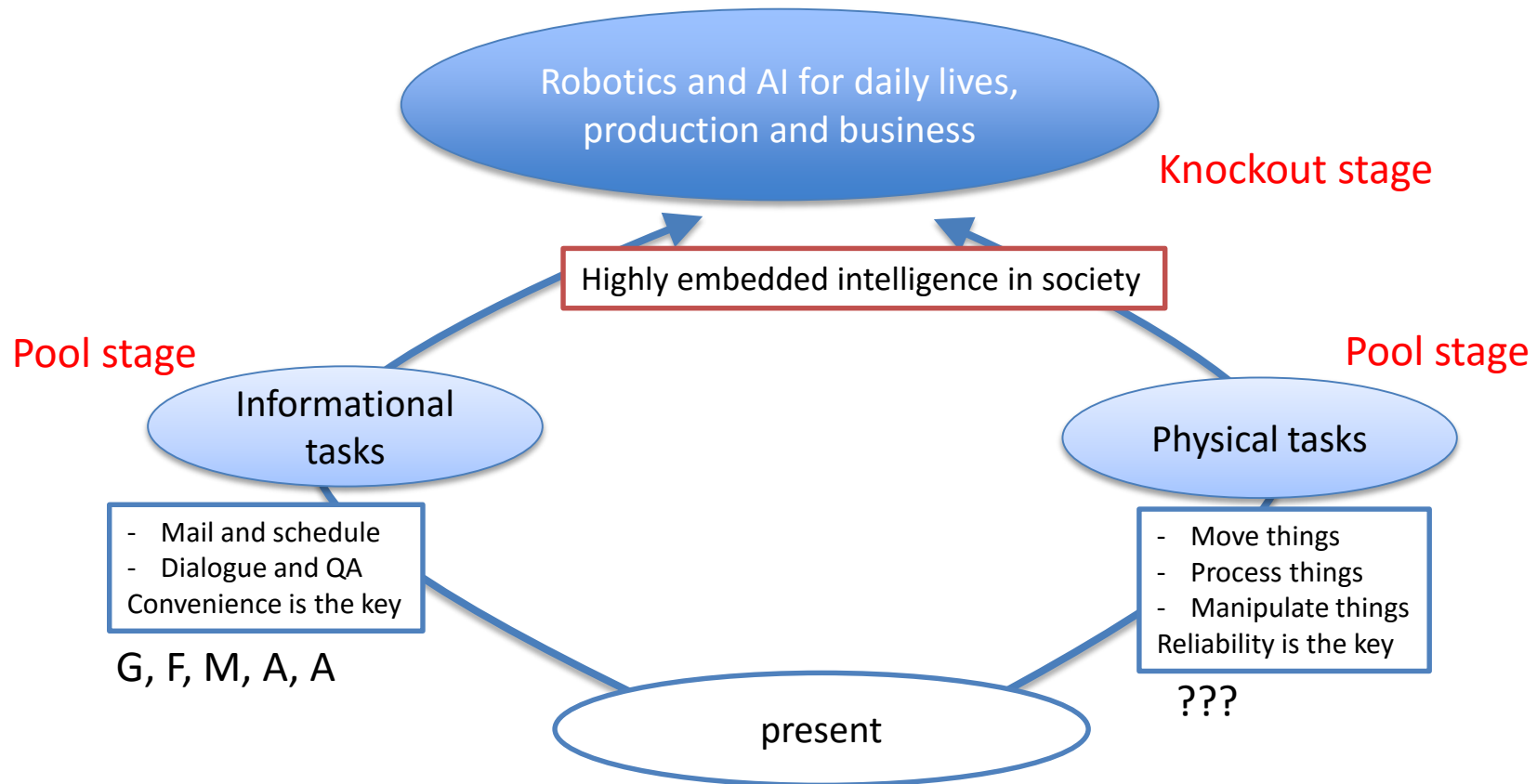
# What is the essence of the change?

- Image/Video recognition
  - Numerous tasks are being done by humans only because image/video recognition is impossible for computers
  - Costs will be reduced: for monitoring tasks, less than 1% of the current cost.
  - Many businesses will suddenly be feasible.
- Action proficiency
  - We believe that machines can do only “machine-like” behaviors, and robots can play only “robot-like” actions. This situation will change.
  - Industries handling natural objects, such as agriculture, construction, and food processing will change to an enormous degree because natural entities (such as plants, soil, and foods) mutually differ in size, color, and shape, necessitating high-level recognition for handling.



# Direction of Industries

- Finally the integration of AI, robotics, and big data is used in our daily lives and business. Many tasks (including informational tasks and physical tasks) are highly automated.
- Two ways to achieve goals: Information labor-oriented approaches and physical labor-oriented approaches
- Pool stage -> knockout stage



# Toward revival of manufacturing in Japan by AI

- Problem of Japan: Aging population and low birth rate
  - Strong needs for physical labor
  - e.g.: workforce shortage for agriculture, construction, logistics, elderly care, decommissioning of nuclear reactors, and successors of skilled labor
- Child AI is a good fit to manufacturing.
  - Japan's strength is useful for competence.
- Many good environmental factors
  - Many AI researchers in Japan: JSAI 3500 vs. AAAI 7000
  - Understanding of AI through different generations: Large organizations can make decisions promptly
  - No language barriers: algorithms and (physical) products matter
- Our actions are important
  - We start an educational program of DL in UT from Nov. 2015.
  - Various opportunities for each industry: We must find and exploit them
  - We should portray a bright picture of a future society using AI