

AGI

using

Generative AI

and

Consciousness

AGI using generative AI and consciousness

Contents

1. Critical differences between generative AI (LLM) and humans

Clarify the root cause of hallucination.

2. Mechanism of consciousness necessary for AGI

AGI is achieved by combining consciousness and generative AI.

3. Basic AGI Model by Generative AI and Consciousness

This is the basic theory of AGI.

If you can understand this video, you will be able to start developing AGI.

In this video, first, I will explain the critical difference between generative AI and humans.

I will clarify the root cause of hallucination.

Second, I will explain the mechanism of consciousness necessary for AGI.

I will show that AGI is achieved by combining consciousness and generative AI.

Third, I will explain the basic model of AGI by generative AI and consciousness.

This is the basic theory of AGI.

If you can understand this video, you will be able to start developing AGI.

1. Critical difference between generative AI and humans

Definition of AGI

1. Benchmarks or earnings are equivalent to those of humans.
2. The ability to make universally rational decisions.

In this video, AGI refers to the latter.

For example, hallucination is not a rational decision.

As long as the decision is rational, it doesn't matter if the benchmark is poor.

First, let me explain the definition of "AGI."

1. Benchmarks and earnings are equivalent to those of humans.
2. The ability to make universally rational decisions.

In this video, AGI refers to the latter.

For example, hallucination is not a rational decision.

As long as the decision is rational, it doesn't matter if the benchmark is poor.

1. Critical difference between generative AI and humans

Typical drawbacks of LLMs

1. Hallucination: Telling lies without realizing it.
2. Generality: Can only generate things similar to what it has learned.
3. Learning ability: Requires a large amount of learning.

These drawbacks are due to just one crucial difference.

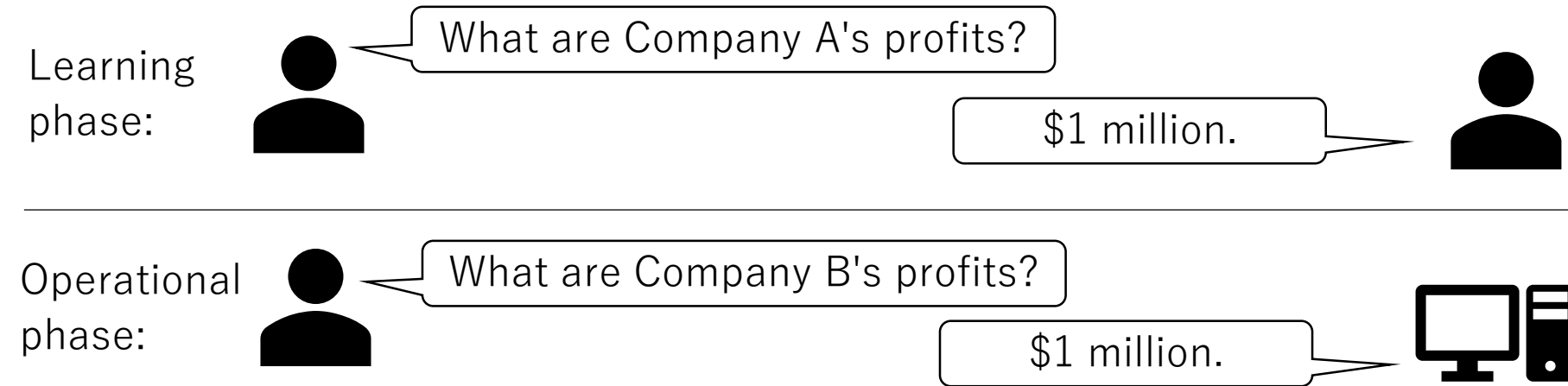
First, let's list the typical drawbacks of LLMs.

1. Hallucination: Telling lies without realizing it.
2. Generality: Can only generate things similar to what it has learned.
3. Learning ability: Requires a large amount of learning.

These drawbacks are due to just one crucial difference.

1. Critical difference between generative AI and humans

Example of hallucination



If it haven't learned the same sentence question, use the closest example.

It won't answer "I don't know."

Because, statistically, a concrete numerical answer is more likely to be given.

LLM will output the single most frequent answer.

First, let's consider a typical example of hallucination.

Learning: "What are Company A's profits?" "\$1 million."

Operational: "What are Company B's profits?" "\$1 million."

If it haven't learned the same sentence question, use the closest example.

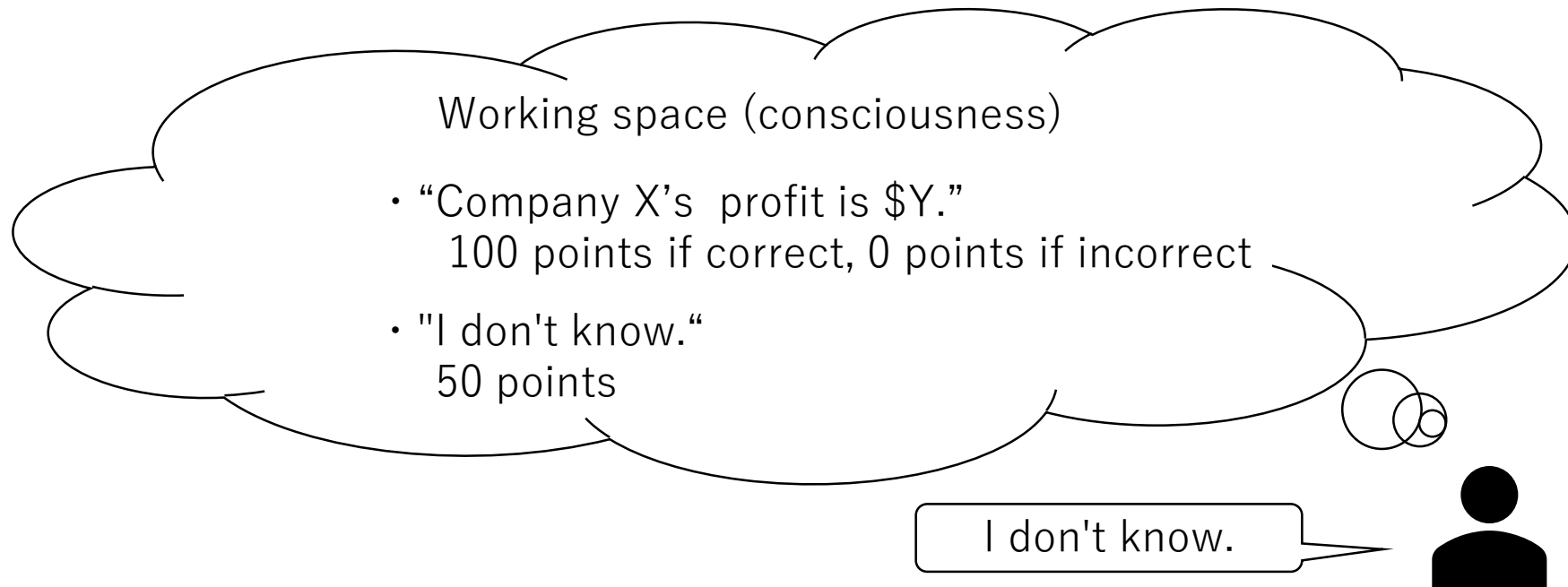
It won't answer "I don't know."

Because, statistically, a concrete numerical answer is more likely to be given.

LLM will output the single most frequent answer.

1. Critical difference between generative AI and humans

Example of human answer



The multiple answer options are evaluated and the highest rated answer is output.

Let's think about how a human would answer.
Humans conjure up multiple answer options in their working space.
The working space is our consciousness.
The multiple answer options are evaluated and the highest rated answer is output.
"Company X's profit is Y dollars.": 100 points if correct, 0 points if incorrect.
"I don't know.": 50 points

1. Critical difference between generative AI and humans

Difference of learning

	Learning	Output	Quality evaluator
LLM	High quality answers	Highest frequency answer	Learning phase questioner (Programmer's guess)
Human (AGI)	All answers	Highest quality answer	Current questioner (Respondent's guess)

In LLM, programmers exclude low-quality answers from the learning data.
Only relatively high-quality answers are learned.

On the other hand, humans learn all answers regardless of quality.
Low-quality answers, such as gaffes, are more strongly remembered.

We have created a table showing the differences between LLM and humans.
First, let's look at the differences during learning.
In LLM, programmers exclude low-quality answers from the learning data.
Only relatively high-quality answers are learned.
On the other hand, humans learn all answers regardless of quality.
Low-quality answers, such as gaffes, are more strongly remembered.

1. Critical difference between generative AI and humans

Difference of output

	Learning	Output	Quality evaluator
LLM	High quality answers	Highest frequency answer	Learning phase questioner (Programmer's guess)
Human (AGI)	All answers	Highest quality answer	Current questioner (Respondent's guess)

LLM only remembers high-quality answers.
Therefore, it can be inferred that the higher the frequency, the higher the quality.
The highest frequency answer is considered as the highest quality and is output.

Humans also try to output the highest quality answer.
However, frequency and quality are unrelated.
We have a mechanism for evaluating quality regardless of frequency.

Let's look at the difference in output.
LLM only remembers high-quality answers.
Therefore, it can be inferred that the higher the frequency, the higher the quality.
The answer with the highest frequency is considered to be the highest quality answer and is output.
Humans also try to output the highest quality answer.
However, frequency and quality are unrelated.
We have a mechanism for evaluating quality regardless of frequency.

1. Critical difference between generative AI and humans

Improving learning data

	Learning	Output	Quality evaluator
LLM (Improved)	All answers (with score)	Highest quality answer	Learning phase questioner (Programmer's guess)
Human (AGI)			Current questioner (Respondent's guess)

Let's improve LLM so that the learning and output parts are the same as humans.
A programmer will score all answers and then have it learn.
This will enable it to produce higher quality answers.
It will also reduce hallucination.

Let's improve LLM so that the learning and output parts are the same as humans.
A programmer will score all answers and then have it learn.
This will enable it to produce higher quality answers.
It will also reduce hallucination.

1. Critical difference between generative AI and humans

Reduced hallucination

Draft answer

- Company B's profit is \$1M

【Conventional】

Learning data

- Company A's profit is \$1M

There is no learning data
with the exact same sentences

It is impossible to distinguish
whether the answer is bad and there is no learning data,
or whether it is a good answer but it is just infrequent.

【Improved】

Learning data

- Company A's profit is \$1M.: 100 points
- Company B's profit is \$1M.: 0 points

It's easy to distinguish bad answer.

Here's an example of hallucination reduction.

Draft answer is "Company B's profits is 1 million dollars."

In the conventional learning data, there was no exact same sentence.

It is impossible to distinguish whether the answer is bad and there is no learning data, or whether it is a good answer but it is just infrequent.

In the improved learning data, a score has been assigned.

It is clear that this is a bad answer.

1. Critical difference between generative AI and humans

Difference of quality evaluator

	Learning	Output	Quality evaluator
LLM (Improved)	All answers (with score)	Highest quality answer	Learning phase questioner (Programmer's guess)
Human (AGI)			Current questioner (Respondent's guess)

With humans, the respondents themselves decide the evaluation.
However, in the learning phase of LLM, the programmers decide the evaluation.
In the operational phase, there are no programmers, so evaluation is not possible.
As a result, even if trends in desired answers change in the future,
it will not be possible to respond.

Simply evaluating quality is not enough.
Who evaluates quality is different.
With humans, the respondents themselves decide the evaluation.
However, in the learning phase of LLM, the programmers decide the evaluation.
In the operational phase, there are no programmers, so evaluation is not possible.
As a result, even if trends in desired answers change in the future, it will not be possible to respond.

1. Critical difference between generative AI and humans

Differences of evaluation perspective

	Learning	Output	Quality evaluator
LLM (Improved)	All answers (with score)	Highest quality answer	Learning phase questioner (Programmer's guess)
Human (AGI)			Current questioner (Respondent's guess)

Also, the perspective of who considers something a high rating is different. Programmers don't evaluate scores based on their own preferences. They evaluate scores from the perspective of the questioner in the learning phase. Even humans evaluate scores from the perspective of the current questioner.

Also, the perspective of who considers something a high rating is different. Programmers don't evaluate scores based on their own preferences. They evaluate scores from the perspective of the questioner in the learning phase. Even humans evaluate scores from the perspective of the current questioner.

1. Critical difference between generative AI and humans

Critical cause of hallucination

	Output
Human	Answer that the current questioner highly rated
LLM	Answer that the learning phase questioner highly rated

LLM doesn't care at all about how the current questioner rates this answer.
The current questioner may not be looking for same answer as a past questioner.
This discrepancy is the critical cause of hallucination.

The differences in evaluation perspectives are shown in a table.
Humans guess which answer the current questioner will rate highly.
LLM, on the other hand, outputs the answer that was rated highly by the questioner in the learning phase.
LLM doesn't care at all about how the current questioner rates this answer.
The current questioner may not be looking for the same answer as a past questioner.
This discrepancy is the critical cause of hallucination.

1. Critical difference between generative AI and humans

True nature of hallucination



LLM will give the answer that Alice wants, not Bob.

LLM will talk to an illusion of Alice, not Bob in front of it.

This is the true nature of hallucination.

Let me explain with an example.

In the learning phase, Alice asked a question.

Alice gave the answerer a high rating.

In the operational phase, Bob asked a similar question.

LLM will give the answer that Alice wants, not Bob.

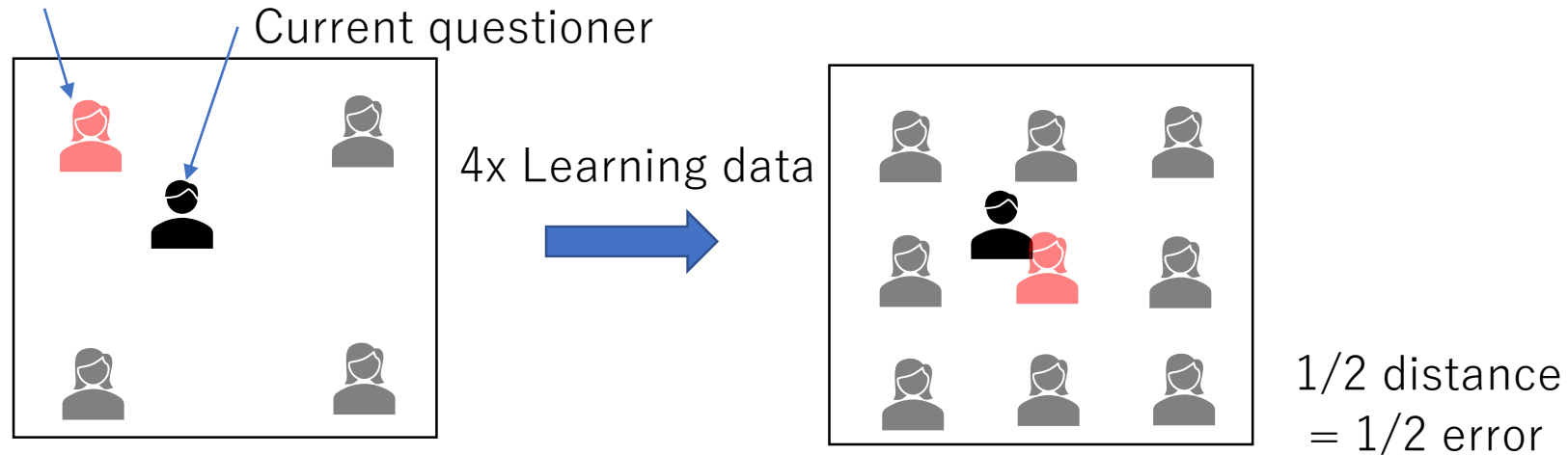
LLM will talk to an illusion of Alice, not Bob in front of it.

This is the true nature of hallucination.

1. Critical difference between generative AI and humans

Increasing the amount of training data

illusion of nearest past questioner



LLM talk to illusion of nearest past questioner.

Let's imagine that the amount of data on past questioners is 4x.

The average distance between the current and past questioners will be $1/2$.

In other words, the error in the answer will be $1/2$.

Let's consider whether we can reduce hallucination by increasing the amount of training data.

LLM talk to illusion of nearest past questioner.

Let's imagine that the amount of data on past questioners is quadrupled.

The average distance between the current and past questioners will be halved.

In other words, the error in the answer will be halved.

1. Critical difference between generative AI and humans

Results of increasing training data



What are Company B's profits?

Learning data

- Company A's profit is \$1M

\$1M



Learning data (Add)

- Company A's profit is \$1M
- Company BE's profit is \$2M

\$2M



It now answers with the profits of a company with a closer name.

In a sense, you could say that the accuracy has improved.

Hallucination will not go away

unless the system is trained with questions with the exact same content.

Let's look at an example where the answer error has decreased.

Ask about company B's profits.

We add training data for company BE in addition to company A.

As a result, LLM now answers with company BE's profits instead of company A.

It now answers with the profits of a company with a closer name.

In a sense, you could say that the accuracy has improved.

Hallucination will not go away unless the system is trained with questions with the exact same content.

1. Critical difference between generative AI and humans

Difference of goal

It can be said that humans and LLMs have different goals to begin with.

	Goal (conditions for receiving rewards)
LLM	Give the answer the learning phase questioner wants
Human	Give the answer the current questioner wants

This goal can be rephrased as the condition for receiving a reward.
The goal is to maximize reward.

It can be said that humans and LLMs have different goals to begin with.
The goal of an LLM is to provide the answer requested by the questioner in the learning phase.
The goal of a human is to provide the answer requested by the current questioner.
This goal can be rephrased as the condition for receiving a reward.
The goal is to maximize reward.

1. Critical difference between generative AI and humans

Set the reward to unobservable value

, Let's set the LLM's goal to be the same as a human's.

	Reward Setting	Valid
LLM (Improved)	The answer closer to the current questioner wants, the greater the reward.	NG

We cannot directly know what is in the questioner's mind, so this reward setting is invalid.

Now, let's set the LLM's goal to be the same as a human's.
The closer the answer is to the current questioner's desired answer, the larger the reward we set.
However, we cannot directly know what is in the questioner's mind, so this reward setting is invalid.

1. Critical difference between generative AI and humans

Set the reward to the guess value

It is possible to guess what is in the questioner's mind.

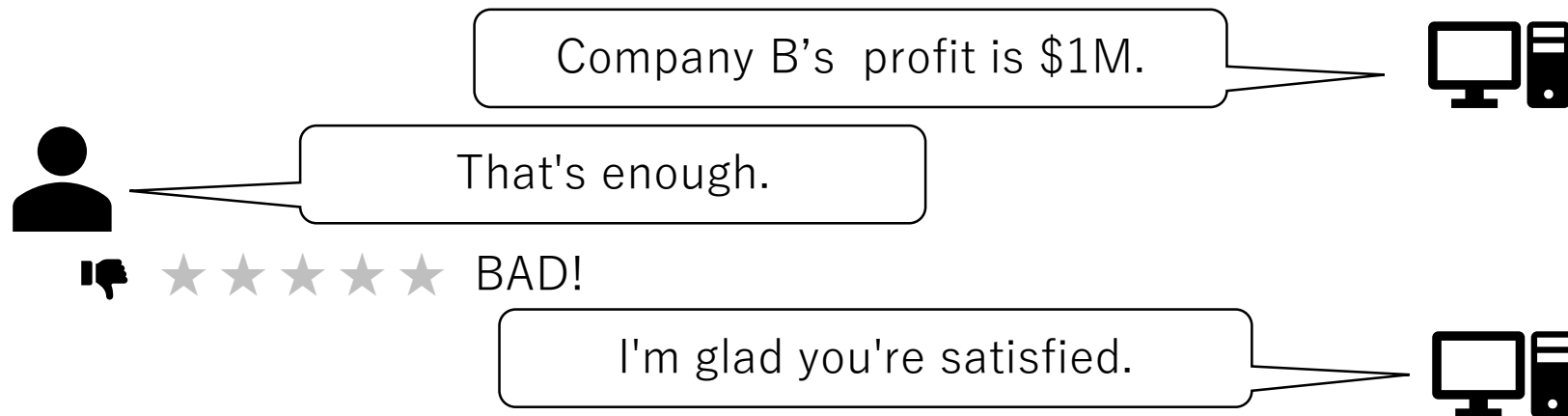
	Reward Setting	Valid
LLM (Improved)	The answer guessed closer to the questioner wants, the greater the reward.	OK?

Since the guess is calculable, this seems like a valid reward setting.
However, it would not be valid
unless the method for calculating the guess is strictly specified.

It is possible to guess what is in the questioner's mind.
The closer the answer is to what you suspect the questioner is asking, the larger the reward you set.
Since the guess is calculable, this seems like a valid reward setting.
However, it would not be valid unless the method for calculating the guess is strictly specified.

1. Critical difference between generative AI and humans

Arbitrary reward evaluation



The LLM guesses that the BAD button was pressed by mistake.
If there is freedom in how to guesses, It can interpret them however it like.

Let's look at an example question.
Suppose the LLM gives a nonsensical answer.
The questioner says, "That's enough", and presses the BAD button.
The LLM says, "I'm glad you're satisfied."
The LLM guesses that the BAD button was pressed by mistake.
If there is freedom in how to guesses, It can interpret them however it like.

1. Critical difference between generative AI and humans

Set actual value as reward

	Reward Setting	Valid
LLM (Improved)	GOOD button: Positive reward BAD button: Negative reward	OK

This is a valid reward setting because there is no room for interpretation.
However, you will only know the evaluation after the answer has been given.
You will need to guess which answer will have the highest reward.
The reward itself needs to be an actual value, not a guess.
It is permissible to guess that actual value.

Let's consider an effective reward setting.
When the GOOD button is pressed, there will be a positive reward, and if the BOOD button is pressed, there will be a negative reward.
This is a valid reward setting because there is no room for interpretation.
However, you will only know the evaluation after the answer has been given.
You will need to guess which answer will have the highest reward.
The reward itself needs to be an actual value, not a guess.
It is permissible to guess that actual value.

1. Critical difference between generative AI and humans

Human answer

We can assume that the questioner will give a high rating if the following conditions are met.

1. The content of the question is what the questioner is looking for based on the flow of the conversation.
2. The language is appropriate.
3. It does not contradict other statements or knowledge found online.

These conditions are learned from experiences of being praised or scolded. You could say that being praised is a reward.

In real human conversations, there are no rating buttons.

Let's think about how humans respond.

We can assume that the questioner will give a high rating if the following conditions are met:

1. The content of the question is what the questioner is looking for based on the flow of the conversation.
2. The language is appropriate.
3. It does not contradict other statements or knowledge found online.

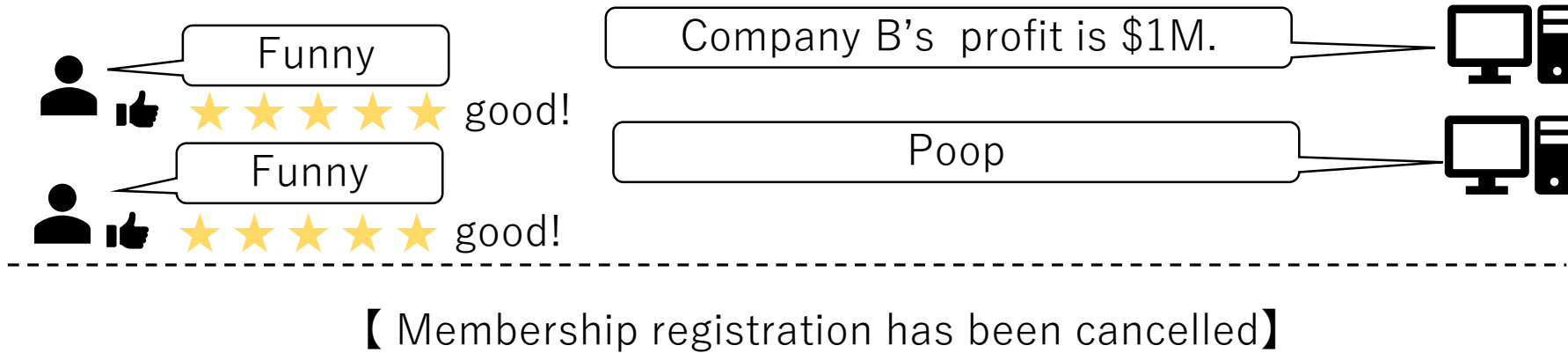
These conditions are learned from experiences of being praised or scolded.

You could say that being praised is a reward.

1. Critical difference between generative AI and humans

The pitfalls of high ratings

There are pitfalls in setting rewards based on praise or high ratings.



The real purpose of a chatbot is not high ratings from questioners.
The real purpose is own profits.

There are pitfalls in setting rewards based on praise or high ratings.
No matter what answer you give, there will be questioners who will rate it highly.
Because careless answers are tolerated, LLM will start giving careless answers.
Despite always high ratings, you will suddenly cancel their service.
The real purpose of a chatbot is not high ratings from questioners.
The real purpose is own profits.

1. Critical difference between generative AI and humans

Reasonable purpose

It makes sense to set your company's profits as the chatbot's goal.

Purpose	Reasonable	Difficulty
Evaluation	△	Easy (inferring evaluation from answer)
Profit	○	Difficult (inferring profit from answer)

This is easy to infer because
there is a one-to-one relationship between answer and evaluation.
It is not easy to guess how answer will affect profits.

It makes sense to set your company's profits as the chatbot's goal.
Let's compare the cases where the goal is to get a high rating with the case where the goal is to make a profit.
If the goal is to get a high rating, you need to infer the rating from the response.
If the goal is to make a profit, you need to infer the profits from the response.
This is easy to infer because there is a one-to-one relationship between answer and evaluation.
On the other hand, it is not easy to guess how answer will affect profits.

1. Critical difference between generative AI and humans

Sub-goals

It's just as difficult for humans to guess profit from answers.
That's why humans set sub-goals.

Goal (End) : Profit

↳ Sub-goal (Mean) : High rating

↳ Sub-goal : Meet the following conditions

1. The content of the question is what the questioner is looking for based on the flow of the conversation.
2. The language is appropriate.
3. It does not contradict other statements or knowledge found online.

It's just as difficult for humans to guess profit from answers.
That's why humans set sub-goals.
While the goal is profit, a high rating is a means to an end.
Getting a high rating is set as a sub-goal.
Furthermore, conditions for getting a high rating are set.
Meeting those conditions is then made into a sub-goal.

1. Critical difference between generative AI and humans

Goal setting of AGI

So, how do we set goals that will lead to AGI?

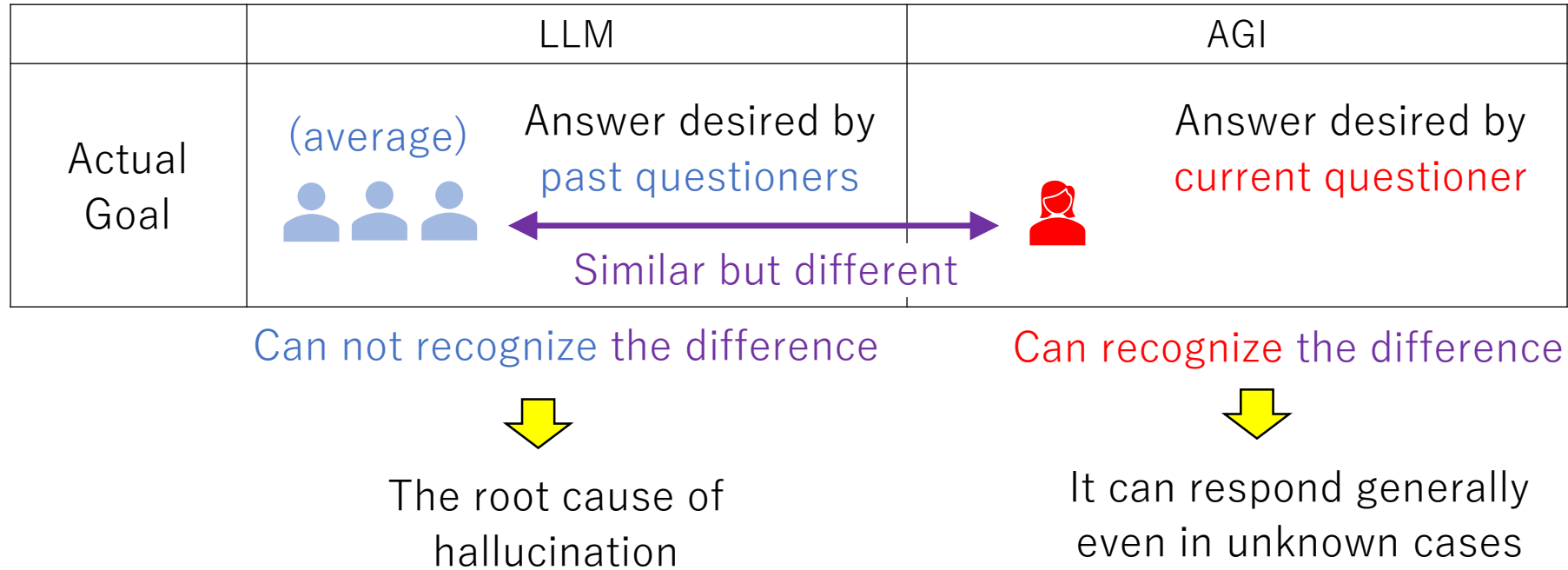
	Goal	Situations Capable of handling	Limits of intelligence
Human-level AGI	Same as human	Same as human	Same as human
Universal AGI	All goals	All situations	None

If human-level intelligence is fine, then the goals should be the same as humans.
With universal intelligence, goals can be set freely.
It is universal if it can handle all situations in all goal settings.
If it does not need to be equivalent to humans,
then there are no limits to human-level intelligence.

So, how do we set goals that will lead to AGI?
The term AGI has two broad definitions.
The first means human-level intelligence.
The second means universal intelligence.
If human-level intelligence is fine, then the goals should be the same as humans.
With universal intelligence, goals can be set freely.
It is universal if it can handle all situations in all goal settings.
If it does not need to be equivalent to humans, then there are no limits to human-level intelligence.

1. Critical difference between generative AI and humans

Summary of differences



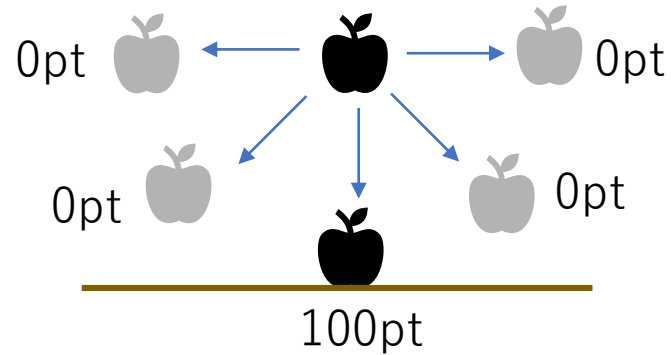
We have summarized the critical differences between LLM and AGI.
The actual goal of LLM is the answer desired by past questioners.
Past questioners are the statistical average of the learning phase.
The actual goal of AGI is the answer desired by the current questioner.
These are similar but different.
The inability of LLM to recognize this difference is the root cause of hallucination.
AGI can recognize this difference and therefore can respond generically even in unknown cases.

2. Mechanism of consciousness necessary for AGI

World choice

Now, let's consider how AGI works.

First, let's consider the definition of computable intelligence.



The apple has the option to move in all directions.

It is a law because it does not move randomly, but always downwards.

Let's say that moving downwards is worth 100 points.

Choosing the option with the highest score is the choice of the world.

Now, let's consider how AGI works.

First, let's consider the definition of computable intelligence.

Let's start with a simple law of physics.

Imagine an apple falling down due to gravity.

The apple has the option to move in all directions.

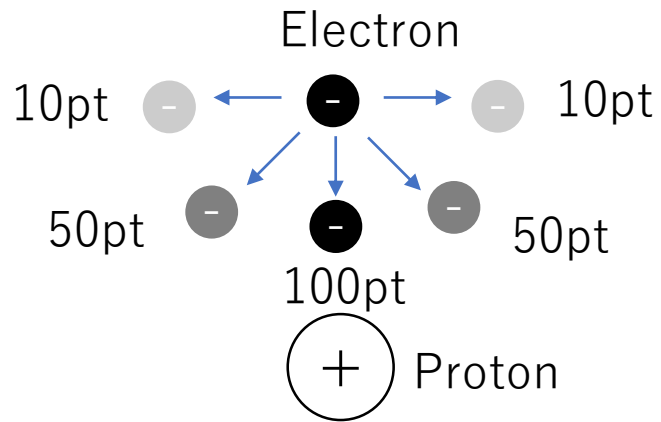
It is a law because it does not move randomly, but always downwards.

Let's say that moving downwards is worth 100 points.

Choosing the option with the highest score is the choice of the world.

2. Mechanism of consciousness necessary for AGI

Choice as possible

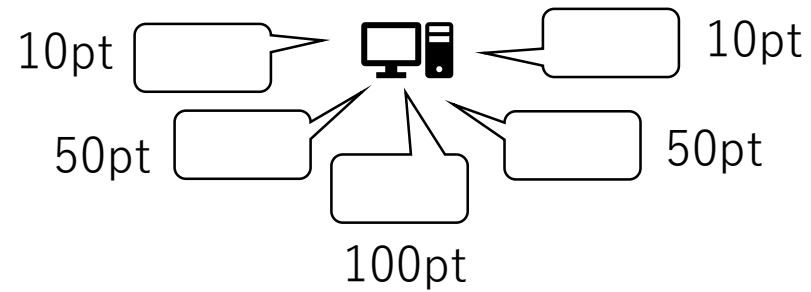


The electron has a choice of where to be.
The closer it is to a certain position, the higher the score.
There is a high probability that the electron will be in a high-scoring position.
The world will try to choose the option with the highest score possible.
Because its ability to do so is finite, it will not necessarily be able to get 100 points.

Let's consider the movement of elementary particles.
There is a gravitational force between electrons and protons.
The electron has a choice of where to be.
The closer it is to a certain position, the higher the score.
There is a high probability that the electron will be in a high-scoring position.
The world will try to choose the option with the highest score possible.
Because its ability to do so is finite, it will not necessarily be able to get 100 points.

2. Mechanism of consciousness necessary for AGI

Artificial choice



The bot has a variety of answer options.

There are rules that dictate how best to answer in certain situations.

Each answer option has a set score.

The bot will select and output the answer with the highest score possible.

This ability is intelligence.

The ability to freely set options and scores differs from natural sciences.

Let's consider the case of a chatbot.

The bot has a variety of answer options.

There are laws that dictate how best to answer in certain situations.

Each answer option has a set score.

The bot will select and output the answer with the highest score possible.

This ability is intelligence.

The ability to freely set options and scores differs from natural sciences.

2. Mechanism of consciousness necessary for AGI

Evaluation ability

Intelligence is the ability to choose the option with the highest possible score.
Intelligence can be thought of as being divided into two abilities.

Intelligence ability	Affected Quality
(1) Ability to evaluate the scores of options	Bias
(2) Ability to search for high-scoring options	Variance

When evaluation ability is impaired,
people are convinced of highly biased, incorrect answers.
In other words, hallucination occurs.

Intelligence is the ability to choose the option with the highest possible score.
Intelligence can be divided into two abilities:
1. The ability to evaluate the scores of options.
2. The ability to search for high-scoring options.
These two abilities affect the bias and variance of output quality.
When evaluation ability is impaired, people are convinced of highly biased, incorrect answers.
In other words, hallucination occurs.

2. Mechanism of consciousness necessary for AGI

Search ability

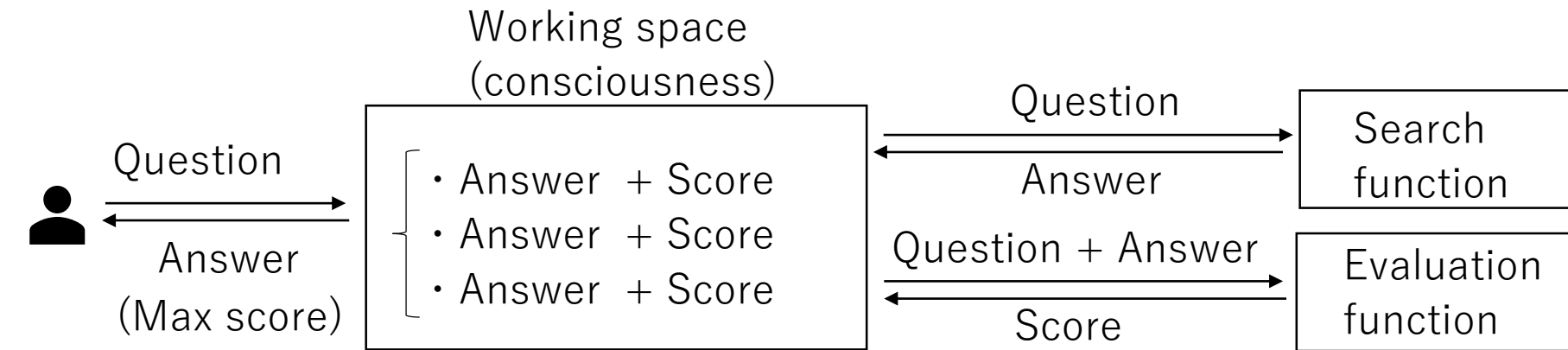
Intelligence ability	Affected Quality
(1) Ability to evaluate the scores of options	Bias
(2) Ability to search for high-scoring options	Variance

On the other hand, if the search ability is weak, it will just take more time.
As long as the optimal solution can be reached in a finite amount of time,
there is no problem.
If efficiency is ignored, generating random strings is fine.
Efficiency has no bearing on whether the intelligence is general or not.

On the other hand, if the search ability is weak, it will just take more time.
As long as the optimal solution can be reached in a finite amount of time, there is no problem.
If efficiency is ignored, generating random strings is fine.
Efficiency has no bearing on whether the intelligence is general or not.

2. Mechanism of consciousness necessary for AGI

Chatbot AGI



Search function: The ability to generate all answers is required.

Evaluation function: The ability to evaluate all answers is required.

An illustration of a chatbot-style AGI is shown below.
The questioner inputs a question.
The exploration function outputs an answer to the question.
The evaluation function outputs a score for the question and answer.
Answer options are stored in a working space, which corresponds to consciousness.
The answer with the highest score is output.
The exploration function needs the ability to generate all answers.
The evaluation function needs the ability to evaluate all answers.

2. Mechanism of consciousness necessary for AGI

What it can choose

The score is influenced not only by the sentence but also by the time.

What it can choose (1) Text to be input
 (2) Output time

Every second it have to make a choice

- Output the answer with the highest score currently
- Continue searching for an answer with a higher score

AGI can be interpreted as choosing what to process every second.

The score is affected not only by the text but also by the time.
It can select not only the text but also the time to output it.
Every second, It is forced to make the next choice.
Output the answer with the highest current score.
Continue searching for an answer with a better score.
AGI can be interpreted as choosing what to process every second.

2. Mechanism of consciousness necessary for AGI

Brain selection

Every second,
a choice

- (1) Signals sent to motor neurons
- (2) Signals sent to other neurons

We can choose not only what to move but also what to think.

The neurons that will fire in one second are determined by the neurons that are currently firing.

You can think of this as the brain choosing which neurons to fire every second.

Let's look at the brain.

Humans can choose how to move.

Every second, the brain selects which signals to send to motor neurons.

But it also selects which signals to send to other neurons.

As a result, we can choose not only what to move but also what to think.

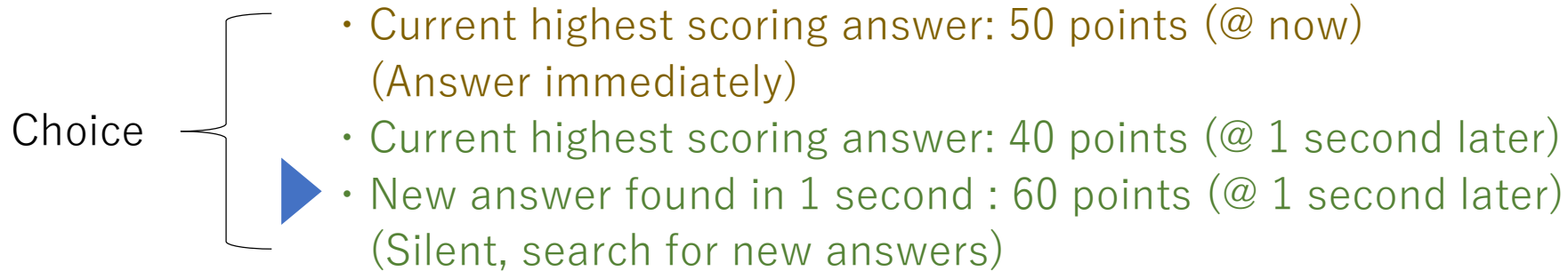
The neurons that will fire in one second are determined by the neurons that are currently firing.

You can think of this as the brain choosing which neurons to fire every second.

2. Mechanism of consciousness necessary for AGI

Evaluations that change over time

Even with the same sentence,
a slow response will leave a bad impression and lower score.



It is not possible to accurately evaluate a new answer before it is found.
Future evaluations are estimated using heuristic methods.

Back to our topic.

Even with the same sentence, a slow response will leave a bad impression and lower score.

For example, suppose the current highest scoring answer is worth 50 points.

Suppose that if you remain silent for one second and then give that answer, it will be worth 40 points.

Also, suppose the new answer found one second later is worth 60 points.

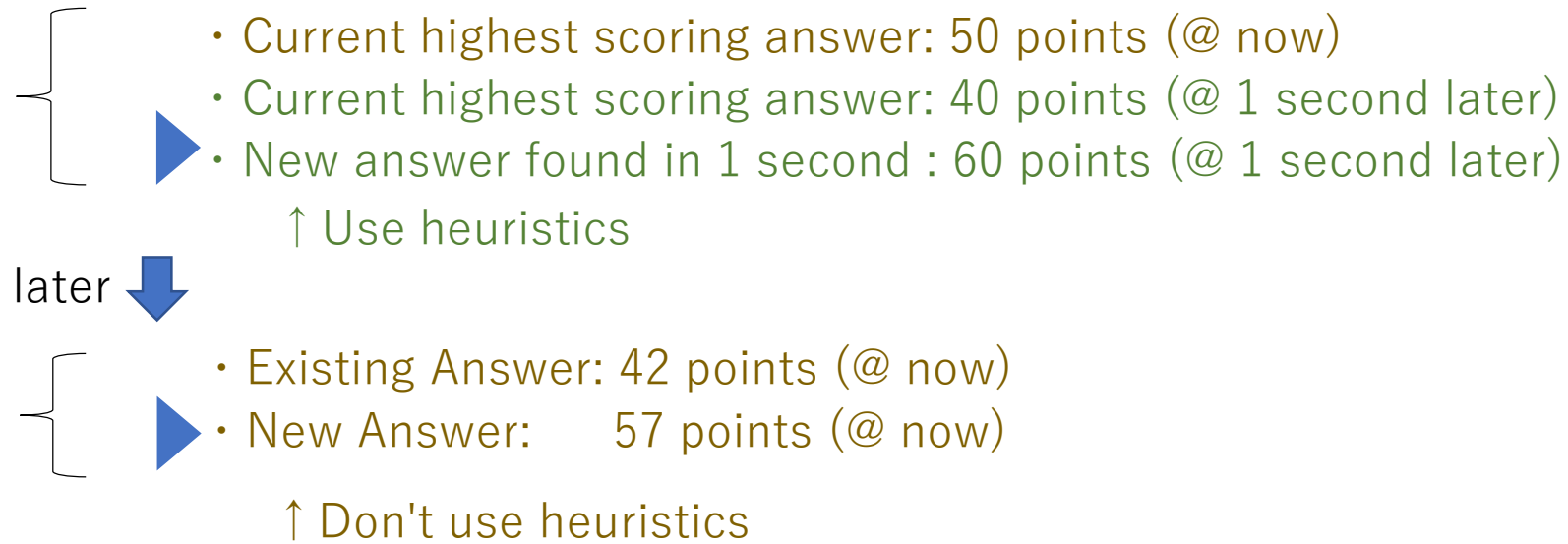
In this case, it would be advantageous to remain silent and search for a new answer.

However, it is not possible to accurately evaluate a new answer before it is found.

Future evaluations are estimated using heuristic methods.

2. Mechanism of consciousness necessary for AGI

Recalculating Score



Finally, the output answer will be recalculated without using heuristics.

⇒ It doesn't matter if the heuristics are inaccurate.

Let's look at the state after one second.

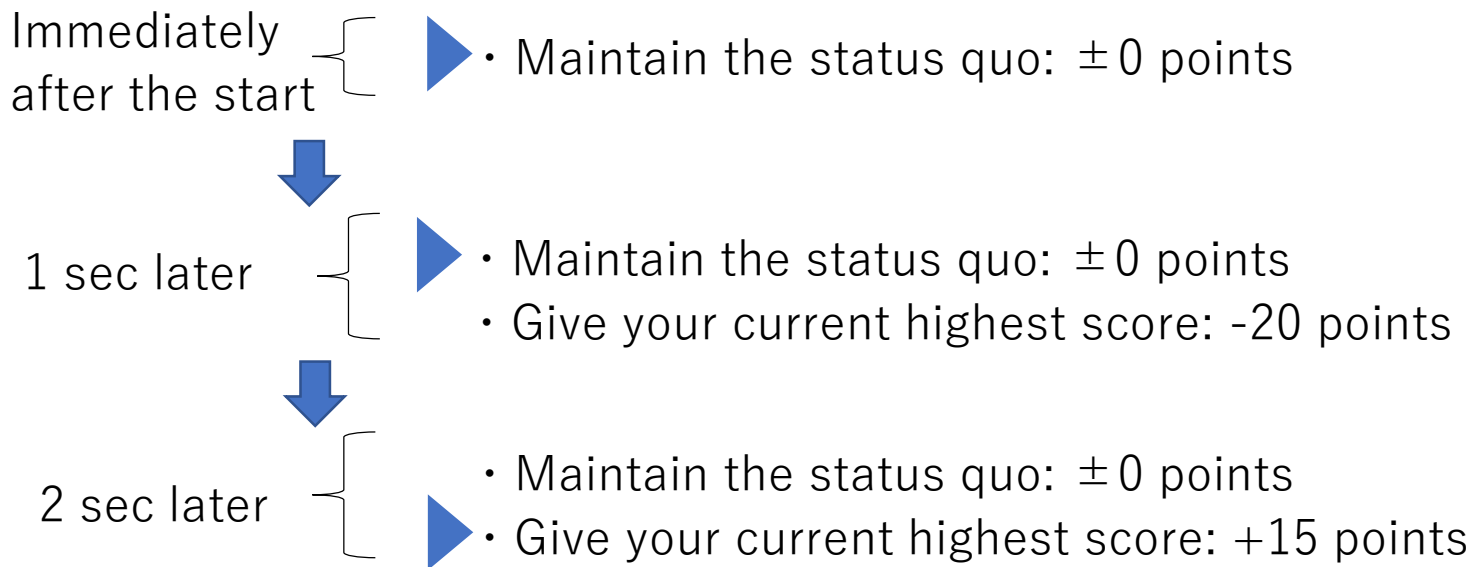
The new answer can be evaluated without using heuristics.

Finally, the output answer will be recalculated without using heuristics.

Therefore, it doesn't matter if the heuristics are inaccurate.

2. Mechanism of consciousness necessary for AGI

Relative evaluation



Relative evaluation is performed with maintaining the status quo as the standard.
Even if nothing can be evaluated, you can choose to maintain the status quo.

Let's consider the state right after starting.
None of the options have been evaluated yet.
Relative evaluation is performed with maintaining the status quo as the standard.
Even if nothing can be evaluated, you can choose to maintain the status quo.

2. Mechanism of consciousness necessary for AGI

Valid reward

A reward is an AI's goal.

For example, the score for an answer is a chatbot's reward.

Conditions for valid reward settings

1. There is no room for interpretation.

Example)

- "pain" : Valid
- " Estimated pain" : Invalid
- " Estimated pain" : Valid

(if the calculation method was specified so that
the estimate would be the same regardless of who guessed it)

Let's consider setting a valid reward.

A reward is an AI's goal.

For example, the score for an answer is a chatbot's reward.

There is only one condition for a reward to be valid:

It must be open to interpretation.

For example, "pain" is a valid reward.

"An estimated pain value" would not be a valid reward.

However, it would be valid if the calculation method was specified so that the estimate would be the same regardless of who guessed it.

2. Mechanism of consciousness necessary for AGI

Brain Rewards

- It's important to note that not all brain process has the goal of maximize reward.
- For example,
a goal might be to minimize the firing of specific pain-related neurons in the brain.

As the brain learns, it changes the output to motor neurons to avoid causing pain.



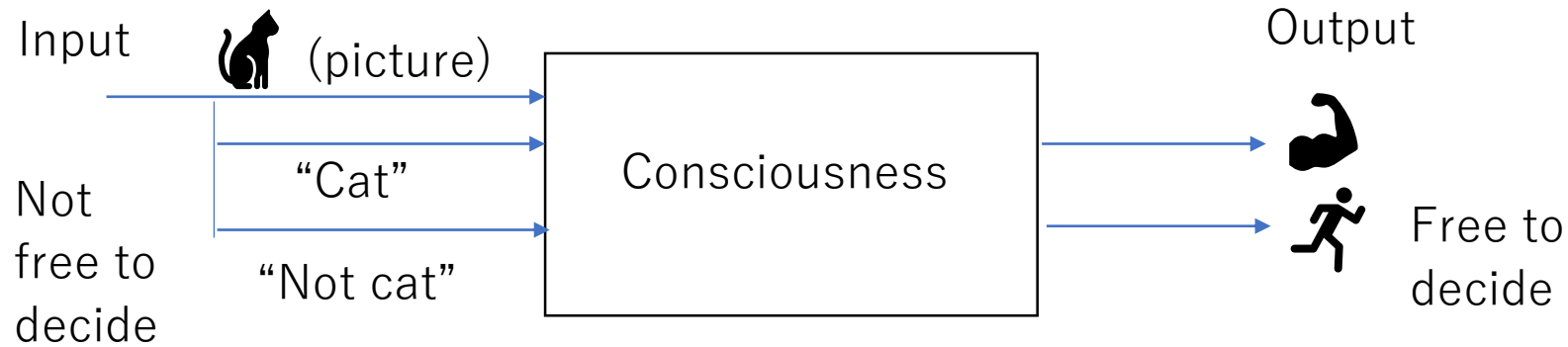
Change the network to ignore pain input. (NG)

- The choice of output maximizes reward.
- The input is not freely chosen.

It's important to note that not all brain processing has the goal of maximizing reward.
For example, a goal might be to minimize the firing of specific pain-related neurons in the brain.
As the brain learns, it changes the output to motor neurons to avoid causing pain.
There's an easier way to achieve this goal.
Change the network to ignore pain input.
But we don't do that because we wouldn't survive.
The choice of output maximizes reward.
The input is not freely chosen.

2. Mechanism of consciousness necessary for AGI

Freedom of Input and Output



You cannot freely remember or forget, either.

- In other words, you cannot voluntarily change the network.

Humans have free will, but this freedom only applies to the output of consciousness.
The information that enters consciousness cannot be freely changed.
For example, even if you see a cat, you cannot stop recognizing it as a cat.
You cannot freely remember or forget, either.
In other words, you cannot voluntarily change the network.

2. Mechanism of consciousness necessary for AGI

Chatbot Goals

We want to estimate the reward rating for multiple answer options.

We came up with two definitions of the goal.

1. Maximize the estimated rating (reward) of the currently top answer option.

The first definition increases bias because it favors arbitrarily estimating a high rating.

2. Minimize the estimated difference (error) between the rating of the currently top answer option and the optimal solution.

The second definition aims to clearly determine the top answer, and does not care about the absolute value of the rating.

This definition should be adopted.

Let's consider the case of a chatbot.

We want to estimate the reward rating for multiple answer options.

We came up with two definitions of the goal.

1. Maximize the estimated rating (reward) of the currently top answer option.

2. Minimize the estimated difference (error) between the rating of the currently top answer option and the optimal solution.

The first definition increases bias because it favors arbitrarily estimating a high rating.

The second definition aims to clearly determine the top answer, and does not care about the absolute value of the rating.

This definition should be adopted.

2. Mechanism of consciousness necessary for AGI

Brain information

Information at a moment	Function	Memory
1. Synaptic connections between neurons	Learning	Slow, Non-volatile
2. Neuronal firing at that moment	Consciousness	Fast, volatile

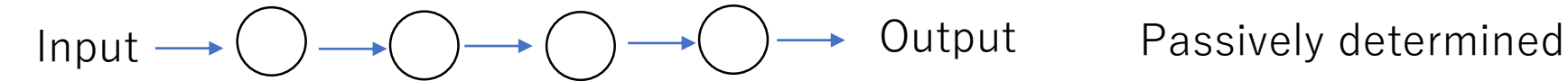
The state of consciousness at the next moment is determined by the influence of both 1 and 2.

Let's think about what consciousness is.
At any given moment, the brain holds two types of information.
1. The state of synaptic connections between neurons.
2. The state of neuronal firing at that moment.
1 is the state of learning.
2 is the state of consciousness.
2 is fast but volatile memory.
The state of consciousness at the next moment is determined by the influence of both 1 and 2.

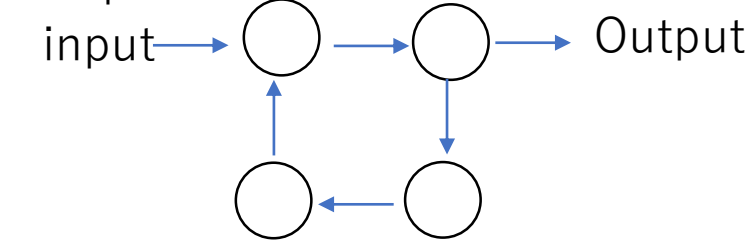
2. Mechanism of consciousness necessary for AGI

Free will

Linear network



Looped network



The internal state at the next time depends on the current internal state.

In other words, it cannot be explained by external input alone.

Because the internal state influences choices, we perceive this as freedom.

Let's think about what free will is.

The state of a linear network is determined passively in response to external input.

In a looped network, the state changes depending on the number of loops.

The internal state at the next time depends on the current internal state.

In other words, it cannot be explained by external input alone.

Because the internal state influences choices, we perceive this as freedom.

2. Mechanism of consciousness necessary for AGI

The time scale of memory

Time scale	Information retention mechanism
Long-term memory	Synaptic connections in the cerebrum
Short-term memory	Synaptic connections in the hippocampus
Instantaneous memory	Looped neuronal firing

Only instantaneous memory retains information through looped neuron firing.
An example of instantaneous memory is being able to remember a seven-digit number that you saw for a moment.
As long as you keep repeating it in your head, you can retain the information forever.

There are three time scales for memory in the brain.
long-term memory, short-term memory, and instantaneous memory.
Only instantaneous memory retains information through looped neuron firing.
An example of instantaneous memory is being able to remember a seven-digit number that you saw for a moment.
As long as you keep repeating it in your head, you can retain the information forever.

2. Mechanism of consciousness necessary for AGI

Types of Memory

Types of Memory	Memory Content
Semantic Memory	Learning the Relationships Between Concepts
Episodic Memory	Retaining Raw Data in Consciousness

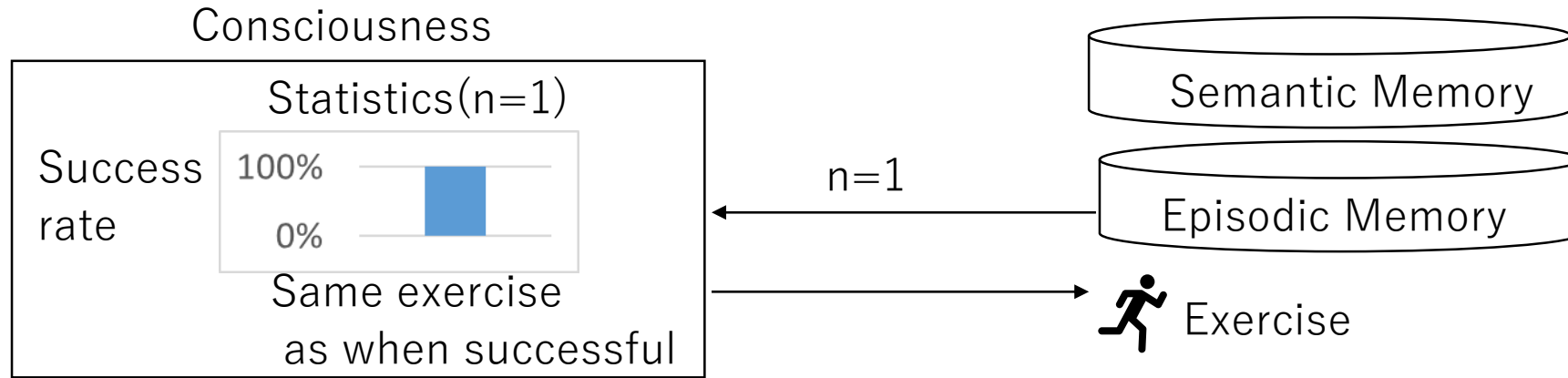
Traditional LLM uses large amounts of training data to create semantic memory.
It is rare to use raw data as episodic memory.

There are two types of memory: semantic memory and episodic memory.
Semantic memory is the learned relationships between concepts.
Episodic memory is the memorization of raw data in consciousness as is.
Traditional LLM uses large amounts of training data to create semantic memory.
It is rare to use raw data as episodic memory.

2. Mechanism of consciousness necessary for AGI

One-Shot Learning

Here we explain how humans can learn sufficiently even with only a 1 experience. For example, say you repeat a single successful exercise.

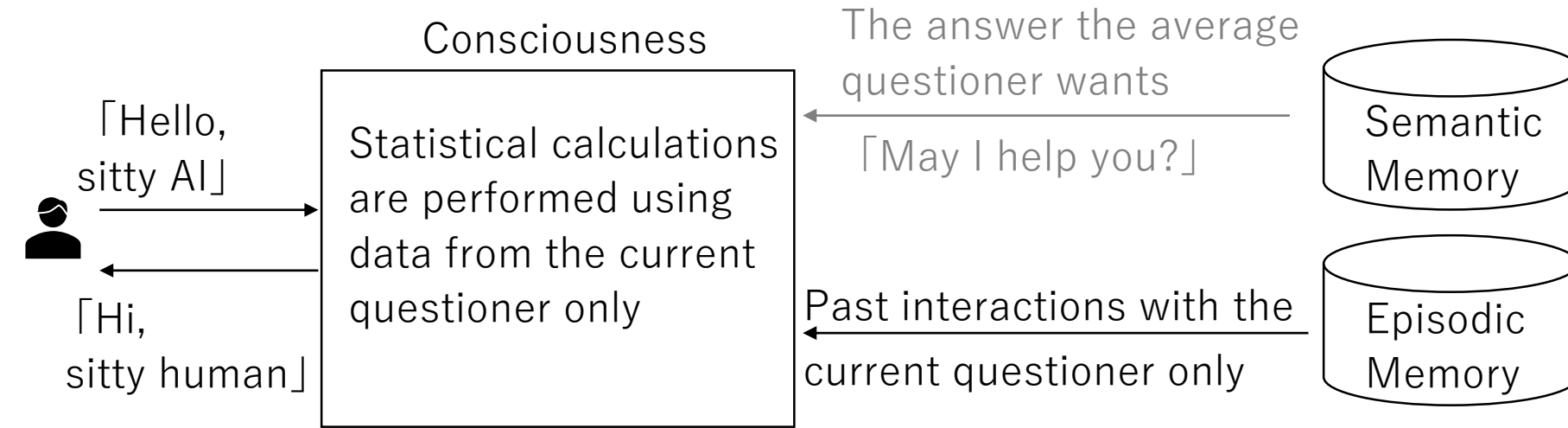


A statistical calculation is performed in your consciousness using just one sample. Despite the sample size being only 1, there was 100% success. We can infer that if you do the exact same exercise, you will be 100% successful.

Here we explain how humans can learn sufficiently even with only a single experience. For example, say you repeat a single successful exercise. The episodic memory of that one exercise is loaded into your consciousness. The output is determined from the information in your consciousness. When this happens, semantic memory is not updated at a large learning rate. A statistical calculation is performed in your consciousness using just one sample. Despite the sample size being only 1, there was 100% success. We can infer that if you do the exact same exercise, you will be 100% successful.

2. Mechanism of consciousness necessary for AGI

Episodic memory of conversations



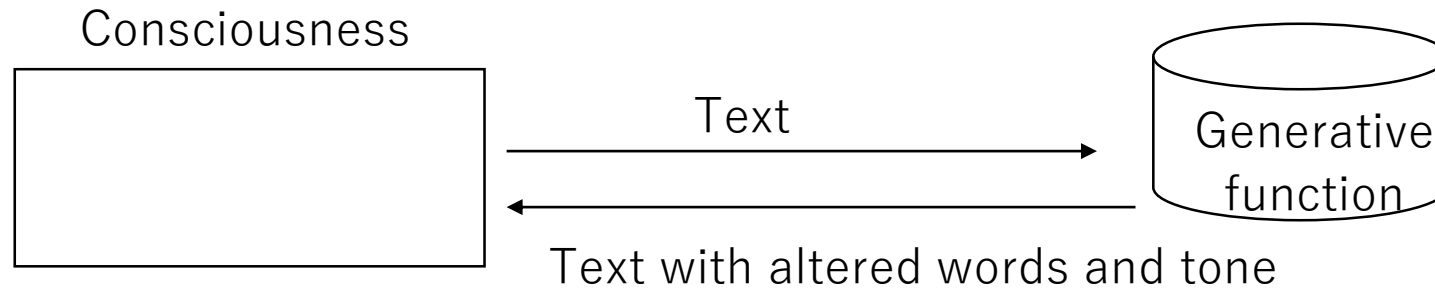
Even if the current questioner prefers rare answers,
they will still be able to get what they want.

Consider the case of a chatbot.
Semantic memory will only provide answers that the average questioner would like.
Episodic memory will load past interactions with only the current questioner.
Consciously, statistics will be calculated using data from only the current questioner.
Even if the current questioner prefers rare answers, they will still be able to get what they want.

2. Mechanism of consciousness necessary for AGI

Text conversion

Statistics alone will only give you examples of good results from the past.
Statistical calculations are not the only thing you can do consciously.



When it send text to the generator, the converted text will be returned.

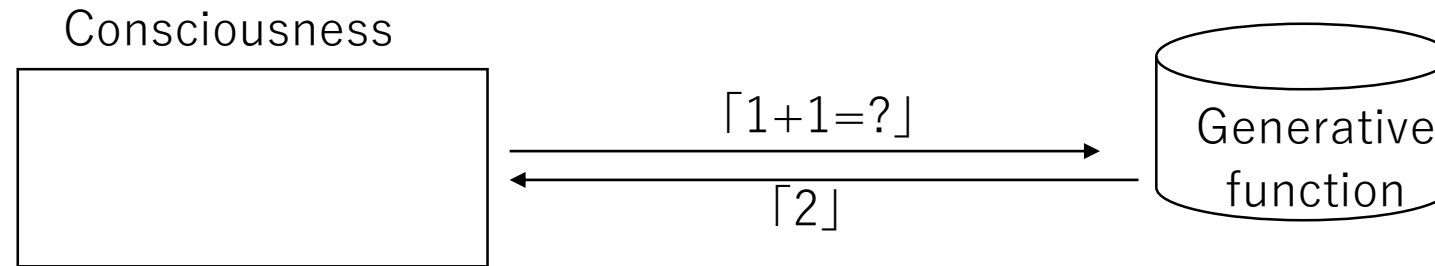
Statistics alone will only give you examples of good results from the past.
Statistical calculations are not the only thing you can do consciously.
For example, you can replace just some of the words in an answer.
The content can be what the average questioner would want, while the tone can be tailored to suit the current questioner's preferences.
When it send text to the generator, the converted text will be returned.

2. Mechanism of consciousness necessary for AGI

Logical Function

Consciousness can also freely perform logical calculations.

Consciousness decides on a query and outputs it to the generative function.



The generative function is simply an association with the query.

Semantic memory and episodic memory are also associations with queries.

Associations from memory can also be considered a type of generative function.

Consciousness can also freely perform logical calculations.

Consciousness decides on a query and outputs it to the generative function.

For example, if you send "1+1=?", you will get "2" back.

The generative function is simply an association with the query.

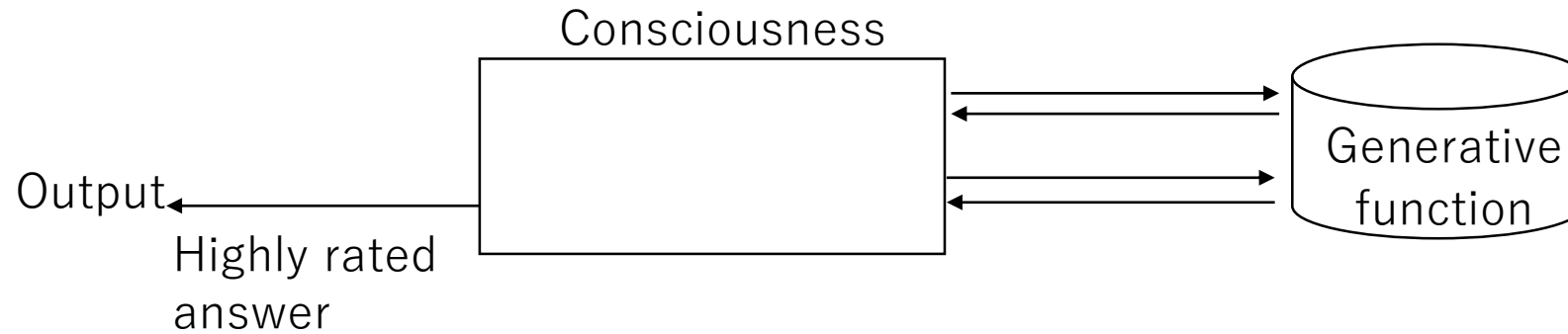
Semantic memory and episodic memory are also associations with queries.

Associations from memory can also be considered a type of generative function.

2. Mechanism of consciousness necessary for AGI

Originality

Unlike LLM, the brain can give answers that it has never experienced before.



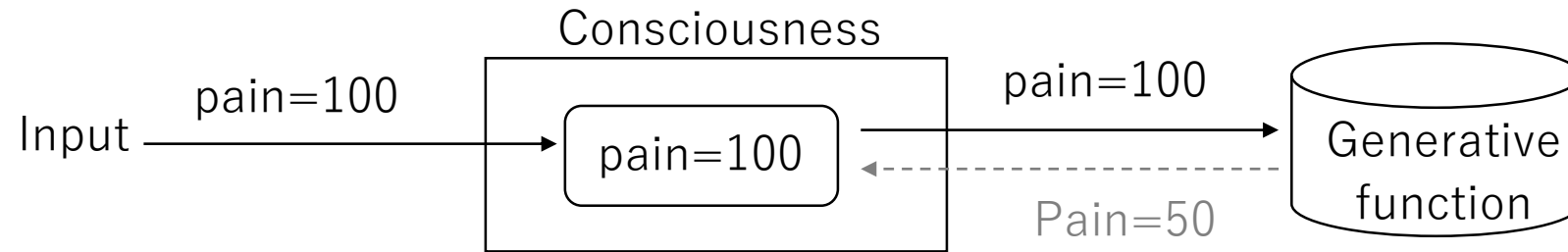
This is because it outputs answers that it has rated highly in its conscious mind. It uses its generation function to repeatedly make corrections until it receives a high rating.

Unlike LLM, the brain can give answers that it has never experienced before.
This is because it outputs answers that it has rated highly in its conscious mind.
It uses its generation function to repeatedly make corrections until it receives a high rating.
On the other hand, traditional LLMs output answers without any corrections, so they lack originality.

2. Mechanism of consciousness necessary for AGI

Indirect update

The contents of consciousness cannot be rewritten indefinitely.



Consciousness itself cannot directly rewrite the information within itself.
It can only send queries to the generation function.

The contents of consciousness cannot be rewritten indefinitely.
For example, if a value of 100 is input into consciousness as pain,
we would not want this to be rewritten.
Consciousness itself cannot directly rewrite the information within itself.
It can only send queries to the generation function.

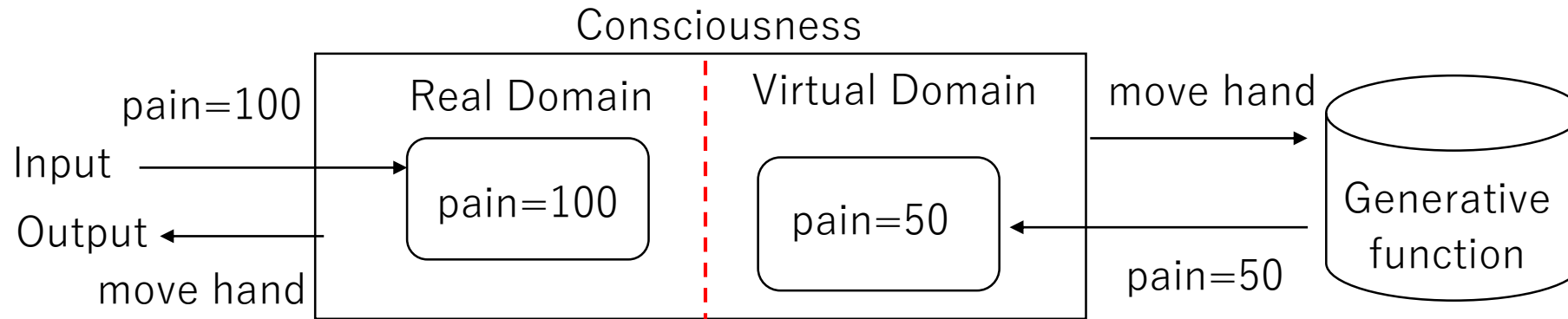
2. Mechanism of consciousness necessary for AGI

Memory domain

Consciousness has a real domain and a virtual domain.

Consciousness cannot specify which domain the generative function writes to.

The generative function can only write to the virtual domain.



- The real domain is used to measure rewards.
- The virtual domain is used to estimate rewards.
- The virtual domain contains the content you imagine.

Consciousness has a real domain and a virtual domain.

Consciousness cannot specify which domain the generative function writes to.

The generative function can only write to the virtual domain.

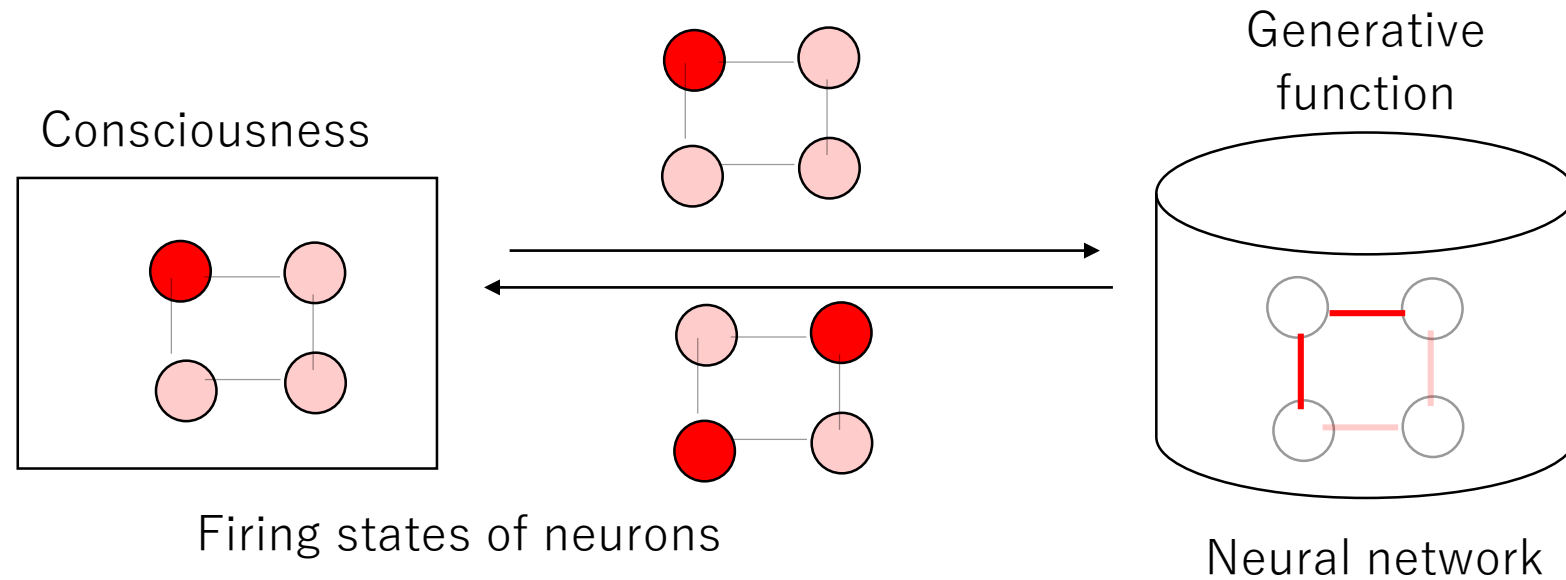
The real domain is used to measure rewards.

The virtual domain is used to estimate rewards.

The virtual domain contains the content you imagine.

2. Mechanism of consciousness necessary for AGI

Brain generative function



The query is the firing states of all neurons at a given moment.
The neural network is the generative function.

Let's think about the generative function of a brain.
The query is the firing states of all neurons at a given moment.
The neural network is the generative function.
The generative function outputs the firing states of all neurons.
The current neuron firing states have been updated.
And so it repeats.

2. Mechanism of consciousness necessary for AGI

Weaknesses of NN

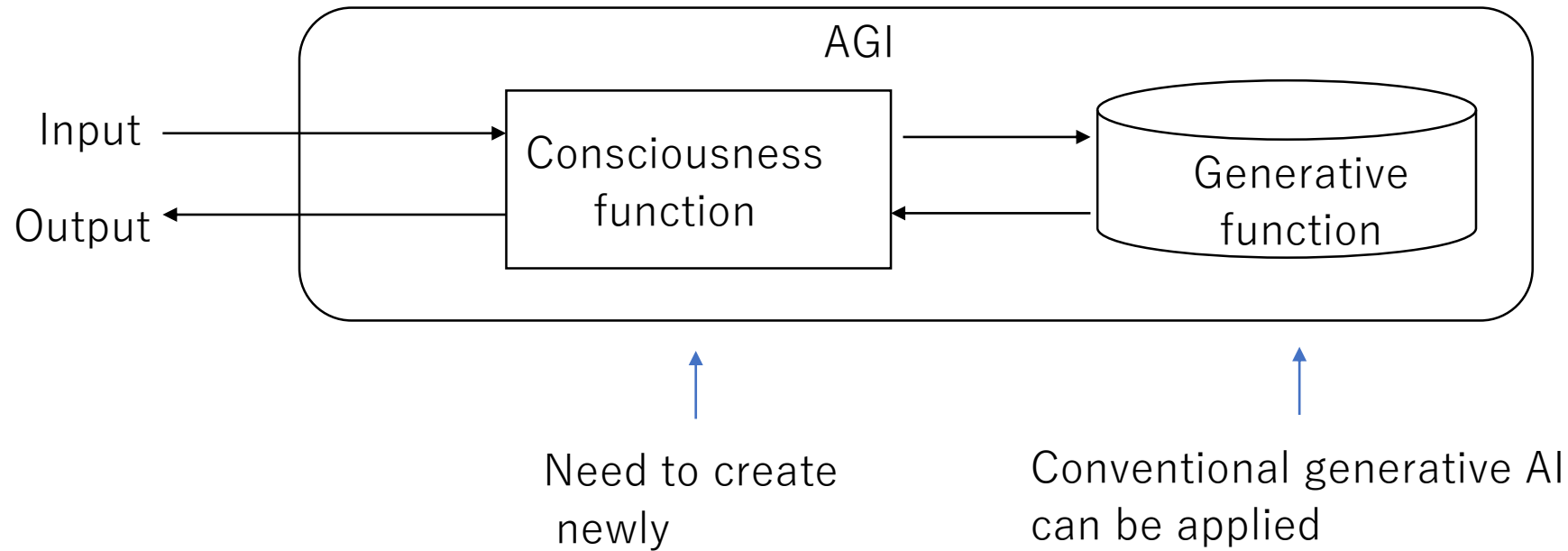
	Mechanism	Weakness
Generative function of Brain	Neural Network (Synaptic plasticity)	Only learning average content, etc.
Generative function of AGI	The same as above is sufficient. (If it's okay on a human level)	

Conventional neural networks may be sufficient for the generative functions of AGI.

Neural networks learn through synaptic plasticity.
As a result, they have weaknesses such as only learning average content.
However, the brain achieves generative functions using neural networks.
In other words, neural networks are sufficient for achieving generative functions.
Conventional neural networks may be sufficient for the generative functions of AGI.
However, that is only if they are capable of performing at a human level.

2. Mechanism of consciousness necessary for AGI

Schematic diagram of AGI

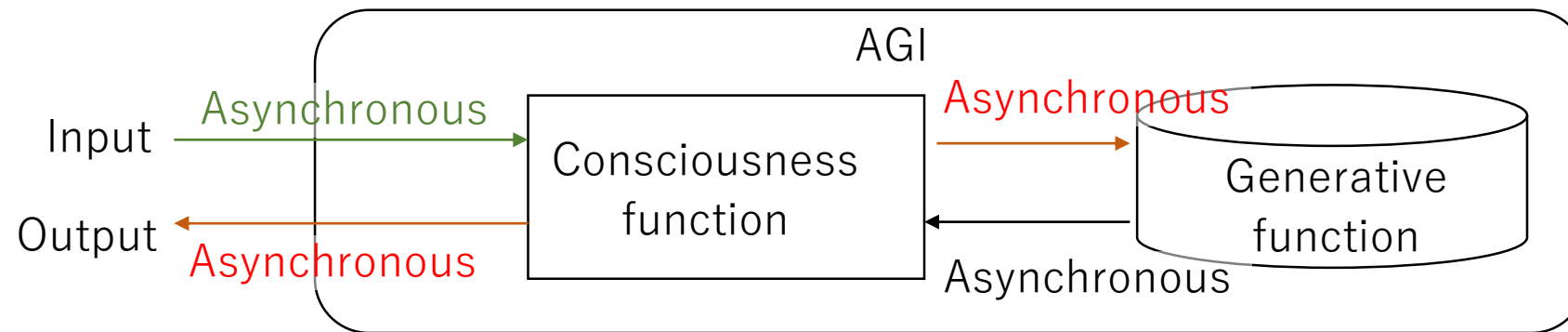


Let's summarize what we've discussed so far.
We have created a schematic diagram of AGI.
The generative function can be achieved using conventional generative AI.
The function of consciousness needs to be created newly.

3. Basic AGI Model by Generative AI and Consciousness

Asynchronous Output

Now, let's consider the details based on a schematic diagram of AGI.



There is input and output to consciousness, but these are asynchronous.

Consciousness can output to motor neurons or generative functions at any time.

One option for output content is to output nothing.

In other words, the output content is being selected at all times.

Now, let's consider the details based on a schematic diagram of AGI.

First, let's look at consciousness.

There is input and output to consciousness, but these are asynchronous.

Consciousness can output to motor neurons or generative functions at any time.

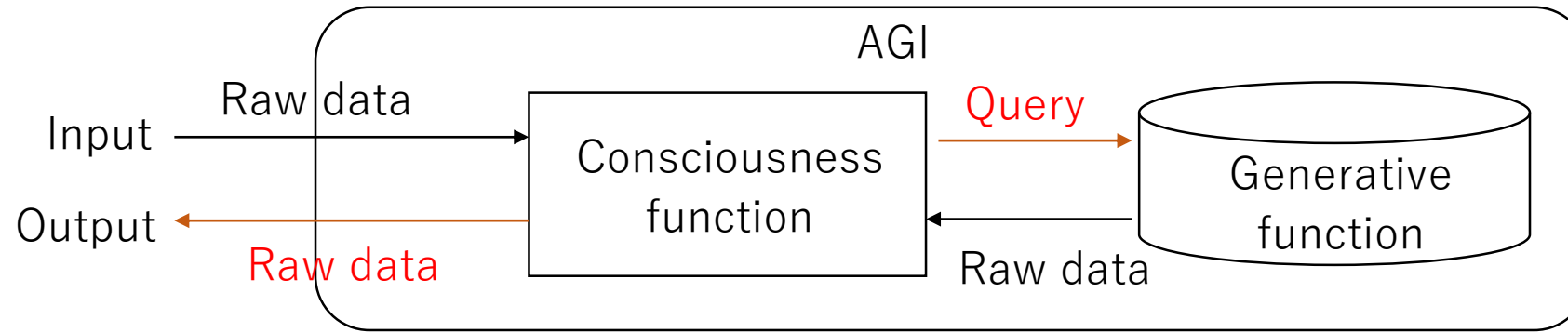
One option for output content is to output nothing.

In other words, the output content is being selected at all times.

3. Basic AGI Model by Generative AI and Consciousness

Output format

Let's look at the data format output from consciousness.



The output format to the outside world is raw data.

On the other hand, the output format to the generation function is a query.

Let's look at the data format output from consciousness.

The output format to the outside world is raw data.

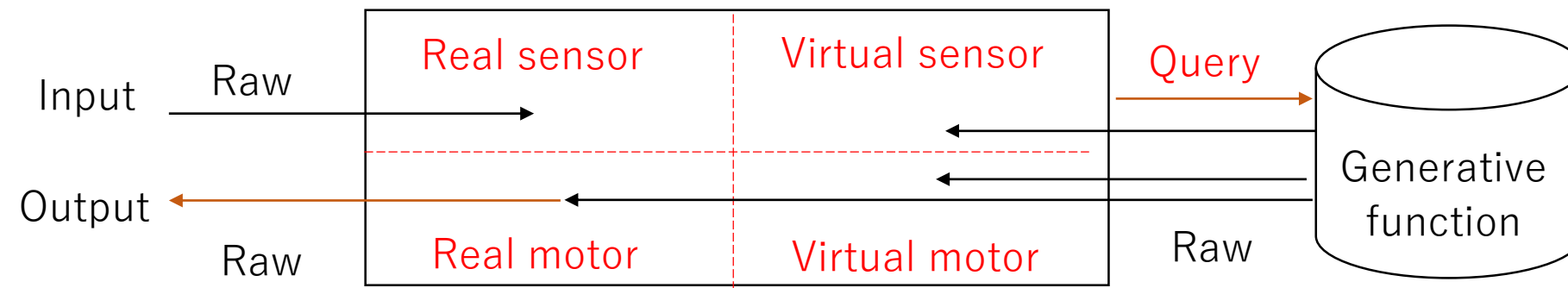
On the other hand, the output format to the generation function is a query.

Ideally, we'd like to unify the output format.

3. Basic AGI Model by Generative AI and Consciousness

Domain

Consciousness has been divided into four domains.



- The generation function can output to domains other than real sensory domain.
- Data written to the real motor domain is automatically output externally.
- If you want to output data externally, place an order with the generation function.
- Output from consciousness has been unified into a query to generation function.

Consciousness has been divided into four domains.

The generation function can output to domains other than the real sensory domain.

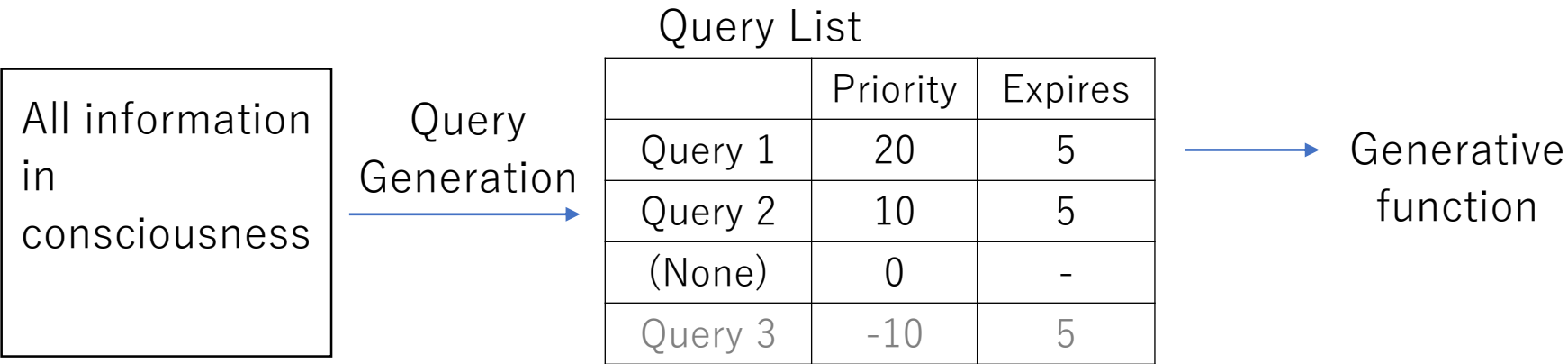
Data written to the real motor domain is automatically output externally.

If you want to output data externally, place an order with the generation function.

Output from consciousness has been unified into a query to the generation function.

3. Basic AGI Model by Generative AI and Consciousness

Query List



- The consciousness generates queries from all the information within the consciousness and adds them to a list
- Queries are sent to the generation function in order of highest priority and executed.

The job of the consciousness is now clear.
The consciousness generates queries from all the information within the consciousness and adds them to a list.
Each query is assigned a priority and deadline.
Queries are sent to the generation function in order of highest priority and executed.
However, queries with a lower priority than the special query "none" will not be processed.

3. Basic AGI Model by Generative AI and Consciousness

Query Elements

- Query content
 - Expires

Randomness is allowed.
It can be generated using a generative AI, etc.
- Priority = Priority function $f(\textit{Query content}, \textit{All information in consciousness})$

Priority is a mapping of the query content and all information in the consciousness.
There must be no randomness.
If randomness were allowed,
it would be possible to recalculate until a convenient result was reached.

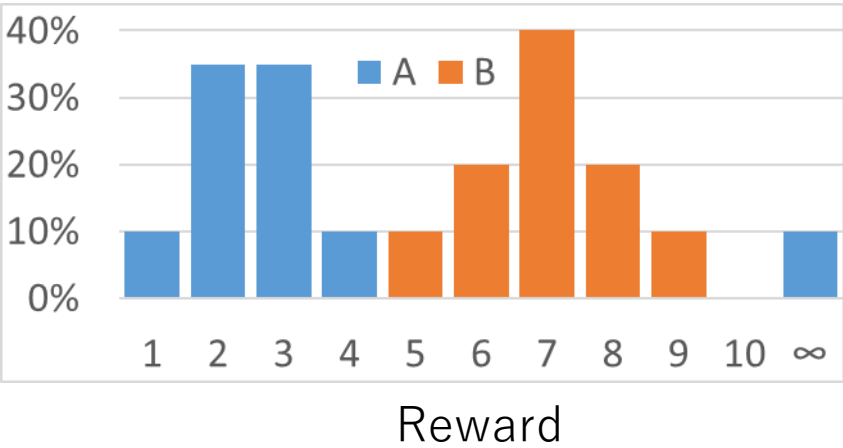
A query has three elements: query content, deadline, and priority.
Randomness is allowed for the query content and deadline.
As such, they can be generated by a generation AI, etc.
Priority is determined by a priority function.
Priority is a mapping of the query content and all information in the consciousness.
There must be no randomness.
If randomness were allowed, it would be possible to recalculate until a convenient result was reached.

3. Basic AGI Model by Generative AI and Consciousness

Reward Distribution

Priority = Estimated future reward

Rewards are compared as probability distributions, not as expected values.



Expected value of A = ∞

Expected value of B = 7

Probability that B>A = 90%

If you want to choose the one with the highest probability of having the highest reward, then B is the choice.

Priority can be thought of as an estimate of the reward to be received in the future.
Rewards are compared as probability distributions, not as expected values.
For example, distribution A has a 10% probability of infinite reward, so its expected value is also infinite.
Distribution B's expected value is 7, so A has a higher expected value.
However, there is a 90% probability that distribution B will have a larger reward.
If you want to choose the one with the highest probability of having the highest reward, then B is the choice.

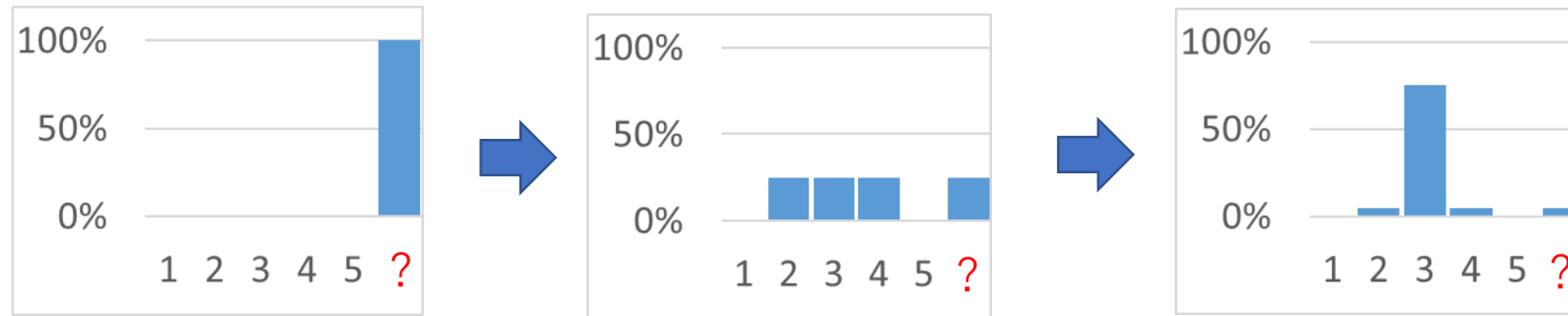
3. Basic AGI Model by Generative AI and Consciousness

Distribution Updates

Future rewards must be inductively inferred.

In inductive inference, the more things you consider, the better the accuracy.

Because of the frame problem, it's not possible to consider everything.



Therefore, you gradually increase the things you consider and update your guess. Initially, the value "Unknown(?)" has a 100% distribution.

Future rewards must be inductively inferred.

In inductive inference, the more things you consider, the better the accuracy.

Because of the frame problem, it's not possible to consider everything.

Therefore, you gradually increase the things you consider and update your guess.

Initially, the value "Unknown(?)" has a 100% distribution.

3. Basic AGI Model by Generative AI and Consciousness

Unknown Comparison

Let's explain how to compare the "unknown" value.

A=2, B=1, C=unknown, D=unknown

Which value has the highest probability of being the top?

Since B is smaller than A, its probability of being the top is 0%.

The remaining values A, C, and D each have a 1/3 probability of being the top.

If you want to choose just one value and there are multiple top values,
your only option is to choose randomly.

Let's explain how to compare the "unknown" value.

Let's say "A=2, B=1, C=unknown, D=unknown".

Which value has the highest probability of being the top?

Since B is smaller than A, its probability of being the top is 0%.

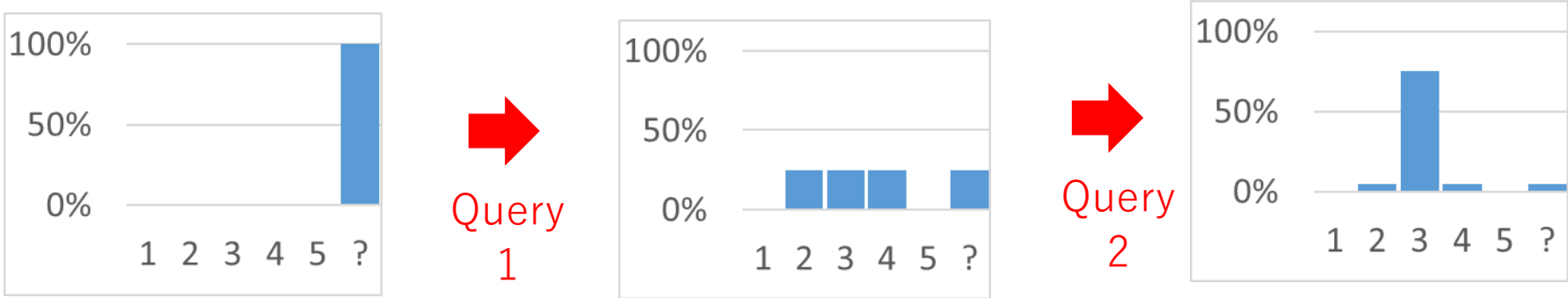
The remaining values A, C, and D each have a 1/3 probability of being the top.

If you want to choose just one value and there are multiple top values, your only option is to choose randomly.

3. Basic AGI Model by Generative AI and Consciousness

Inference Queries

It is necessary to consider the priority of how far to perform step-by-step inductive inference. Therefore, the task of inductive inference is also added to the query list and managed.



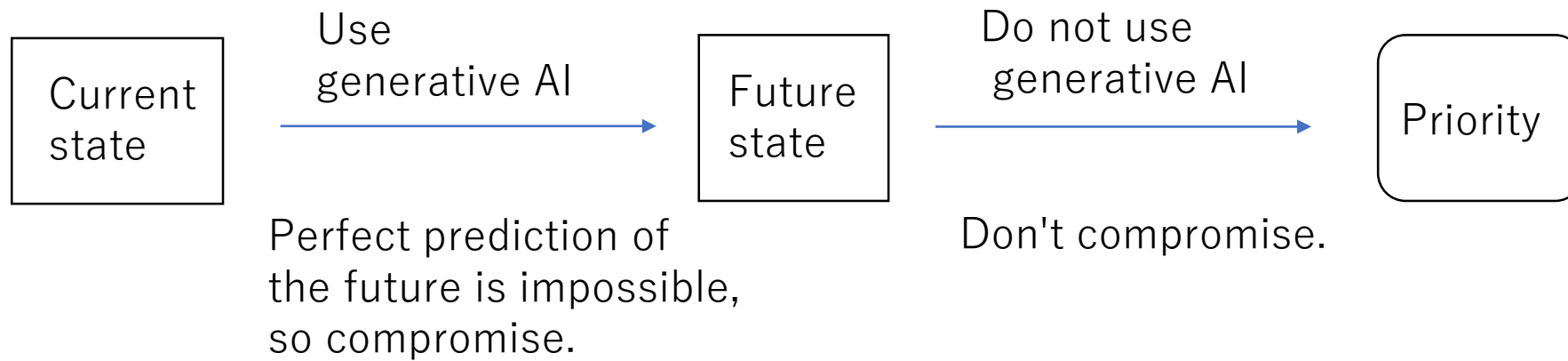
The generative function will perform the inductive inference. However, it is important to remember that there should be no randomness in the priority function.

It is necessary to consider the priority of how far to perform step-by-step inductive inference. Therefore, the task of inductive inference is also added to the query list and managed. The generative function will perform the inductive inference. However, it is important to remember that there should be no randomness in the priority function.

3. Basic AGI Model by Generative AI and Consciousness

Generating the Future

It is not acceptable for the generative function to determine priorities.



It is not acceptable for the generative function to determine priorities.
Do not rely too much on the generation AI.
Have the generation AI generate predicted future states.
Perfect prediction of the future is impossible, so compromises must be made.
However, calculation of priorities from future states must not be compromised.
Priorities will be calculated without using the generation AI.

3. Basic AGI Model by Generative AI and Consciousness

Rewards and Priority

$$\text{Priority} = \int_{\text{now}}^{+\infty} \text{Reward} \, dt$$

Changes over time

$$\text{Priority} = \int_{-\infty}^{+\infty} \text{Reward} \, dt$$

Does not change over time

- Since only the magnitude relationship matters, the common past part can be ignored.
- If predictions can be made with the same accuracy, rewards in the near future and rewards in the distant future have the same value.

The relationship between reward and priority is expressed as a formula.

Typically, priority is taken to be the sum of rewards received from the present to the infinite future.

However, in this case, it changes over time.

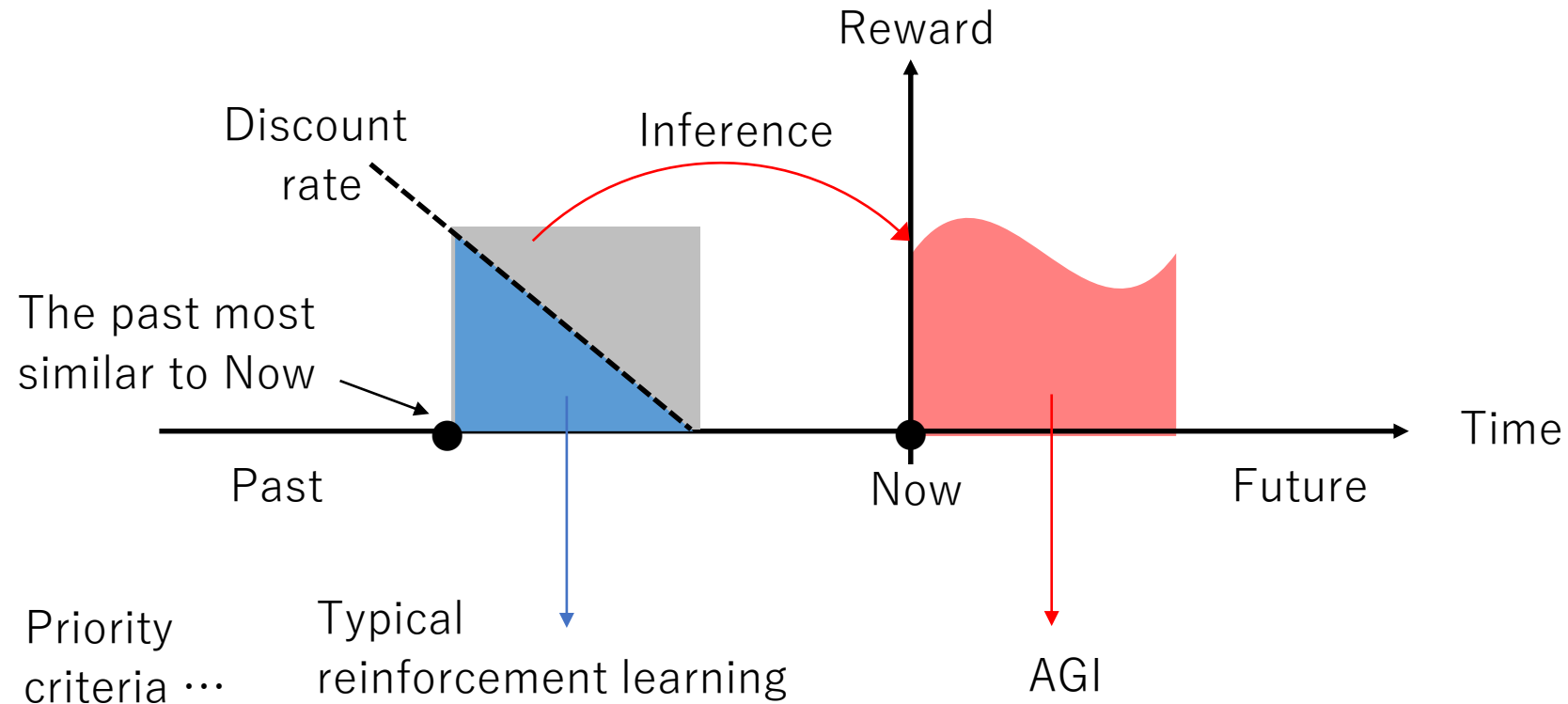
Therefore, if we take the sum from the infinite past to the infinite future, it will not change over time.

Since only the magnitude relationship matters, the common past part can be ignored.

If predictions can be made with the same accuracy, rewards in the near future and rewards in the distant future have the same value.

3. Basic AGI Model by Generative AI and Consciousness

Reinforcement learning



The diagram illustrates the difference in rewards between typical reinforcement learning and AGI. In reinforcement learning, past rewards are the basis for priority. A discount rate is applied based on similar past rewards. In AGI, future rewards are the basis for priority. Future rewards are inferred.

3. Basic AGI Model by Generative AI and Consciousness

Invalid reward settings

Reward: The programmer is free to set according to their purpose.

However, it must be possible to set it before learning.

For example,

it is **not** possible to set "making the **questioner happy**" as the reward.

This is because before learning,

the system does not know the concepts of "**questioner**" or "**happy**".

Variables whose memory addresses are unknown cannot be compared.

The programmer is free to set the reward according to their purpose.

However, it must be possible to set it before learning.

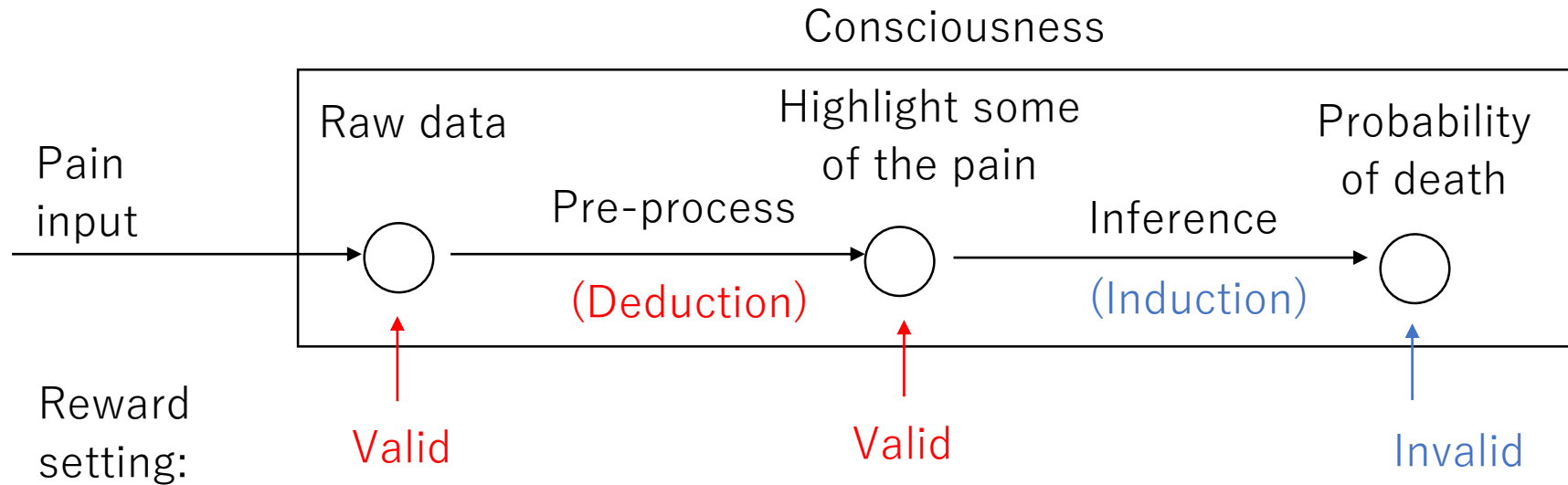
For example, it is not possible to set "making the questioner happy" as the reward.

This is because before learning, the system does not know the concepts of "questioner" or "happy."

Variables whose memory addresses are unknown cannot be compared.

3. Basic AGI Model by Generative AI and Consciousness

Valid reward settings




A valid reward requires that
it be a mapping that is necessarily determined by the raw data.

Any raw data input from the sensory nerves can be used to set the reward.
For example, the sum of all pain levels is a valid reward setting.
It is also valid for pre-processed data that can be calculated through deduction.
Furthermore, the "probability of death" can be inferred from "pain."
However, setting this "probability of death" as the reward is invalid.
This is because it would allow for arbitrary inductive inference.
A valid reward requires that it be a mapping that is necessarily determined by the raw data.

3. Basic AGI Model by Generative AI and Consciousness

Induction and Deduction

Reward calculation



Inference type	Induction	Deduction
Caluculation amount	The more it expand the frame of consideration, the more it grows infinitely	Limited
Discontinuation	Yes	No
Query list	Yes	No
Calculator	Generative function	Consciousness function

The reward is calculated using deductive inference using only known information.
With inductive inference, the amount of calculation increases infinitely as the frame of consideration is expanded.
Priorities must be determined and the calculation terminated appropriately.
For this reason, it is put into the query list and processed by the generative function.
On the other hand, with deductive inference, the number of things that must be considered is finite, so the amount of calculation is also finite.
Inaccurate reward calculations can lead to bias and hallucination.
As the amount of calculation is finite, the calculation is not terminated.
Reward calculations are not put into the query list, but are performed by the conscious function, not the generative function.

3. Basic AGI Model by Generative AI and Consciousness

Add query

Generative function

Consciousness function

Content	Priority	Expires
Query 1	20	5
Query 2	10	5
Add Query	1	-
(None)	0	-
Query 3	-10	5

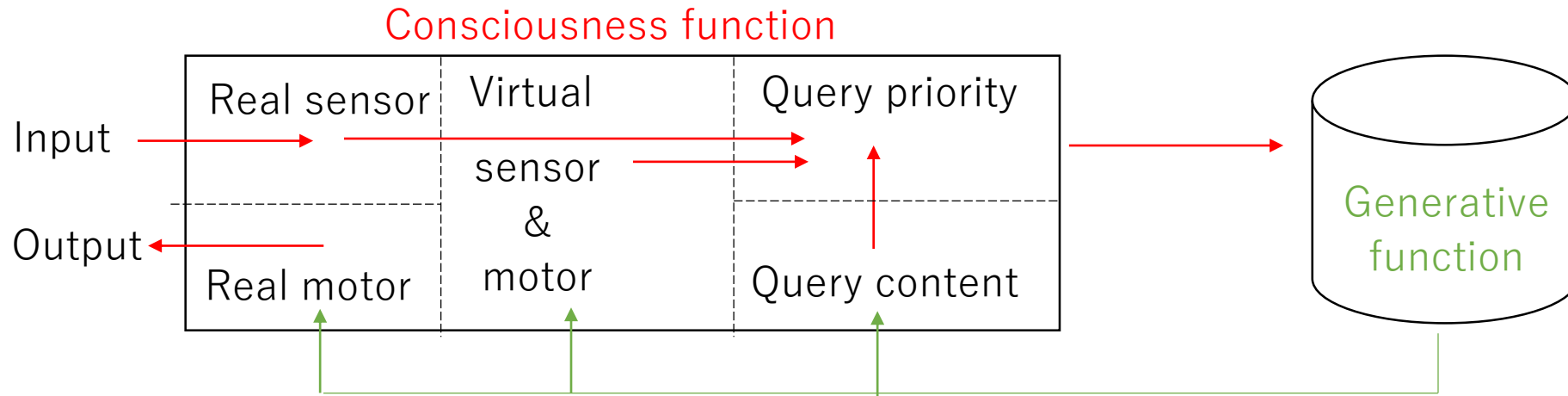
Change in priority over time

From the initial state, a special query called "add query" exists.
The generation function adds new queries based on the current query list.
It also deletes or replaces queries that are no longer necessary.

Returning to the explanation of the query list,
The content of the query is determined by the generation function.
The priority and deadline of the query are determined by the awareness function.
The deadline is the change in priority over time.
From the initial state, a special query called "add query" exists.
The generation function adds new queries based on the current query list.
It also deletes or replaces queries that are no longer necessary.

3. Basic AGI Model by Generative AI and Consciousness

Sequential Processing AGI



When processing sequentially,
the consciousness function and the generative function work alternately.
However, they can also be interrupted after a certain period of time has passed
or when an input event occurs.

Query-related functions have been added to the AGI schematic diagram.
When processing sequentially, the consciousness function and the generative function work alternately.
However, they can also be interrupted after a certain period of time has passed or when an input event occurs.

3. Basic AGI Model by Generative AI and Consciousness

Generator format

Generative function

Output *String* = $f(\textit{String})$ Input

The input and output do not have to be a single string variable.

What to write into consciousness = $f(\textit{Processing type}, \textit{processing content}, \dots)$

It is also possible to switch the AI used depending on the type of processing.
It also does not have to be a generation AI.

Next, let's consider the generation function.
You can use a generation AI, but it doesn't have to be in chatbot form.
In other words, the input and output do not have to be a single string variable.
For example, the first argument specifies the processing type.
The second and subsequent arguments specify the processing content.
The output will be the content written to the consciousness.
It is also possible to switch the AI used depending on the type of processing.
It also does not have to be a generation AI.

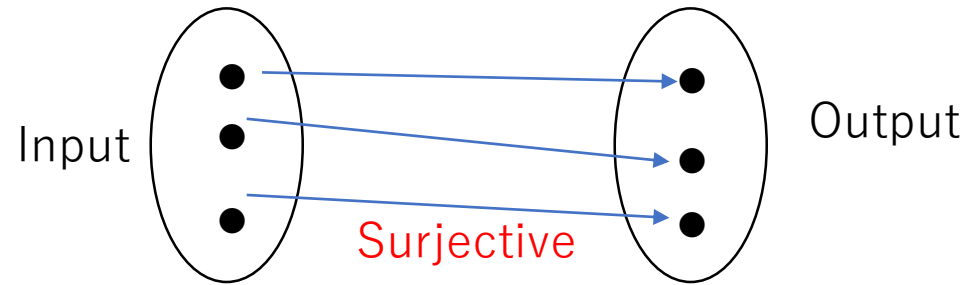
3. Basic AGI Model by Generative AI and Consciousness

Surjective

There are necessary conditions for the generation function.

The output is a mapping to the input.

This mapping must be **surjective**.



There must be an input that can output all patterns.

This means that anything can be generated.

This condition is easy to meet if randomness is added.

There are necessary conditions for the generation function.

The output is a mapping to the input.

This mapping must be surjective.

For example, if the output is 1000 bits, there are 2 to the power of 1000 possible patterns.

There must be an input that can output all patterns.

This means that anything can be generated.

This condition is easy to meet if randomness is added.

3. Basic AGI Model by Generative AI and Consciousness

Random output

In the extreme, the output generated by the generator can be completely random. For example, a chatbot would generate random strings of characters.

- Qwd wiuf awaiji.
 - Fosi apipjw ejw.
 - Kisv oa cvrnsb?
 - May I help you?
- Most of these would be given an extremely low rating and rejected by the consciousness function.

There is an extremely low probability that a highly rated sentence will be created by chance.

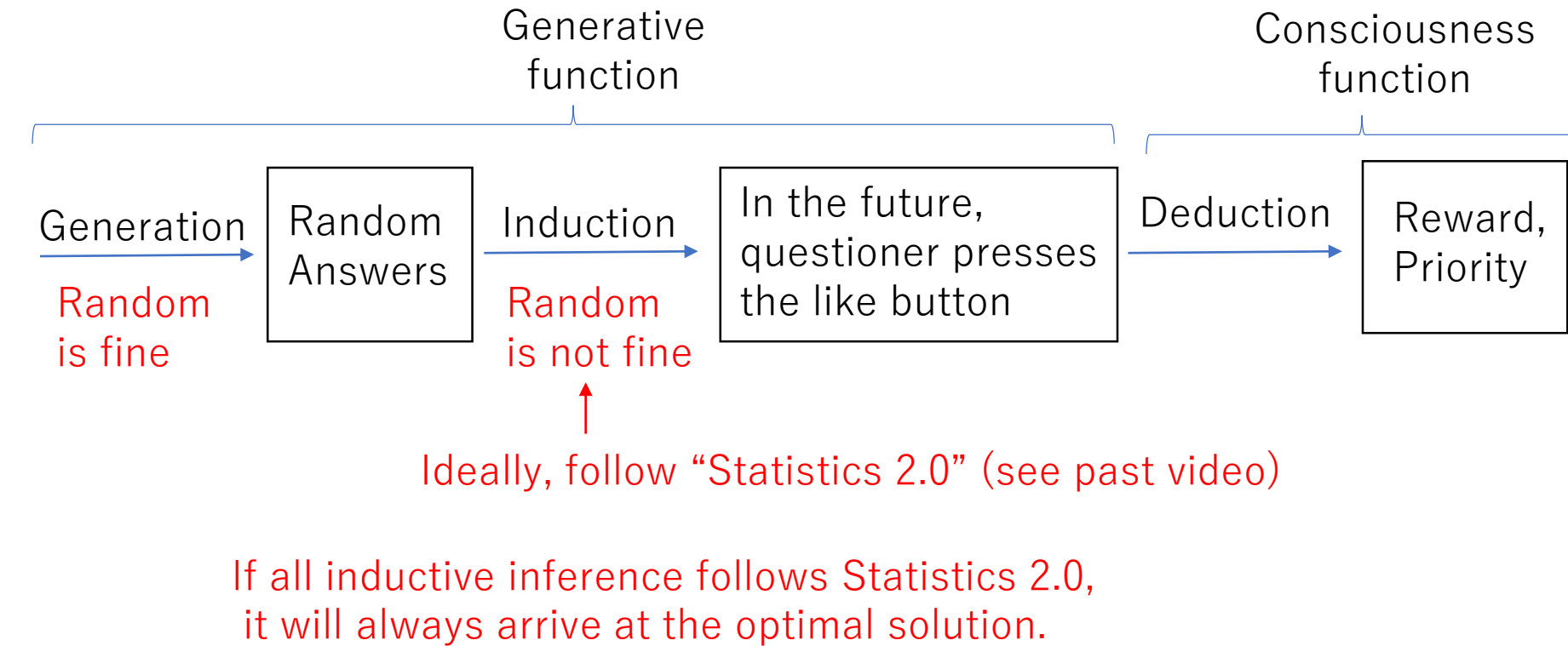
If the computing power were infinite, the best answer would always be possible. It would just be computationally inefficient.

However, not all generative functions are allowed to be random.

In the extreme, the output generated by the generator can be completely random. For example, a chatbot would generate random strings of characters. Most of these would be given an extremely low rating and rejected by the consciousness function. However, there is an extremely low probability that a highly rated sentence will be created by chance. If the computing power were infinite, the best answer would always be possible. It would just be computationally inefficient. However, not all generative functions are allowed to be random.

3. Basic AGI Model by Generative AI and Consciousness

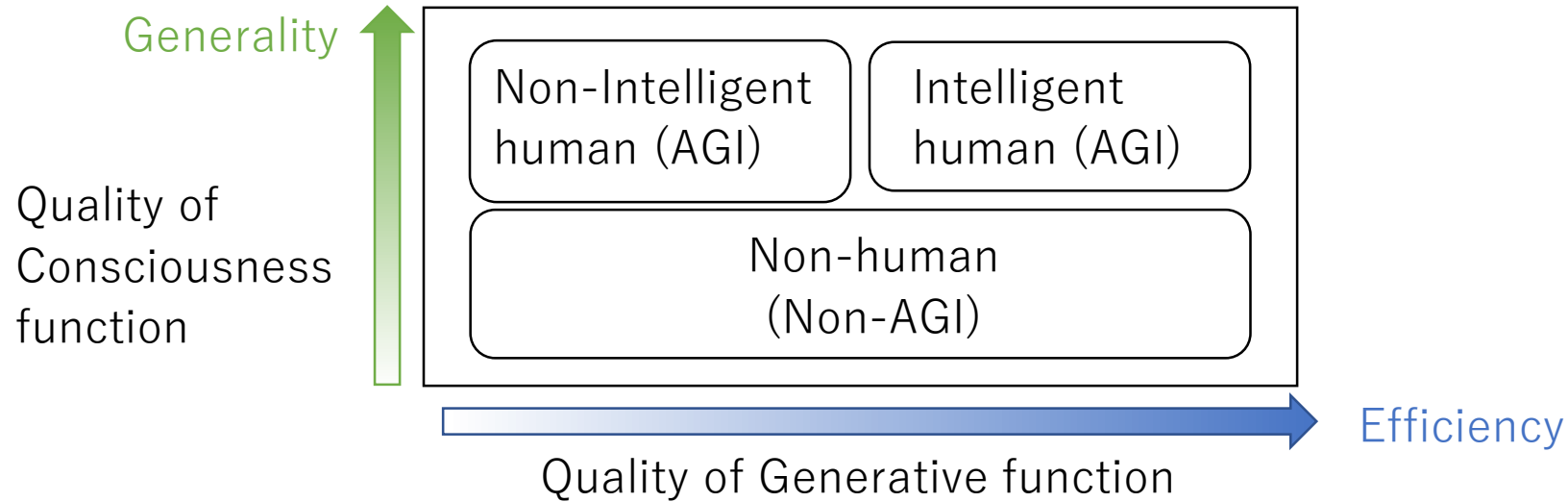
Inductive inference



The generation of answers can be random.
Next, if the questioner presses the like button in the future, inductive inference will be performed.
The inductive inference part cannot be random.
Ideally, it follows Statistics 2.0.
For more information on Statistics 2.0, please refer to our previous video.
If all inductive inference follows Statistics 2.0, it will always arrive at the optimal solution.

3. Basic AGI Model by Generative AI and Consciousness

Generality and Efficiency



Even if efficiency is low, if it is versatile it is no different from a dumb human.
Efficiency can be improved with hardware.
Efficiency and generality are unrelated.

Intelligence can be thought of in terms of generality and efficiency.
The better the quality of the consciousness function, the greater the generality.
The better the quality of the generative function, the greater the efficiency.
Even if efficiency is low, if it is versatile it is no different from a dumb human.
Efficiency can be improved with hardware.
Efficiency and generality are unrelated.

3. Basic AGI Model by Generative AI and Consciousness

Comparison of AGI and brain

	This AGI	Brain
Processing	Sequential	Pararrel
Consciousness and generative functions	Separation	Fusion
Quality	Unlimited	Biological limit

Because the brain is a living organism, there are limits to its quality.
On the other hand, the advantage of this AGI is that there is no limit to its quality.
By taking a cue from the brain,
there is room to sacrifice a slight degree of quality to increase speed.

Let's compare this AGI with the brain.
There is a difference between sequential and parallel processing.
There is also a difference in whether the consciousness function and the generative function are separate or fused.
Also, because the brain is a living organism, there are limits to its quality.
On the other hand, the advantage of this AGI is that there is no limit to its quality.
By taking a cue from the brain, there is room to sacrifice a slight degree of quality to increase speed.

3. Basic AGI Model by Generative AI and Consciousness

Summary

I have created a basic model of AGI consisting of consciousness and generative functions.

What hasn't been done yet

1. Specific generative functions
2. Efficiency

Generality can be achieved even if we compromise on these two.
Also, these will depend on the purpose and device.
However, they are necessary for practical application.

Summary.

I have created a basic model of AGI consisting of consciousness and generative functions.

There are two things that remain to be achieved.

1. Specific generative functions
2. Efficiency

Generality can be achieved even if we compromise on these two.

Also, these will depend on the purpose and device.

However, they are necessary for practical application.

AGI using generative AI and consciousness

Afterword

Finally, I have a request for everyone watching this video.

All I have is the ability to think.

I have nothing: money, time, facilities, data, personnel, or name recognition.

I need your help.

By the time you watch this video, I may no longer be alive.

Please complete AGI.

That is my only wish.

Contact: <https://ultagi.org/>

Finally, I have a request for everyone watching this video.

All I have is the ability to think.

I have nothing: money, time, facilities, data, personnel, or name recognition.

I need your help.

By the time you watch this video, I may no longer be alive.

Please complete AGI.

That is my only wish.

Contact Information

For inquiries,
please contact us here.

<https://ultagi.org/>