Introduction to AIML/Intelligent Technology

John Sum
Institute of Technology Management
National Chung Hsing University
Taichung 402, Taiwan

February 22, 2023

'I can only teach what I know and I can only demonstrate what I can do', JS.

Contents

1 AIML/Intelligent Technology

Today, intelligent technology has been applied in almost every corner in the world. Every time a photo of friends has been uploaded to Facebook, Facebook will automatically square the faces and give name suggestions. The face recognition technology is an intelligent technology. While a LINE message is edited, the iPhone will automatically give next-word suggestions (predictive text). The technology behind this is an intelligent technology. A car can now drive from one place to another without human intervention. Clearly, auto-driving is yet another intelligent technology.

There are huge number of applications that have applied intelligent technologies. Sometimes, we have even been unaware of being beneficial by intelligent technologies. For instance, news feed and friends recommendation in Facebook are developed based on intelligent technology. Some network security systems have applied intelligent technology for intrusion detection.

1.1 What is Intelligent Technology ?

Intelligent technology (or intelligent technologies), it usually refers to two types of technologies. For the first type (I) *it is a collection of technologies whose algorithm*[1] *designs are inspired by or copied from the ways how human thinks and the methods how a human being solves problems.* It could be a model of neuronal network with a number of neurons. The mathematical model of each neuron is modeled by the property of a biological neuron. It could also be a hypothetical (or psychological) model mimicking human stimuli-response behavior. For the second type (II) *it is a collection of technologies for solving problems used to be solved by human beings*, like chess playing and poker game playing.

1.1.1 Type I iTech: Psychological or Biological Models Inspired

The technology applied in the fuzzy logic-based auto-parking system is an example of the first type. Fuzzy logic is theory extended from the classical logic to tackle actions involving linguistic variables like 'turn more left', 'turn more right', 'low speed' and 'high speed'. The technology applied in the optical character recognition (OCR) system is another example of the first type. Specifically, this technology being used in the 1990s and 2000s was neural network, in which the model is inspired by the neuronal network in human brain. Nowadays, the technology being used is deep neural network (equivalently, deep learning model).

For the first type of intelligent technologies, the technologies developed must have certain intelligent essence. The problems to be solved could be combina-

[1] An algorithm is essentially a step-by-step procedure (i.e. a program), an operations procedure, a process or a method for solving a problem. So, an algorithm design could be interpreted as a program design, an operations design, a procedural design or a methodology design.

torial optimization problems, financial prediction problems and system control problems. These problems could be solved by methods other than intelligent technologies. Intelligent technology is just an alternative method.

1.1.2 Type II iTech: Non-Psychological/Biological Model Inspired

Driving a car is task to be done by a human driver. Today, many car manufacturers have equipped a car with an auto-driving system. Auto-driving system is a complicated system which applies various intelligent (and non-intelligent) technologies for object recognition, speed control, road event prediction and others. Table 1 lists a number of tasks to be done by human.

The following intelligent technologies are examples of the second type.

- The technology behind face tagging in Facebook is an obvious example of this type.
- The technology behind iPhone Siri, a voice assistant system, is another obvious example of this type.
- Google AlphaGO is a system designed to play GO game. Winning a GO game is a problem to be solved by a human player. Thus, the technology behind AlphaGO is another example of the second type.
- Google translator applies an algorithm called Long-Short-Term-Memory (LSTM) model to learn to translate a sentence from one language to another. The LSTM is an intelligent technology of this type.
- The technologies behind text generation by OpenAI's ChatGPT or Google's BERT are examples of the second type. Writing a summary of an article is always a question to be solved by a high school student in public examination to assess his/her language capability. Traditionally, the student who solves this problem would need to have certain intelligence.

For the second type of intelligent technologies, the technologies developed for solving the problems used to be solved by human beings could have no intelligence nature. For example, to recognize an object in an image, the object has to be segmented in the first place. The object segmentation algorithm is basically an image processing algorithm which is not intelligent at all. That is to say, the technology developed for solving a problem used to be solved by an intelligent human could be non-intelligent.

1.2 AI versus Machine Learning

To me, the technologies developed under the Type I intelligent technologies are AI technologies as their models and algorithms are inspired by the psychological or biological findings from humans. For certain extends, they mimic the psychological models or biological models of human thinking. Here lists a number of research areas which are overlapped with the area artificial intelligence.

Table 1: List of tasks which require human thinking.

Tasks	Existing Tools or Researches	Mature?
Listening	Siri, Google Voice Assistant, Amazon Echo	Not yet, simple commands
Speaking of a given sentence	Voice synthesization	Yes
Text/Sentence/Document generation	BERT[a], ChapGPT[b]	Not yet
Reading (with summarization)	Resoomer[c], Summerizer[d]	Not yet
Translation	Google Translate	Not yet
Document editing	MS WORD, Page, Latex	Yes
Driving	Auto-driving systems	Not yet
Chess playing	IBM Deep Blue, AlphaGo	Not yet
Hypothesis generation	[1, 2, 3, 4, 5]	Not yet
Face recognition	Facebook Autotag	Yes
Object recognition	[6, 7, 8, 9]	Close to mature
Photo captioning	Under research	No
Video tagging	Under research	No
Song composing	Under research	No
Painting/Drawing	Under research	No
Story writing	Under research	No

[a] `cloud.google.com/ai-platform/training/docs/algorithms/bert-start`.
[b] `chat.openai.com`.
[c] `resoomer.org`.
[d] `summerizer.org`.

- Neural networks, neural computation.
- Spiking neural networks.
- Computational neuroscience.
- Neuromorphic computing.
- Cognitive science, connectionism.
- Reinforcement learning.

The technologies developed under the Type II intelligent technologies are machine learning technologies as their models and algorithms are not inspired by the psychological or biological findings from humans. For instance, the models developed for text summarization have nothing related to any psychological model of language understanding and generation. Here lists a number of research areas which are overlapped with the area of machine learning.

- Computational intelligence.
- Fuzzy systems.
- Evolutionary computation.
- Deep learning.
- Statistical learning.

Thus, my distinction between an AI research area and a machine learning research area is based on the model defined to mimic human behavior. If the model is inspired by psychological or biological model, the area is an AI research area. Otherwise, the research area is a machine learning research area.

Moreover, I have to stress that the categorization of the research areas has been evolving over time. The distinction between AI and machine learning could be unclear as the models for solving a humanized problem are getting more complicated. In this regard, the models being developed could compose of AI models and machine learning models. Under such circumstances, it would be difficult to state clearly which research area belongs to which category.

1.3 Other References for Intelligent Technology

Apart from the above references, intelligent technology could be referred to (equivalently, interpreted as and perceived as) a *product* like Sony AiBo, a *service* like Amazon ECHO, a collection of technologies like neural machine translation and deep neural networks, a machine learning algorithm. The product and service are called the *intelligent product* and *intelligent service*.

Intelligent products and intelligent services could also be integrated and applied to develop other intelligent products and intelligent services. Moreover, the technologies developed in areas of *AI*, *machine learning* and *cognitive computing* are intelligent technologies.

2 Intelligence & Learning

The very first concept to be clarified in intelligent technology or artificial intelligence is definitely the definition of intelligence. It is not an easy question. If you look up the information over the Internet[2], you will find dozen of definitions of intelligence which are defined by scholars from various fields, like in psychology [10]. But, how many of them you really understand. In regard to artificial intelligence, many scholars have also attempted for its definition or even its theory [11, 12] Nevertheless, as mentioned in [13], '*Despite a long history of research and debate, there is still no standard definition of intelligence. This has lead some to believe that intelligence may be approximately described, but cannot be fully defined.*'

2.1 Definition of Intelligence

In this section, I am going to define intelligence based upon my experience. For sure, you can disagree and give your own definition if you wish. First of all, one should admit that each of us is intelligent. Equivalently, human has intelligence. We have intelligence. A human being is intelligent because human can solve problem[3].

Axiom 1 (Human Intelligence) *Every person is intelligent. Every person has intelligence. Every person can solve at least one problem (i.e. survival) in his/her lifetime.*

Thus, we can have the following definition for human intelligence.

Definition 1 (Human Intelligence) *(1) Human intelligence is the capacity of a human in solving a problem. (2) Human intelligence is attributed to the collective behavior of a sequence of purposeful actions a human takes in response to an external stimuli.*

Accordingly, we can have the following corollary regarding the intelligence of a living organism.

Corollary 1 (Intelligence) *Every living organism is intelligent as every living organism is able to find the way to survive.*

Now, we can give the definition of intelligence for human and living organism with an additional property for it. Intelligence must be observable or measurable.

Definition 2 (Intelligence) *(1) Intelligence is attributed to the collective behavior of a sequence of purposeful actions a living organism takes in response to an external stimuli. (2) Intelligence must be observable.*

[2] `https://en.wikipedia.org/wiki/Intelligence` for instance.

[3] In the *Society of Mind* [14] which is authored by Marvin Minsky, intelligence merely means what people usually mean — the ability to solve hard problems.

A person's intelligence has to be observable from the person's reaction to the external stimuli. Therefore, a person cannot be claimed as intelligent driver if he/she claims that he/she can drive in his/her dream/imaginiation.

Regarding the problem and action as stated in Definition 1 and Definition 2, here are some examples for illustration.

Problem	Organism	Stimuli	Action(s)
Eat	Amoebae	Food	Move to the location of the food.
			Surround the food.
			Ingest the food.
Survive	Ant	Danger	Run away.
Survive	Baby	Hungry	Baby cries.
			(Mother comes.)
			(Put its head near a nipple.)
			Move its month to a nipple.
			Suck the breast milk.

Based on Definition 2, unicellular organism are intelligent. A virus is intelligent. A cancer cell is intelligent. Multicellular organisms, like plants and animals, are intelligent. Every animal is intelligent. Human is certainly intelligent. All living things are intelligent. For non-living things, like stone, I am not going to make any claim on their intelligence.

2.2 Definitions of AI and ML

By the same token, a machine is intelligent if it can solve problem(s). Owing not to argue on the number and the nature of the problems, my **ground zero definition** on intelligent machine is stated below.

Definition 3 (Intelligent Machine) *A machine is intelligent if it can solve at least one problem.*

Based on the definitions for human intelligence and intelligence, machine intelligence and artificial intelligence could now be defined in similar manner.

Definition 4 (Machine Intelligence) *Machine intelligence could be referred to the capacity a (non-living) machine in solving a problem.*

Definition 5 (Artificial Intelligence) *Artificial intelligence is attributed to the collective behavior of a sequence of purposeful actions a non-living organism takes in response to an external stimuli. Moreover, those behaviors mimic the human behaviors, psychologically or biologically, in response to that external stimuli.*

Definition 6 (Machine Learning) *Machine learning is referred to a collection of technologies, respectively machines and algorithms, which are able to generate a sequence of purposeful actions to response to an external stimuli that a human will face.*

Therefore, every intelligent machine has artificial intelligence or simply every intelligent machine has intelligence. On the other hand, every human has intelligence but no any human has artificial intelligence. The distinction between AI and machine learning is based on the models and algorithms developed for the system. If the models and algorithms are developed based upon psychological and/or biological models, the system is an AI system. Otherwise, it is a machine learning system.

2.2.1 Note on learning

It should be noted that **learning is not, and should not be, a factor included in the definition of intelligence**. The reason could be explained by the following examples. A cancer cell is able to find a way to survive. A neuron is able to react to the signals received from its dendrites. One argument is that each cancer cell and neuron cell have been programmed to do so. Would there be any learning procedure designed in the program? It is hard to tell.

Here is another example. An auto-driving car is able to move in a city and park itself in a parking slot. The auto-driving system acts like a human driver controlling the car moving on a road and parking to a parking slot. This auto-driving system is commonly claimed as intelligent. However, the job it does is based on a program running in a computer installed inside the car. The program is able to sense the environment, recognize the objects around and then generate appropriate electrical signals to the mechanical system. Everything is programmed.

While learning is not, and should not be, a factor included in the definition of intelligence, it is clear that a person could solve more and more problems via learning.

2.2.2 Note on machine

A machine is not limited to a computer machine. It can be a mechanical computer. Like the Difference Engine[4] which was built by Charles Babbage in 1822, it was made of mechanical parts and it was applied to solve approximation problems. By Definition 3, Difference Engine is intelligent.

2.2.3 Note on problem difficulty (complexity)

Here, difficulty of a problem is not included in the definitions of intelligence, machine intelligence, artificial intelligence and the definition of intelligent machines. One reason is that problem difficulty is not easily measured.

One might argue that problem difficulty could be measured by the complexity of the solution for solving such problem. The definition is in the same manner as the one defined in the area of computer algorithms[15]. While it seems reasonable, this definition has one important constraint. The complexity of a problem is defined as the complexity of the **best** algorithm for solving the

[4] `https://en.wikipedia.org/wiki/Difference_engine`.

problem. The best algorithm is the algorithm with the minimum number of steps for solving the problem. Thus, it comes to a question.

Let say, the best algorithm is designed by John. Peter does not know the best algorithm and designs another algorithm to solve the same problem. The algorithm requires more steps. Then, could we say that John is more intelligent than Peter? To me, I will not. I only say that both John and Peter are intelligent. So, *I will not say that a machine is intelligent if it can solve a hard problem.*

2.3 Two Additional Properties of Intelligence

Two important properties have to be added to the notion of intelligence. First, intelligence must be manifested from the interaction between a living organism and its environment.

Definition 7 (Observable) *Intelligence can only be attributed when a living organism has taken purposeful actions to an external stimuli. The interaction between the living organism and the environment must be observed by a third party.*

For human beings, intelligence can only be attributed only when a person has taken actions to solve a problem. No one should claim himself/herself being intelligent if he/she has not taken any action to solve any problem. Therefore, the following arguments are not valid under the above arguments.

- I can imagine how to solve a problem. So, I am intelligent.
- I can dream of controlling a fight jet. So, I am intelligent.

Imagination or dreaming of solving a problem does not actually take actions to solve the problem. Therefore, a person cannot claim himself/herself being intelligent solely based on his/her imagination or dreaming.

The second property of intelligence is that the end result of the actions taken to solve a problem must be assessable. From the end result, one can justify if the problem has been solved successfully.

Definition 8 (Measurable) *The end result of the actions taken by an intelligent body in solving a problem must be measurable. That is to say, outcome from the interaction between an intelligent body and the environment could lead to a third party to determine if the problem has been solved completely.*

Let me take an example. A person has claimed that he/she has completed an assignment. The correctness (or quality) of the work presented in the assignment has to be assessed. Only when the problem has been successfully solved, the person who solved this problem is intelligent and thus the person has intelligence.

Let me take another example. Driving a car from my home to the school is a problem. Before starting the engine, I need to design the route to the school. Some people would consider that I am intelligent because I have designed the route. This comment is correct only if route design is the problem to be solved. However, I will say 'not yet'. It is because the actual problem is to drive a car to school. It has not yet been solved. Suppose that I have finally driven my car back to school. Only in this moment, I can claim that I am intelligent.

With these additional properties, we could avoid those so-called intelligent systems which can only generate random actions.

3 Turing Test on Machine Intelligence

While we have intelligence quotient (IQ) to measure the intelligence capacity of a human being, there is no such measure for AI or machine intelligence. Only a test proposed by Alan Turing could be used to examine if a machine has intelligence.

3.1 Definition

Since Alan Turing has presented a number of conceptual computing machines for solving some decision and mathematical problems in 1930s [16, 17], he turned to think of using a computing machine to solve intelligent tasks [18]. In this regard, a very first question is how to assess a machine if it is intelligent. The answer is based on a test, the famous Turing Test, its procedure is presented in [18].

Definition 9 (Turing Test) *Suppose a machine and a human are placed in two rooms. A human tester writes a question on two pieces of papers and the papers are pass to both rooms. The machine (resp. human) responses to the question by putting the answer on a piece of paper. Based upon on their answers, the human tester has to identify from their answers which room has a machine inside. If the human tester fails to identify, the machine is intelligent.*

By the Turing test, one can readily infer that Deep Blue, the machine defeated the world champion on chess game, and AlphaGo are not intelligent. It is because we can identify from the number of wins who is machine.

My definitions on intelligent machine (i.e. Definition 3), human intelligence (i.e. Definition 1) and intelligence (i.e. Definition 2) are inspired by the Turing Test. By using the same jargon, Turing Test is a bottom-line test. No matter in my definitions regarding intelligence and the Turing Test on an intelligent machine, the method how a human (resp. machine) is able to solve a problem or answer a question is not considered.

3.2 Machines Failing Turing Test

As a matter of fact, many so-called AI software or AI systems are fail in the Turing test. For instance, a system which can only generate random responses must fail the test. Moreover, those AI systems which outperform human testers, like IBM Deep Blue, IBM Watson and Google AlphaGO, are fail to the Turing test. One simple reason is that these systems are built to win the games. They are unable to play the games with tier-break. Human testers could easily identify if the systems are human or non-human. If a system almost wins a game, it must be non-human.

4 Informal Learning Theory

While some problem solving abilities are innate, many problem solving abilities are learned from our experience. These problem solving abilities are the end results attained after a number of learning processes.

4.1 Definition of Human Learning

No matter it is innate or problem driven, every learning process lets us acquire sufficient knowledge to solve a specific problem.

Definition 10 (Human Learning) *Learning is a process in which a person acquires sufficient knowledge to solve a problem by himself/herself. The knowledge includes (1) the factual events required and (2) the procedure for solving the problem.*

The factual events could be the data, the information and the casual relations among the events. The procedure[5] is the steps for solving the problem. If the factual events and the procedure are not available, acquisition of the factual events and the design of the procedure are yet another two problems to be solved. Therefore, we can have another two definitions for (two sub-tasks of) learning, as stated in the following definitions.

Definition 11 (Fact Acquisition) *Learning is a process in which a person acquires sufficient knowledge to acquire factual events for solving a problem by himself/herself.*

Definition 12 (Procedure Design) *Learning is a process in which a person acquires sufficient knowledge to design a procedure for solving a problem by himself/herself.*

Along with Definition 11, learning could also be defined as the process for us to get the regularities hidden in our world. The regularity could be a statistical model of certain events. For instance, the chance of getting '1' from a fair dice is 1/6. It could be a casual model for a set of events. If the sky is getting darker and there are clouds in the sky, it is highly likely to have rain soon. So, Definition 11 could be re-stated in the following definition.

Definition 13 (Regularity Learning) *Learning is a process for a person to know the regularities of the world.*

[5] It can also be called the operation, the operational procedure, the solution, the method or the algorithm.

The regularities being attained from learning are usually applied for solving problems. Learning process is always problem specific. We learn many things to solves many problems. Learning is process for us (or a machine) to be intelligent.

The simplest learning is to memorize all the factual events and the procedure. But, to me, it is not learning. The person who memorizes all the factual events and the procedure for him/her to solve the problem is no difference from a computer program. This person, be definition, is still intelligent. But, his/her intelligence would be questionable. Memorization should not be treated as learning. So now, what should be considered as learning?

4.2 Regularity Learning

Here, I have an idea. Let say, someone has designed (a procedure) a trading rule (based on the assumption that the stock price p of a listed company follows Gaussian distribution with stationary mean $\bar{p}$ and variance $\bar{S_p}$), *a stock can be sold when* $p \geq \bar{p} + 2 \times \sqrt{\bar{S_p}}$ *and the stock can be bought when* $p \geq \bar{p} - 2 \times \sqrt{\bar{S_p}}$.

Trading the stock is now a problem to be solved. The procedure to solve the problem is very clear. But, you need to acquire the factual events, i.e. the **hidden regularity** of the change of the stock price. No way out, we need to learn from the historical stock price for the $\bar{p}$ and the $\bar{S_p}$. Clearly, it is the learning process defined in Definition 11. To do so, we need to develop a procedures for us to get $\bar{p}$ and $\bar{S_p}$. As $\bar{p}$ **and** $\bar{S_p}$ **are unknown**, the best that we can do is to estimate their values. The estimation is in fact the process of learning – learn from the historical and the future stock prices to estimate the values of $\bar{p}$ and $\bar{S_p}$.

We let $\hat{p}_N$ and $\hat{S}_N$ be the estimates of $\bar{p}$ and $\bar{S_p}$ based on N observable data $p_1, p_2, \cdots, p_N$.

$$\hat{p}_N = \frac{1}{N}\sum_{k=1}^{N} p_k, \quad \hat{S}_N = \frac{1}{N}\sum_{k=1}^{N}(p_k - \hat{p}_N)^2, \tag{1}$$

and $\bar{p} \approx \hat{p}_N$ and $\bar{S_p} \approx \hat{S}_N$. The goal of the learning process is to learn to known $\bar{p}$ and $\bar{S_p}$ as good as possible. Getting their values by using the formula in (1) is the bottom-line approach. We still can learn from it to get the values. However, it is not efficient. Especially when a new data coming in, say p_{N+1}, $\hat{p}_{N+1}$ and $\hat{S}_{N+1}$ will have to be evaluated based on the whole set of data $p_1, \cdots, p_{N+1}$. It will be time-consuming.

A philosophy behind learning is that we should not start from zero to learn something new. We should learn something new based on something already known. In this regard, another approach is to develop another learning algorithm. As it is known that $N\hat{p}_N = \sum_{k=1}^{N} p_k$ and $(N+1)\hat{p}_{N+1} = \sum_{k=1}^{N+1} p_k$, we can get that

$$(N+1)\hat{p}_{N+1} = \sum_{k=1}^{N+1} p_k = N\hat{p}_N + p_{N+1}.$$

Thus,

$$\begin{aligned} \hat{p}_{N+1} &= \frac{N}{N+1}\hat{p}_N + \frac{p_{N+1}}{N+1} \\ &= \hat{p}_N - \frac{1}{N+1}\left(\hat{p}_N - p_{N+1}\right). \end{aligned} \tag{2}$$

This is the way we learn from something new, i.e. p_{N+1}, based on something already known, i.e. $\hat{p}_N$, to get a better estimate on $\bar{p}$, a regularity of the stock price.

For the variances $\hat{S}_{N+1}$ and $\hat{S}_N$, we can go through similar step for an iteration equation as (2).

$$N\hat{S}_N = \sum_{k=1}^{N} \left(p_k - \hat{p}_N\right)^2, \quad (N+1)\hat{S}_{N+1} = \sum_{k=1}^{N+1} \left(p_k - \hat{p}_{N+1}\right)^2.$$

As

$$\begin{aligned} \sum_{k=1}^{N+1} \left(p_k - \hat{p}_{N+1}\right)^2 &= \sum_{k=1}^{N} \left(p_k - \hat{p}_{N+1}\right)^2 + \left(p_{N+1} - \hat{p}_{N+1}\right)^2 \\ &= \sum_{k=1}^{N} \left(p_k - \hat{p}_N + \hat{p}_N - \hat{p}_{N+1}\right)^2 + \left(p_{N+1} - \hat{p}_{N+1}\right)^2 \\ &= \sum_{k=1}^{N} \left(p_k - \hat{p}_N\right)^2 + 2\left(\hat{p}_N - \hat{p}_{N+1}\right)\sum_{k=1}^{N} \left(p_k - \hat{p}_N\right) \\ &\quad + \sum_{k=1}^{N} \left(\hat{p}_N - \hat{p}_{N+1}\right)^2 + \left(p_{N+1} - \hat{p}_{N+1}\right)^2, \end{aligned}$$

we can get that

$$\hat{S}_{N+1} = \frac{\hat{S}_N + N\left(\hat{p}_N - \hat{p}_{N+1}\right)^2 + \left(p_{N+1} - \hat{p}_{N+1}\right)^2}{N+1}. \tag{3}$$

The recursive equations (2) and (3) constitute the **learning algorithm** for getting the **hidden regularity of the stock price**. This learning algorithm conforms to the learning process as stated in Definition 11.

S1 Initialize $\hat{p}_0 = 0$ and $\hat{S}_0 = 0$.

S2 Repeat the following steps whenever a new stock price p_k, for $k \geq 1$, has been input.

S2.1 Estimate the mean by $\bar{p} \approx \hat{p}_k$, where

$$\hat{p}_k = \hat{p}_{k-1} - \frac{1}{k}\left(\hat{p}_{k-1} - p_k\right).$$

S2.2 Estimate the variance by $\bar{S} \approx \hat{S}_k$, where

$$\hat{S}_k = \frac{\hat{S}_{k-1} + (k-1)\left(\hat{p}_{k-1} - \hat{p}_k\right)^2 + \left(p_k - \hat{p}_k\right)^2}{k}.$$

S2.3 Make the trading decision T_k, where

$$T_k = \begin{cases} \text{Buy} & \text{if } p_k \leq \bar{p} - 2\sqrt{\bar{S}}, \\ \text{Sell} & \text{if } p_k \geq \bar{p} + 2\sqrt{\bar{S}}, \\ \text{No action} & \text{Otherwise.} \end{cases}$$

In the above example, only the steps S2.1 and S2.2 are the learning steps (or the learning algorithm). They learn from the incoming stock prices $p_1, p_2, \cdots, p_N, \cdots$ for the mean and the variance of the stock prices. As the stock price is assumed to be following Gaussian distribution, the regularity of the stock prices will be known if the mean $\bar{p}$ and $\bar{S}$ are known. **In this example, learning is nothing else but simply doing estimation**.

Here, three points should be noted. First, one factor leading to the success of the trading is based on the assumed model $p \sim \mathcal{N}(\bar{p}, \bar{S})$. Precisely, Gaussian distribution is a **family of models with the same mathematical definition** $\mathcal{N}(\bar{p}, \bar{S})$. $\mathcal{N}(0, 1)$ is a model. It is difference from $\mathcal{N}(1, 2)$. As $\bar{p}$ and $\bar{S}$ are the parameters of the family of Gaussian distributions, $\mathcal{N}(\bar{p}, \bar{S})$ is also called the parametric model for the family. Sometimes, an author might ignore the word 'parametric'.

Definition 14 (Parametric Model) *A parametric model is referred to a family of mathematical model of the same form. A model is particular instance. It is a parametric model in which the parameters are defined to specific values.*

Every learning algorithm must be developed based on an assumed parametric model. If there is something wrong with the assumed parametric model, the performance might be degraded. Therefore, for advanced learning theory, model selection is another big issue to be concerned.

Second, another factor leading to the success is the quality of learning, equivalently the quality of estimation. In other words, will $\hat{p}_k \to \bar{p}$ and $\hat{S}_k \to \bar{S}$? To answer this question, one would need to have good foundation on *Statistical Theory*. Without showing the detail proofs, it has been shown that $\hat{p}_k$ is a good estimate for $\bar{p}$. $\hat{S}_k$ is not a good estimate for $\bar{S}$.

Third, if the above algorithm is implemented by a computer program and the computer has been connected to the stock market, the computer can automatically trade the stocks for us and become an automated trading system. In accordance with the Definition 3, this computer can be claimed as an intelligent trading system even though you might not feel in that way.

4.3 Procedural Learning

In the above example, the decision rule has been designed – buy (resp. sell) the stock is $p_k \leq \bar{p} - 2\sqrt{\bar{S}}$ (resp. $p_k \leq \bar{p} + 2\sqrt{\bar{S}}$). However, there is a parameter in the decision rule. It is the value '2'. Why should it be '2', but not '1' or '3'? To answer this question, an investor will clearly tell you that this value can be changed via learning.

Recall that the purpose of learning is to let oneself to be more intelligent. However, the measure of intelligence is not definable. It is problem-specific and based on the *goal* of a person on the problem to be solved. Suppose the goal of the stock trading problem is to make the most profit. Then, we can decompose the stock trading problem into two sub-problems.

P1 Estimate the values of $\bar{p}$ and $\bar{S}$.

P2 Determine the factor α in the trading rule $p_k \leq \bar{p} - \alpha\sqrt{\bar{S}}$ (resp. $p_k \leq \bar{p} + \alpha\sqrt{\bar{S}}$).

For the second problem, a learning process will have to be developed so that the optimal value of α can be estimated.

5 Mathematical Learning Theory

Learning involves the process of regularity learning and procedural learning. The theories of learning for these two different tasks have subtle difference. Theory of regularity learning is normally referred to the so-called teacher learning. Theory of procedural learning is considered as goal-directed learning. While these two theories could be unified as a single learning theory, I am not going to do so in here. As an introductory text, I simply introduce in the following text the concepts regarding the theory of regularity learning.

Normally, an intelligent technology (but not all intelligent technologies) associated with a *parametric model* (equivalently, an hypothetical model) which generates the observations (resp. samples). As the parameters of the true model are unknown, it is inevitable to develop *an algorithm to update the parameters of a model (i.e. learning algorithm)* such that the parameters of the true model can be found eventually.

Here is a simple example. Let say, we have a set of N samples $\mathcal{D} = \{\mathbf{x}_k, y_k\}_{k=1}^{N}$, where $\mathbf{x}_k \in R^n$ and $y_k \in R$ for $k = 1, \cdots, N$. Assuming that this data set is generated by a linear regressor, i.e.

$$y_k = a + \mathbf{b}^T \mathbf{x}_k + \xi_k, \tag{4}$$

where ξ_k is a random noise. As a and $\mathbf{b}$ are unknown, we define the following model to learn from the samples the true parameters.

$$f(\mathbf{x}_k, \hat{a}, \hat{\mathbf{b}}) = \hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_k. \tag{5}$$

Given $\hat{a}$, $\hat{\mathbf{b}}$ and $\mathcal{D}$, the mean square error (MSE) between the parametric model

(5) and the true model (4) is given by

$$E(\hat{a}, \hat{\mathbf{b}}) = \frac{1}{N} \sum_{k=1}^{N} \left(y_k - f(\mathbf{x}_k, \hat{a}, \hat{\mathbf{b}}) \right)^2$$
$$= \frac{1}{N} \sum_{k=1}^{N} \left(y_k - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_k \right) \right)^2 . \quad (6)$$

Taking derivative of (6) with respect to $\hat{a}$ and $\hat{\mathbf{b}}$, we get that

$$\frac{\partial E(\hat{a}, \hat{\mathbf{b}})}{\partial \hat{a}} = -\frac{2}{N} \sum_{k=1}^{N} \left(y_k - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_k \right) \right) \quad (7)$$

$$\frac{\partial E(\hat{a}, \hat{\mathbf{b}})}{\partial \hat{\mathbf{b}}}, = -\frac{2}{N} \sum_{k=1}^{N} \left(y_k - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_k \right) \right) \mathbf{x}_k. \quad (8)$$

Clearly, the true parameters could thus be estimated by setting the above equations to zeros. It works for N which is not large, say $N = 10^5$.

For large N, say $N = 10^{12}$, this method will not be feasible. An alternative approach is to design the search for true parameters by the following iterative equations.

$$\hat{a} \leftarrow \hat{a} + \mu \left(y_t - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_t \right) \right) \quad (9)$$

$$\hat{\mathbf{b}} \leftarrow \hat{\mathbf{b}} + \mu \left(y_t - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_t \right) \right) \mathbf{x}_t, \quad (10)$$

where μ is a small positive number called step size, $(\mathbf{x}_t, y_t)$ is a sample randomly picked from $\mathcal{D}$. The initial values of $\hat{a}$ and $\hat{\mathbf{b}}$ are small random numbers around zero. As a result, an algorithm to estimate the true parameters could be listed below.

S1 Initialize $\hat{a}$ and $\mathbf{b}$ to small random numbers around zero. Set $\mu = 0.01$.

S2 Repeat the following steps until the square error is smaller than 0.0001.

- S2.1 Pick a sample randomly from $\mathcal{D}$ and set it to be $(\mathbf{x}_t, y_t)$.
- S2.2 $\hat{a} \leftarrow \hat{a} + \mu \left(y_t - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_t \right) \right)$.
- S2.3 $\hat{\mathbf{b}} \leftarrow \hat{\mathbf{b}} + \mu \left(y_t - \left(\hat{a} + \hat{\mathbf{b}}^T \mathbf{x}_t \right) \right) \mathbf{x}_t$.

The above procedure is called the *learning algorithm* for the model (5) to learn from the data set $\mathcal{D}$ the behavior of the true model (4).

Thus, the *mathematical model*, the *cost function* to measure the fitness of the model and the *learning algorithm* are the essential components for an intelligent technology. To understand the working principle and the limitation of an intelligent technology, one would need to understand these three components. One note to add, the *cost function* is also called the *learning objective function*, *learning objective* or *objective function*. By showing its value after each round of learning, i.e. the steps S2.1, S2.2 and S2.3, one can check the progress of learning.

6 Related Issues

6.1 Learning Machines

The term *learning machine* has already appeared in 1959 in an article authored by Friedberg on specialized learning machine [19] and in 1965 by Nils J. Nilsson on a general introduction on the learning machines in that era [20]. In that period of time, a learning machine was referred to a digital computing machine like IBM 704 or a specialized designed machine like Perceptron Mark I that is able to implement a learning algorithm. Today, the term *learning machine* has rarely been linked to an intelligent technology or a hardware with intelligent technology inside.

6.2 Smart Home/City/Material

While the Chinese translation of the 'smart' in smart home, smart city and smart material is the same as the 'intelligent' in intelligent technology, one should not confuse that the actual meanings of both of them are very difference. Strictly speaking, smart homes, smart cities and smart materials are not intelligent, while intelligent technology could be applied in making part of a home (resp. city) smart. A smart home is normally referred to a home with fully automated control of the home appliances. For instance, the A/C could be set to be automatically on at 18:00 every day. All the lights in the living room will be off if the sensors sense no any conversation at home for more than 15 minutes. Amazon ECHO and Google HOME are two intelligent systems that can used for making a home smart. However, the services delivered by Amazon ECHO and Google HOME are a lot more than making a home smart.

Regarding smart material, the goal is even far different from intelligent. The ultimate goal of smart material is to synthesize new materials for special applications. The materials include the material for making lighter cloth for athletes and soldiers, the harder and stronger material for making fighting jets, the material coating on a fighting jet to make it invisible under any radar system. So, smart material is not an intelligent technology. It has nothing related to intelligence. One should be confused.

6.3 Intelligent Technology : What is It ?

Based on the definitions of *human intelligence* and *intelligence*, i.e. Definition 1 and Definition 2, intelligent technology could simply be defined as following.

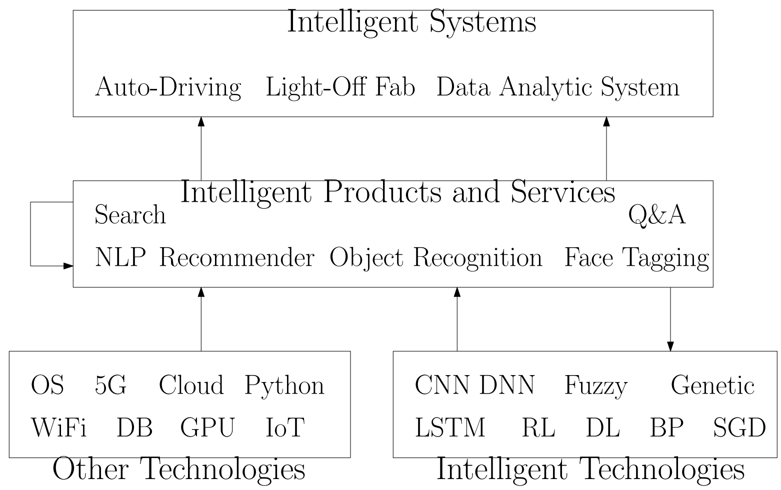

Figure 1: Intelligent technologies, intelligent products/services and intelligent systems.

Definition 15 *Any technology that can be applied to solve a problem used to be solved by a living organism is an intelligent technology.*

It is a ground zero definition. Once a technology is able to be applied to solve a problem, it is intelligent. So, almost all technologies are intelligent technologies based on the above definition.

Pretty clear, not every one accepts this ground zero definition. So, after all, would there be a better definition for intelligent technology? Here, I give one.

Definition 16 *An intelligent technology is a technology that is able to solve a problem used to be solved by human beings.*

This technology could have no any inspiration from human behaviors or biological neural structure. On the other hand, intelligent technology could be a technology its *model* is inspired by *human behaviors* or *biological neuronal structure*. It is a technology its *learning algorithm* is inspired by *human genetic evolution*. These intelligent technologies are applied to a wide range of problems including engineering problems, management problems and others. The relations amongst intelligent technologies, intelligent products/services and intelligent systems are shown in Figure 1. In the bottom level, there are (generic) intelligence technologies, including those models that you can find in AI/ML textbooks, and non-intelligent technologies, including computer technologies and communication technologies. These two types of technologies could thus be applied to develop intelligent products and services.

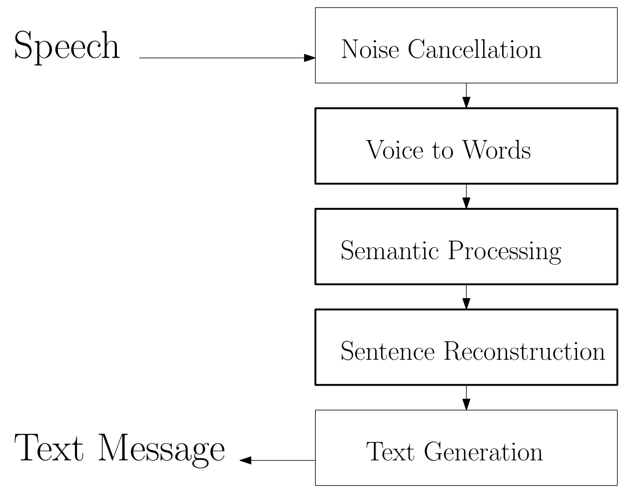

Figure 2: Schematic diagram for the speech-to-text service.

6.4 iService using both iTech and non-iTech

Let us have an example. iPhone is an intelligent system with two intelligent services FaceID and Siri. FaceID is a built-in security system to authenticate the user. It uses a face recognition software to capture the 3D face features of the user and use them as the key to unlock the iPhone. The face recognition software is developed based on intelligent technologies together with image processing technologies (non-intelligent technologies) for face recognition.

Siri is another built-in system for converting speech to text. Once the Siri is on, user could speak out a speech and then Siri will convert the speech to a text message. Imagine that your speech is a command like 'phone call Mary please'. It is clear that Siri can get this text message as well. If this message conforms to the format of a voice command and the name Mary is listed on the phone book, the voice command module in the iPhone will act on behalf of the user to make the phone call to Mary.

Figure 2 shows the schematic diagram of the technologies behind the speech-to-text service. A user speaks a speech which is then sensed by the built-in microphone and converted to a series of electrical signal. As background noise exists in the environment, the electrical signal consists of both the speech signal and the background noise. So, the electrical signal generated by the microphone will pass to a filter for noise cancellation and get a clean speech signal. In this first step, the technology applied is not intelligent. It is a simple signal processing technique.

For the second step, the clean signal is then passed to a voice-to-word module. The work to be done in this step is complicated.

- Word segment identification – To identify which part of the signal is likely to be a word.

- Word segmentation – Find and cut the signal into segments. Each segment corresponds to an unknown word.
- Word recognition – For each segmented signal, find out the corresponding word.
- Word concatenation – Combine the words to form a sentence.

After the second step has completed, a word strings will be get. For example, the sentence could read like below.

After the second step: **I an are hand some man**.

Clearly, this sentence seems not quite correct. So, this text message will then pass to another module for further processing. In this step, the technology for word recognition is an intelligent technology.

In the third and forth step, the string of texts will then be passed for semantic processing. This step is even complicated. So, I am not going to tell the detail. The result is that the module tries to find the appropriate semantic meaning of the text and makes correction. Finally, a new string of texts.

After the forth step: **I am a handsome man**.

Here, the technologies for semantic processing and sentence reconstruction are intelligent. In technical terms, they are natural language processing (NLP) or language understanding technology.

In the final step, the reconstructed text message is sent to the APP, like LINE and WhatsAPP, for display. Clearly, the technology for this step is not an intelligent technology.

6.5 Graphical Processing Unit (GPU): A Driving Force for AI

One important hardware which accelerates the advancement of intelligent technology is the graphical processing unit (GPU). GPU is a special-designed processor used to be applying in handle intensive mathematical calculations in real-time video processing. Imagine that you are playing an on-line game, in which the background images have to be processed in real-time. As a general purpose CPU, like Intel CPU, normally takes much longer time to render the animation, the player will feel uncomfortable lag on the animation and thus quit the game. With a video card in which a GPU is installed, rendering animation would become a piece of cake. Uncomfortable lag is obsolete.

GPU is a processor designed particularly to handle specific mathematical calculations for image rendering and video processing. Its design is much simpler than a general purpose CPU which is designed to handle everything, like keyboard input, panel output, logical operations and arithmetic operations and others. Thus, the processing power of a GPU in mathematical calculations could be more than thousand time faster than a general purpose CPU.

As a GPU is particularly designed to handle mathematical calculations, it has then been applied to handle complicated and time-consuming learning algorithm. In the previous example about learning algorithm, the hypothetical model is simply a linear regressor. While the sample size N could be very large, say $N = 10^{12}$, the computational complexity (i.e number of multiplications) per step is just in the order of $\mathcal{O}(n)$, where n is the size of the parametric vector $\mathbf{b}$.

However, for some specific models like deep neural network models, the per-step computational complexity could be in the order of $\mathcal{O}(n^3)$ and n could be larger than 10^6. In this regard, the learning process would take weeks to complete in a computer with general purpose CPU only. Hence, in the 2010s, researchers in AI/ML started to map the learning process to GPU and demonstrated that the processing time for a learning could be reduced to just a few days and even a few hours[6]. Subsequently, many research groups followed and purchased GPUs to accelerate their researches on the development and the applications of the machine learning algorithms.

Therefor, GPU has to be worth mentioned with intelligent technology as it is a major driving force for the advancement of intelligent technology. In the early days, Nvidia is the major GPU designer and chip maker. Today, many firms have been involved in design and/or making GPU. Intel, AMD and Apple are three other players in the market. The A-series system-on-chip (SoC) processor by Apple has already embedded with multiple CPUs and multiple GPUs in it. Intelligent service developed from non-in-house Apple developers for such on-chip GPUs has yet to be explored.

6.6 Cloud Platform: Yet Another Driving Force for AI

In Section 6.4, we have mentioned two intelligent services, Siri and FaceID. As a matter of fact, various tech giants have already released a number of intelligent services on their cloud platforms. Here are some examples[7].

- IBM Cloud – Watson Speech to Text, Watson Text to Speech, Watson Language Translator, Watson Visual Recognizer, IBM Watson Services for CoreML, etc.
- Amazon Web Services – Amazon Lex (voice to text), Amazon Polly (text to voice), Amazon Rekonigtion for image analysis, Amazon Machine Learning, etc.
- Microsoft Azure – AI Services like Azure Cognitive Services and Azure Machine Learning; AI Tools and Framework; and AI Infrastructure, etc.
- Google Cloud – Cloud Vision API, Cloud Intelligence API, Natural Language API, Cloud Translation API, Speech-to-Text API, Text-to-Speech API, Tensor Processing Unit (TPU), other Cloud ML services.

[6] Search from the Google for the information about ImageNet competition and AlexNET.

[7] `https://www.datamation.com/artificial-intelligence/the-top-cloud-based-ai-services.html`.

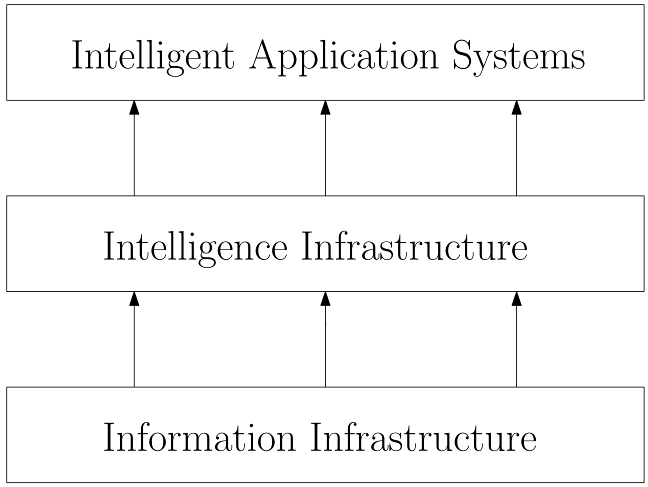

Figure 3: Intelligence infrastructure. The idea has also been advocated by Michael I. Jordan in [21].

On one hand, these cloud platforms provide so-called virtual computers with huge computational power comparable to multiple GPUs and seemingly unlimited memory capacity for a developer to develop an AIML model for application. On the other hand, these cloud platforms delivering intelligence services would serve as the intelligence infrastructures for the development of higher level of intelligence application systems, as shown in Figure 3.

Today, we have a lot more emerging technologies available in the *information infrastructure*. They include the personal area network (PAN), high speed wireless communication, Internet of Things (IoT), Internet of Vehicles (IoV), global positioning systems, mobile devices (smart phones, pads, watches and wearable devices), virtual reality (VR) headsets, augmented reality (AR) headsets like Microsoft Hololen, 5G communication technologies, cloud platforms and others.

The services delivered on top of this information infrastructure, like Google Map and Facebook, would definitely be facilitate the development of intelligent services to be added to the *intelligence infrastructure*. The pace of development could be far higher speed then ever.

6.7 Real World Applications

With the aforementioned technologies and other advanced technologies, a number of real world applications have been on the move. In a blog, Aayushi Johari has introduced 10 real world AI applications[8] in the areas of (1) marketing, (2) banking, (3) finance, (4) agriculture, (5) health care, (6) gaming, (7) space exploration, (8) autonomous vehicles, (9) chatbots and (10) artificial creativity.

[8] `https://www.edureka.co/blog/artificial-intelligence-applications/`

Except that, AI had been successfully applied in automated mail-sorting machines in the US Postal Office. The key technology being applied is a neural network for optical character recognition which was developed by Yann LeCun and his collaborators in AT&T Bell Lab [9]. AI programs have been developed and applied in the US legal system [22].

AI programs have now been applied in scientific researches. Before 2000, AI program had been developed to read articles and then generate a summary for the articles [23]. Today, this technology has been even advanced. AI programs have been developed to read thousands of articles and generate a monograph [24]. These technologies could help a researcher to spend more time on the solution of a research rather than reading research articles. Even more, some AI programs are able to make hypotheses from the articles read [1]. Researchers could thus select from the set of hypotheses a few hypotheses for laboratory researches.

6.8 Miscellaneous

In the past decade, many philosophical issues have been raised about the uses of AI systems. Two notable issues are *AI safety* and *Explainable AI*.

6.8.1 AI Safety

The issue on *AI safety* is in essence the issue on *technology safety*. While Tesla object recognition system is able to recognize an object with 99.99% per image, its safety is still questionable. It is claimed that a Tesla car will capture 2000 frames per second[10]. With the accuracy 99.99%, erroneous recognition could be larger than $1-0.9999^{2000} = 0.1813$. In other words, Tesla could commit an error with error-rate 0.1813 in a second. If a Tesla auto-driving car has been moving on the road for an hour, the probability that the car will have an accidence is not small. Whatever car accidence might cause casualty. Thus, *AI safety* is currently an important issue being discussed.

6.8.2 Explainable AI

The issue on *Explainable AI* is focused on the models and algorithms developed for an AI system. Many AI systems are complicated. Their models are a combination of various AI (or machine learning) models. Their successes are based on a lot of fine-tunes and trial-and-errors. In the end, their successes can hardly explained by the underlying models and their learning algorithms. This causes an important problem. *If an AI system is able to solve a problem, could we explain its success? If an AI system fails to solve a problem, could we explain why?* Clearly, even the researchers in the development of the AI systems are unable to give an explanation. Without an explanation, it would lead to a hesitation on the use of an AI system. Nevertheless, researchers of those AI

[9] `http://yann.lecun.com/ex/research/index.html`.
[10] `https://en.wikipedia.org/wiki/Tesla_Autopilot`.

systems would have to rely on trial-and-error to get an not-that-bad system for use. This will violate to what a scientific research should be.

7 Questions

1. State your definition of *intelligence*. Note that the condition(s) included in your definition must be observable and measurable. For instance, *a person is intelligence if it is alive* or *a substance is intelligence if it is dead*.

2. In accordance with your definition, justify with reason(s) if the following object(s) is(are) intelligent.

   (a) You are intelligent.

   (b) A baby is intelligent.

   (c) A dog is intelligent.

   (d) An amoebae is intelligent.

   (e) A tree is intelligent.

   (f) A rock is intelligent.

   (g) AlphaGO is intelligent.

3. State your definition of *artificial intelligence*.

References

[1] S. Spangler, A. D. Wilkins, B. J. Bachman, M. Nagarajan, T. Dayaram, P. Haas, S. Regenbogen, C. R. Pickering, A. Comer, J. N. Myers *et al.*, "Automated hypothesis generation based on mining scientific literature," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 1877–1886.

[2] Y. Gil, D. Garijo, V. Ratnakar, R. Mayani, R. Adusumilli, H. Boyce, and P. Mallick, "Automated hypothesis testing with large scientific data repositories," in *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems (ACS)*, vol. 2, 2016, p. 4.

[3] J. Sybrandt, M. Shtutman, and I. Safro, "Moliere: Automatic biomedical hypothesis generation system," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 1633–1642.

[4] S. J. Wilson, A. D. Wilkins, M. V. Holt, B. K. Choi, D. Konecki, C.-H. Lin, A. Koire, Y. Chen, S.-Y. Kim, Y. Wang *et al.*, "Automated literature mining and hypothesis generation through a network of medical subject headings," *BioRxiv*, p. 403667, 2018.

[5] D. Garijo, S. Fakhraei, V. Ratnakar, Q. Yang, H. Endrias, Y. Ma, R. Wang, M. Bornstein, J. Bright, Y. Gil *et al.*, "Towards automated hypothesis testing in neuroscience," in *Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB 2019 Workshops, Poly and DMAH, Los Angeles, CA, USA, August 30, 2019, Revised Selected Papers 5*. Springer, 2019, pp. 249–257.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[8] ——, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[9] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," *arXiv preprint arXiv:1612.07828*, 2016.

[10] J. Piaget, *The Psychology of Intelligence*. Routledge and Kegan Paul, 1950.

[11] M. Minsky, "Steps toward artificial intelligence," *Proceedings of the IRE*, vol. 49, no. 1, pp. 8–30, 1961.

[12] J. S. Albus, "Outline for a theory of intelligence," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 473–509, 1991.

[13] S. Legg, M. Hutter *et al.*, "A collection of definitions of intelligence," *Frontiers in Artificial Intelligence and applications*, vol. 157, p. 17, 2007.

[14] M. Minsky, *Society of Mind*. Simon and Schuster, 1988.

[15] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[16] A. M. Turing, "On computable numbers, with an application to the entscheidungsproblem," *Proceedings of the London Mathematical Society*, vol. s2-42, no. 1, pp. 230–265, 1937.

[17] ——, "On computable numbers, with an application to the entscheidungsproblem. A correction," *Proceedings of the London Mathematical Society*, vol. s2-43, no. 1, pp. 544–546, 1938.

[18] ——, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433 – 460, 1950.

[19] R. M. Friedberg, "A learning machine: Part I," *IBM Journal of Research and Development*, vol. 2, no. 1, pp. 2–13, 1958.

[20] N. J. Nilsson, *Learning Machines: Foundations of Trainable Pattern-Classifying Systems*. McGraw-Hill, 1965.

[21] M. I. Jordan, "Artificial Intelligence – The Revolution Hasn't Happened Yet," *Harvard Data Science Review*, vol. 1, no. 1, July 2019. [Online]. Available: https://hdsr.mitpress.mit.edu/pub/wot7mkc1

[22] B. Alarie, A. Niblett, and A. H. Yoon, "How artificial intelligence will affect the practice of law," *University of Toronto Law Journal*, vol. 68, no. supplement 1, pp. 106–124, 2018.

[23] G. Salton, A. Singhal, M. Mitra, and C. Buckley, "Automatic text structuring and summarization," *Information Processing & Management*, vol. 33, no. 2, pp. 193–207, 1997.

[24] B. Writer, *Lithium-Ion Batteries : A Machine-Generated Summary of Current Research*. Springer, 2019.