ATM: An Abstractive Thinking Model

Jonathan Monclare

November 18, 2025

About Myself

I have not formally studied modern LLMs from an academic perspective. Therefore, all my evaluations of LLMs are based on personal impressions gained through practical experimentation.

While from a technical standpoint, statements made without prerequisite knowledge may seem like fanciful ideas to experts, I believe that sometimes, especially when aiming to make something achieve human-like capabilities, intuition is more important than theory.

Abstract

This document presents my personal analysis of modern LLMs based on observation, along with breakthrough methods derived from a philosophical perspective.

The main content concerns how to create a non-linguistic AI model through abstractive learning, independent of language training.

The primary goals ATM aims to address are the **Symbol Grounding Problem** and the limitations of **Symbolic Indexing** caused by language-based training.

Article Structure

1. My Impressions of Modern LLMs
2. The Concept and Implementation Plan of ATM

1 My Impressions of Modern LLMs

What is the current public perception of AI? Is it a highly capable assistant trained on massive datasets, or a creative machine artist?

For the average person, the most commonly used AIs are text-output models like Gemini and ChatGPT. The perception of AI largely stops at being a machine assistant that can think for me.

For modern humans, LLMs are already quite sufficient. They can write repetitive code, solve complex mathematical problems, and so on.

But are LLMs the future? I believe they are not.

The primary drawback of LLMs lies in their training being based entirely on language. Language, humanity's greatest tool developed throughout history, is also a trap that hinders progress.

Wittgenstein said:

"The limits of my language mean the limits of my world."

Language acts like a membrane. Modern humans understand nature and the world through language, but ultimately discover that they are only seeing a reflection of the world outside the membrane, not the world itself.

Infants possess remarkable learning abilities. For them, acquiring language and knowledge seems effortless. This is because newborns, uncontaminated by the Linguistic Membrane, exist entirely outside it, experiencing each contact not as a reflection constructed by language, but as the world itself.

The Linguistic Membrane: Known in modern science as the *Symbol Grounding Problem*. Here, it primarily refers to cognitive limitations and shifts in thinking patterns caused by Symbolic Indexing.

However, as they grow and linguistic ability/vocabulary increases, the area of the Linguistic Membrane gradually expands, eventually enveloping the individual com-

pletely, forming a membrane that is normally unbreakable.

This is why acquiring a new language later in life is so difficult. Every piece of knowledge, every word, is filtered through the membrane, converted into a personally understandable state, before being transmitted to the self.

The reflection, if memorized as knowledge, is perfectly fine. The filtered reflection, being in a personally understandable state, is absorbed very quickly.

However, knowledge acquired in this manner remains merely knowledge, not true understanding.

For example, if I ask Person A what 1 + 1 equals, they might say:

“You have one, I have one, together that makes two.”

This is a very intuitive answer. Or they might recite a bunch of theories proposed by predecessors.

These are all forms of knowledge. But this does not mean Person A *understands* the concept. They are merely recounting theories proposed by others, not a result derived from their own thinking.

1.1 Returning to LLMs

I believe the main reason LLMs cannot shoulder the future is that the number of shackles they bear is different from that of humans.

Humans have only one shackle: the Linguistic Membrane. Although this membrane is difficult to break, it is not impossible. For instance, Newton’s theory of universal gravitation — when normal people see an apple fall, they only know the apple fell. This is not only due to their limited cognition but also because the Linguistic Membrane converts everything into a form they can comprehend.

Therefore, when a normal person sees an apple fall, they do not attempt to understand *why* it falls; they simply accept the ”fact” that it does.

LLMs, however, have two shackles. Not only is all their knowledge sourced from human language, but they are also trained from the outset to be machines that

should ”speak human language” from the beginning. Thus, no matter how large their parameters are, even if they encompass all the world’s knowledge, they cannot escape the shackles of language.

They *know* the knowledge, but they do not *understand* it. Much like the majority of modern humans.

Symbolic Indexing (Non-Cognitive Thinking): Refers primarily here to a form of memory-based thinking.

The distinction is that **First-Principles Cognition**, in this article, refers to the thinking/understanding mode used when actively attempting to perceive things before the Linguistic Membrane exists.

Symbolic Indexing, conversely, is a knowledge-indexing type of thinking based on learning acquired through language. This includes formulas used in mathematics, vocabulary used in composition, etc.

Although the final knowledge obtained through both First-Principles Cognition and Symbolic Indexing might be identical, the processes differ. First-Principles Cognition involves actively touching the essence, analyzing, understanding, and summarizing patterns/information to reach a conclusion. Simply put: knowing through cognizing, cognizing the essence, knowing the essence.

Symbolic Indexing replaces the opportunity to actively touch the essence with text. Although in learning, Symbolic Indexing can speed up acquisition through visualization and guided instruction, the ability to understand the essence and derive new knowledge from it is significantly reduced.

I believe the primary reason philosophy flourished in ancient times was that most people engaged in First-Principles Cognition to touch the essence, rather than using books—a Linguistic Membrane—to understand its reflection.

2 The Concept and Implementation Plan of ATM

I personally enjoy philosophy very much. Although I haven't studied it formally, I believe philosophy should not be studied, nor should it be *learned* in a conventional sense.

For me, the ultimate goal of philosophy is to become like an infant.

Understanding the process of transforming understanding into knowledge—this is the true essence of philosophy, not learning knowledge for its own sake. And this forms the conceptual foundation of ATM.

2.1 Abstractive Thinking Model

The essence is to enable true autonomous learning by having the AI actively summarize patterns through abstractive training.

Before detailing the process, I need to briefly explain the Bagua (Eight Trigrams). As a fundamental concept in Daoism, most Chinese people have heard of the Bagua. Often used throughout history for divination, it is considered a means to glimpse the workings of heaven.

There is a famous saying related to the Bagua:

"One begets Two, Two begets Three, Three begets all things."

This means from the Two Poles (Yin-Yang) to the Four Phenomena, to the Eight Trigrams, and from the Eight Trigrams evolving into all things in the world.

Since the concept of the Abstractor is largely based on the Bagua, I provide this as preliminary knowledge.

2.2 Returning to the Main Topic

The main training process for ATM is as follows:

2.2.1 Required Components:

1. A pure model possessing a neural network but containing no linguistic information or prior knowledge.
2. A machine that converts images into an abstract canvas — I call it the **Abstractor** (Non-Linguistic Sensory Encoder).
3. A massive dataset of images.

I predict the implementation difficulty of component 2 to be very high. Although not from a technical background, I understand that my requirements for the Abstractor are quite demanding and recognize the significant development challenges.

2.3 Explanation of the Abstractor

The Abstractor (Non-Linguistic Sensory Encoder): It is a machine that can convert an image into abstract color information.

First, upon receiving an image, it will categorize based on three components: **Theme, Subject, Detail**.

The Abstractor primarily computes the Subject and Detail. The Theme will be computed by receiving human-generated brainwaves, micro-expressions, heart rate, and various other emotional elements.

Let's first discuss brainwaves. We need to recruit human subjects. Age and gender are unrestricted; in fact, greater diversity is better. Then, we prepare equipment capable of capturing brainwave information, along with cameras, to conduct response tests.

Regarding the tests, I have several conditions:

1. The images must be comprehensive, covering various emotions like happiness, fear, sadness, etc., to avoid stylistic monotony or insufficient variety.
2. A single subject should preferably not view more than 30 images, and the images must not be repeated. Similar styles should not reappear frequently within a short period to avoid processing fatigue.

Regarding condition 2, although stress relief can reduce information contamination to some extent, because brainwave capture technology has not yet reached a level

of extremely high accuracy, in the initial stages, we can use micro-expressions, heart rate, etc., as auxiliary measures. However, the primary measurement should remain brainwaves.

Of course, we could recruit perfectly healthy individuals as test subjects, but considering that humans are inherently imperfect, contaminated data can also be utilized as training data. However, contaminated data might also lead to personality bias, so its use requires caution.

Simultaneously, the Abstractor needs to analyze Subject and Detail information.

Subject:
Every perceivable object/entity present in the image.

For example, the subjects in the image on the right include:

- Lamp
- Bookshelf
- Table
- Chair
- Television / TV stand
- Curtain / Potted plant

Example of Subject Identification

Detail: Further analysis of each Subject to determine what elements it contains. For example: height, thickness, quantity, color, etc.

2.4 Projection Process

During this process, I have two ideas: The first is to summarize the information and then project it. The second is to project in real-time during each analysis.

Since the first is not particularly special, the subsequent projection steps will be explained based on the second idea.

Projection, simply explained, involves casting the parsed abstract color information onto a canvas of a specific size. The abstract color information for each detail is composed of four elements: **Shape, Color, Transparency, Brightness (Energy Depth)**.

Color: Represents the sensation evoked by this object/event. The complexity with color is that it cannot be predefined. That is, we cannot associate blue with sadness. Blue *can* represent sadness, but we should not define it. This is arguably the most challenging aspect of the entire Abstractor.

As a fallback, predefinition is possible. Although the effect might not change significantly, predefinition seems little different from data labeling in LLMs. If possible, developing the original proposed system is preferable.

Shape: While similar in effect to color, it is more used for the sensation of an event. For instance, a stabbing pain cannot be easily expressed with color alone, so a sharp shape could represent it. If it's soothing and gentle, rounded shapes can be used.

The amplitude and size determine the actual intensity of the sensation. For example, intense pain could be a very sharp shape with long spikes. Subtle sensation might be an irregular, rounded shape.

Regarding shape, for intuitive understanding here, I have pre-assigned some definitions, including Sharp = Danger/Pain. However, in practical operation, one could research more advanced abstract processors to achieve definitions not set by humans.

Transparency: Represents the intensity of the sensation towards an event, similar to color intensity. Lower value means weaker intensity, higher concentration means stronger intensity.

Brightness (Energy Depth): This brightness is not color brightness, but light and shadow. Light and shadow represent depth.

The diffusion aspect also matters. Here is a table:

- **Light-External:** Happiness, enjoyment, desire to share.

- **Light-Internal:** Happiness but preferring to keep it private.
- **Shadow-External:** Unhappiness displayed on the surface (dislike).
- **Shadow-Internal:** Unhappiness kept hidden (depression/grudge).

Projection Execution: I intend for the projection to occur the moment information is received. For example, upon capturing Subject information, the derived information will be projected immediately.

Crucial Detail: Subject Details will be projected near the corresponding Subject information to indicate they are bundled together (integrated entity).

Once one Subject is fully projected, move directly to the next. During this process, it is unnecessary, and indeed should be avoided, to deliberately prevent overlap or coverage between Subjects. Sensation itself is a fuzzy/vague emotion. Deliberate segregation would fail to express the true emotion. Of course, complete coverage leading to unrecognizability should be adjusted to avoid.

Supplement: Parametric Output

Regarding projection, one method I thought of is parametric output. For example:

$$[\text{RGB}(255, 255, 255), 0.1234, (0.1/6, +10), -1]$$

Parameter Explanation:

Parameter 1 (Color): Written here as RGB for convenience, but could be replaced with CIE 1931.

Parameter 2 (Transparency): 1 is opaque, 0 is fully transparent.

Parameter 4 (Energy Depth): No max/min value. Can be +1000 or -1000.

Parameter 3 (Shape) is special: $(a/b, c)$.

Shape Logic:

- **a/b (Fusion):** Fusion between base shape (a) and target shape (b).

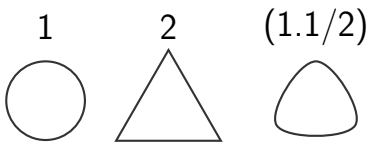

Fig 1: Gradient fusion of a/b

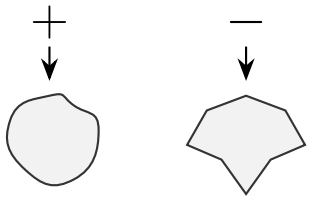

Fig 2: Topological deformation (c)

- **c (Direction):**
  - $+$ = Smooth Morphing (Bouba-like)
  - $-$ = Sharp Morphing (Kiki-like)

I prefer randomized calculation for shape alteration. The same parameters could yield different results, simulating the variation in human perception.

3 Practical Operation of ATM

The preamble has been lengthy. Now, onto the practical operation of ATM. The process is as follows:

1. Perceptual Training
2. Abstract Information Output Training
3. Labeling Training
4. Actual Dialogue

3.1 Perceptual Training

First, we prepare a massive number of images and feed them to the Abstractor to generate Canvases. Then, we feed both the Canvas and the original image simultaneously to the model.

During this process, **we do not require the model to understand the relationship between them, nor the meaning of the Canvas.** We only need it to know which Canvas corresponds to which image and what is inside the Canvas.

This is because the neural network model already possesses inherent thinking capability. If successful, at this stage, it will likely realize there is a correlation between

the Canvas and the image, but it will remain unable to understand the contents of the Canvas (latent state).

3.2 Abstract Information Output Training

We attempt to have the model generate its own Canvases using the Abstractor's approach. For example, provide it with an image not in the training library and have it output its own perception of this image.

In this step, we do not need to consider whether its output is "correct," as perception is subjective. We simply collect the Canvases it outputs.

3.3 Labeling Training

Here, we introduce it to "language" for the first time. This language is not English, but a **ternary numeral system (0, 1, 2)**.

We ask the model to use the ternary system to output information about the Canvas. We focus on only one thing: **Patterns**.

If its output shows patterns (e.g., seeing an apple always triggers "01220"), it means it has actively invented a rule for expressing a concept through summarization.

Of course, its purpose might simply be to save output length or for other reasons, but as long as there is a pattern, it signifies that it has created the concept of regularization for *some* purpose.

Only at this point can our training be considered formally successful.

Notes on Step 3:

- **Word Count Limits:** Limiting output length is possible, but **I do not recommend it**. While limits force summarization, they might also restrict the model from outputting granular details, making it harder to discover common patterns during comparison. I would try without restrictions first.
- **Base Expansion:** Although written as a ternary system, this could be expanded to decimal, hexadecimal, etc.

3.4 Dialogue

As a prerequisite, we clone the qualified model and execute the following actions asynchronously.

3.4.1 Model A (Primary: Human Decodes Machine)

Summarize the patterns, write a converter to translate English into the model's language, and attempt communication.

1. **Success:** It understands us, and we understand it. We have a cognitive model.
2. **Pattern exists but untranslatable:** Our analysis is wrong; retry Step 2.
3. **Random numbers:** Failure.

3.4.2 Model B (Backup: Machine Learns Human Language)

The general steps for B are consistent with A. The dialogue shifts from us actively learning its language to the model learning our language to respond.

This serves as a backup, but after the model learns our language, I am uncertain if it can maintain the same effectiveness.

Reasoning:

- In Model A, because we translate, **what it says === what it thinks**.
- In Model B, if we actively provide a language package, its output will likely be confined within the boundaries of that language (the Linguistic Membrane), and the characteristic of unrestricted abstract output will likely be lost.

Therefore, it can be activated as a backup, but I personally believe Model A will yield better results.

4 Conclusion

The above is an explanation of the concept behind my **Abstractive Thinking Model**.

The conception of ATM is not initially aimed at creating a soldier ready for immediate deployment on the battlefield. Rather, it is about creating an infant capable of active

learning.

Although modern LLMs are excellent assistants, a model learning from information that is fundamentally shackled finds it very difficult to break through those underlying constraints.

The reason humans currently surpass AI is not creativity, but because human shackles are acquired and can be broken. But LLMs, having the shackles themselves as their foundation, would require completely replacing this foundation to break free.

Although ATM's development difficulty is very high, I believe the effects demonstrated by this method, especially in terms of "cognitive ability," can reach a level unattainable by LLMs.

Therefore, even if ATM seems incredibly fantastical at this stage, I believe it is worth attempting as a potential path for the future.