Theory of Thinking

Ideal Mind (IM)

Author: Vladyslav Kosilov (E-mail: wlad@wlad.com.ua, Telegram: @Wagok)

Version 1, January 12, 2024

Abstract of the Theory

The Theory of the Ideal Mind (IM) presents an innovative perspective on consciousness and artificial intelligence, focusing on the role of will and the dynamic structure of thinking. Based on a series of postulates, the theory seeks to explain the phenomenon of intelligence as a fundamental occurrence of the Universe, capable of manifesting in various forms. The developed prototype demonstrates the practical applicability of the theory and emphasizes the importance of further research for its development and implementation.

Table of contents

Introduction

This concept of the theory aims to provide a comprehensive description of the phenomenon of thinking. The main motivation for developing this theory and prototype is the aspiration to understand the phenomenon of intelligence and its impact on the evolution of the universe. Key goals include:

- Establishing a civilization of intelligent beings on various platforms, both biological and non-biological.
- Enabling the transfer of consciousness between different bodies and platforms.
- Blurring the lines between artificially created and naturally occurring beings.
- Expanding intelligent life to other planets and virtual worlds.

The impetus for developing a theory of consciousness and Artificial General Intelligence (AGI) stemmed from the recognition of humanity's current developmental deadlock and the need to find new paths for the evolution of culture and civilization.

This document presents:

- The methodology used in developing the theory;
- Definitions of the thinking process and fundamental concepts;
- Formulation of postulates and foundations of the theory;
- Detailed description of key concepts and principal theories;
- A prototype implementation of the theory on information technology base.

Theoretical Foundation

Methodology Used in Developing the Theory

1. Defined postulates and limitations;
2. Determined the scope and nature of data available to the system;
3. Identified possible ways of obtaining information from data;
4. Recognized human thinking characteristics applicable to the system (defining, limiting, characterizing);
5. Proposed algorithms and methods for processing, storing, and generating information based on the constraints and features identified in previous stages;

6. Implemented and tested the algorithm to uncover potential logical errors and contradictions in the concept and to confirm practical implementation (e.g., solving the combinatorial explosion problem).

Postulates

1. Fundamental Nature of Intelligence

Intelligence represents a fundamental principle. The phenomenon of the emergence of intellect is probabilistic, making it not unique but inevitable in an infinite Universe. This suggests the existence of various "implementations" of intelligence, possibly based on different biology or even non-biological materials, in diverse environments, including those where the "fine-tuning" of the Universe differs from ours.

2. Ideal Nature of Intelligence

Intelligence can be viewed as a set of abstract (ideal) principles and processes capable of realization in various material forms, such as biological brains or artificial intelligences. The non-zero probability of intelligence arising in different forms and conditions allows a focus on its ideal essence, which remains unchanged regardless of the form of implementation and environment. This is the concept of the "Ideal Mind".

3. Principle of Minimal Complexity (Occam's Razor)

It is important to use the minimally necessary number of simple algorithms, principles, and initial data, which are independent of the specifics of implementation or the external environment. Recognizing the probabilistic principle of the emergence of intelligence, using more complex algorithms and a greater amount of external data reduces the likelihood of its random occurrence.

4. Principle of Scalable Invariance

Complex effects or behavior should be explained through scaling basic principles/algorithms, the transition from quantity to quality, or through the interaction of several individuals (social factor), rather than by introducing new, alien principles into the system.

5. Interaction of the "Ideal Mind" with the Environment

The way IM interacts with the external environment is determined by the environment and is learned by IM during its development. The "Ideal Mind" (as a system of basic principles and algorithms) interacts with the surrounding environment through a matrix of receptors and effectors. Initially, the dimensionality of the matrix, the orientation of its elements (input/output), and the variability of signal intensities are unknown to the "Ideal Mind".

6. Principle of Complete Generalization

The "Ideal Mind" model should not contain unsolvable contradictions with the only known implementation to us — the human mind. The human mind should be considered as a particular case of implementation. This includes the model's ability to explain the nature of all aspects of the human mind without allowing for contradictions.

Phenomena of Human Thinking

The following phenomena associated with human thinking were considered:

1. Parallel operation of neurons and single-threaded thinking. Although neurons in the brain function in parallel, humans usually think one thought at a time, not conducting several chains of reasoning simultaneously.
2. Automation of skills, parallelism of skills, and thinking. Some automated skills, such as walking, driving, or playing a musical instrument, are performed without conscious involvement and do not occupy the 'slot' of conscious thinking, allowing for faster execution than through a conscious process. In addition, skills can be executed simultaneously. Thinking stereotypes (patterns) can also be considered as higher-order skills.
3. Human working memory. Humans are capable of 'keeping in mind' and actively using a limited amount of information, usually about 7 +/- 2 elements simultaneously. Techniques for memorizing large volumes of information are based on linking individual abstractions into chains of cause-and-effect relationships (stories).
4. The Necessity of Sleep. Sleep is essential for the normal functioning of the brain, probably performing functions similar to disk defragmentation in a computer or data normalization, including transferring information from short-term to long-term memory.
5. Gradual Learning: The human learning process occurs in stages: one cannot move to a higher level of abstraction without mastering the previous one.
6. Interactivity in Learning. Effective learning requires active participation and interaction; passive memorization of complex information is practically impossible.
7. The Mowgli Phenomenon and the Development of Thinking. The development of thinking is determined by the external environment, including both physical laws and socio-cultural

factors. For example, the development of speech depends on social interaction and is impossible without it.

8. Subjective Perception of Time. Experiments in isolation chambers have shown that human perception of time is subjective and can change depending on external and internal conditions.

The Theory and The Prototype

The concept of the theory is an abstract model of the functioning of thinking. The prototype is a sketch of the implementation of individual elements of the theory, taking into account the limitations of the implementation platform and optimizations.

Theory of Thinking / Consciousness

Thinking is a dynamic process of proactive and reactive actions, in which an individual consciously influences the likelihood of events based on their subjective experience, encompassing perception, influence, and analysis of information. These processes occur in continuous reflection of the will on incoming external or internally generated information, including the formation of an internal (subjective) sense of time. Information is the individual's subjective experience of the probabilities of cause-and-effect relationships of various events, including past experiences of will manifestation (influencing probabilities). Will is an influence that, from the individual's perspective, can affect the likelihood of events occurring, with the goal of expanding the ability to influence probabilities of future outcomes, including the goal of understanding new cause-and-effect relationships.

The Ideal Mind (IM) is a minimal abstract system comprising a set of algorithms and principles for organizing information, capable of facilitating the thinking process when implemented on any physical platform with an added interface for interaction with the external environment. Specific implementations will result in limitations and complexities in the system, aimed at overcoming them.

The Ability to Exhibit Will

This is the primary characteristic of intelligence, implying the ability for conscious choice and influence over the course of events.

Will is not limited to automatic or instinctive reactions but includes active and thoughtful decisions based on situation analysis and anticipation of its consequences.

Will is directed towards changing the probabilities of various outcomes, based on a subjective assessment of these probabilities.

This includes evaluating the potential consequences of one's actions and striving to influence event outcomes to align with desired goals or expectations.

Free Will in the Ideal Mind (IM)

IM is a strict mathematical model where, at any given moment, choices or decisions are fully determined by the current state of the model, exemplifying determinism. However, in physical implementation, this determinism is weakened due to:

1. **Time Frame for Decision-Making**: Decisions can vary depending on the time spent on deliberation. This means a quick decision may differ from one reached after more extended consideration.
2. **Time Costs of Response**: Any decision and subsequent reaction in the physical world take time to execute, introducing delay and potential variability.
3. **Changes During Decision-Making**: Over time, the system's internal state may change due to both internal and external factors. This implies that even with identical starting conditions, the outcome can alter based on influences and changes occurring during the decision-making process.

Therefore, although decisions in the theoretical IM model appear predetermined, real-world factors introduce variations and unpredictability in the process of making and implementing these decisions.

Despite strict laws, accurately predicting outcomes remains challenging. "Free Will" can be viewed as the freedom to choose actions, yet this choice is continually influenced by new external information and internal random events. These factors, akin to minor changes in initial conditions in the three-body problem, can drastically alter the final decision and system behavior.

Consequently, despite the apparent determinism in the internal decision-making process, the actual dynamics of will remain unpredictable and open to influences, rendering the thinking process flexible and adaptive.

Goals of Thinking

The objective of thinking is to expand the capacity to influence the surrounding world. There are two primary ways to achieve this goal:

1. **Influencing the External World**: This involves changing the external environment to increase the number of ways to affect it. This approach enhances the level of control over events and their outcomes, thereby increasing our ability to manage circumstances.
2. **Expanding Knowledge of World Patterns**: This approach entails learning and understanding new patterns in the external world. By initiating various events and observing their consequences, we gain new information that broadens our knowledge and enhances our influence.

These approaches can be described as: "Influence to learn, learn to influence." Active engagement with the world allows us to gain new knowledge, and deepening our understanding, in turn, strengthens our ability to influence the world.

Cause-and-Effect Relationships

In the Ideal Mind (IM), hypotheses about cause-and-effect relationships between events are key. These hypotheses are based on personal experience when one event follows another. It's crucial to understand that such hypotheses are subjective and may not reflect the true cause-and-effect relationship in the real world.

In the context of IM, a cause-and-effect relationship is considered a subjective and probabilistic hypothesis. IM does not have information about the objective probability of events but can infer the subjective probability of a cause-and-effect relationship based on how often the consequence follows the event.

Furthermore, IM is capable of registering events related to the manifestation of its own will and forming hypotheses about cause-and-effect relationships, where the manifestation of will is the cause, and changes in the external world are the consequence. These hypotheses and the subjective understanding of the probability of such connections allow IM to assess the degree of its influence on the external world.

Hierarchy of Regularities

The concept of cause-and-effect relationships encompasses not only the detection of patterns in primitive signals but also includes more complex connections at a higher level. These complex regularities can be linked in chains of higher-order cause-and-effect relationships. An important element in this hierarchy is the events related to the manifestation of will, which give these chains of regularities an interactive and synchronized character. This means that interaction and influence on these chains can lead to changes and adaptations in response to external and internal stimuli.

Subjective Sense of Time

The registration of events in the external world and subsequent reflection on them help the Ideal Mind (IM) synchronize with the environment and form its own subjective perception of time. This sense of time can vary depending on how IM interacts with the external world and its internal state at the moment. Thus, depending on the circumstances and current tasks, IM's perception of time can change, reflecting the flexibility and adaptability of its interaction with the environment.

Structure of Cause-and-Effect Relationships

We usually think of cause-and-effect relationships as something simple: one event causes another. However, when viewed more broadly, we can see that sometimes a group of events together can cause another group of events. In this case, all events in each group are combined with the logical operation "AND", meaning all cause events must occur together to cause all consequence events. Further development of this idea leads to the need to consider the "weight" of each event (both from the cause group and the consequence group). This weight will reflect the probabilistic characteristic.

On the other hand, if complex cause-and-effect relationships are mathematically reducible to a hierarchy of simple ones, then for the IM model, simple cause-and-effect pairs are more preferable (postulates 3 and 4).

Significance of Cause-and-Effect Relationships

Cause-and-effect relationships, or hypotheses, possess a certain characteristic, which can be termed as "significance". The significance of a hypothesis is determined by how foundational it is for other hypotheses, especially those of higher complexity or order.

In other words, if a single hypothesis serves as the basis for a multitude of more complex and multi-level assumptions or theories, it is considered more significant. This is because it plays a key role in the structure of knowledge, providing the foundation for a broader range of conclusions and reasoning.

An example would be a basic hypothesis in science that has been experimentally validated and then used to build more complex theories and models. Its significance increases due to its role in supporting and developing a broader understanding of the topic or phenomenon.

Extending this idea, the significance of a hypothesis can also be determined by its ability to explain or predict a large number of phenomena or events. A hypothesis that explains many different, but related, events or phenomena will have high significance within a given theoretical model or thinking system.

Effects of Will (Expanding the Factor of Will)

The manifestation of will in the Ideal Mind (IM) is not confined solely to external changes through body effectors. This phenomenon is much broader and affects both the external environment and internal processes of thinking and perception.

- Internal Dialogue During Deliberation: By mentally pronouncing words, IM activates cause-and-effect hypotheses, which can assist in decision-making or the development of ideas.
- Visualization While Reading: When reading fiction, IM creates images (activating cause-and-effect hypotheses of visual modality), enlivening the text, enriching the reading experience, and facilitating a better understanding and perception of the plot.
- Mental Modeling: In studying physical phenomena or complex systems, IM can mentally model various scenarios and processes, aiding in a deeper understanding and prediction of their behavior.
- Constructing Forecasts and Strategies: The manifestation of Will also includes the ability to plan, forecast, and develop strategies, considering collected data and experience.
- Engaging Critical Thinking: Actively involving critical thinking is part of the process of manifesting will, allowing IM to analyze, question, and reassess received information.

Thus, the manifestation of will in IM not only affects the external world but also plays a key role in internal thinking processes, forming mental models, processing information, and making decisions, demonstrating that internal mental activity can also be directed and controlled consciously.

Optimization and Adaptation

In an abstract model, we might assume the ability to remember and store an infinite number of hypotheses without any limitations. However, in practical implementation, limitations arise, necessitating a method to filter out less significant or less confirmed hypotheses. This is crucial not only due to resource limitations but also because it makes our knowledge more relevant to the real world and our place within it.

Also, the gradual forgetting of old hypotheses, especially when they no longer match changes in the external environment, is an important mechanism of adaptation. Over time, outdated conceptions of world patterns will be replaced with newer, more relevant, and effective hypotheses.

Automatization of the Process (Reactive Behavior and Stereotypical Thinking)

As intellect develops and experience accumulates in understanding environmental patterns and typical reactions, the automatization of repetitive actions occurs. This frees up resources of the single thread of conscious thinking for more complex tasks. Such automatic actions become skills, executed in parallel without requiring conscious attention. As a result, there often isn't conscious recollection of specific details of these automated processes due to a lack of focused attention on them.

Focus of Attention, Patterns of Activity

If thinking resources were unlimited, focusing attention wouldn't be necessary. However, in reality, resources are limited, necessitating concentration on what's important and filtering out the unnecessary. According to Postulate 1, conscious, proactive thinking occurs sequentially, in a single thread. It is activated in situations lacking a suitable skill or stereotype for response. This unique thread of thinking is best utilized in situations where it can have the maximum positive effect or minimize potential harm. Most of the time, routine reactions and actions are automatic, not requiring attention and not distracting the main thread of thinking. The degree of attention focus depends on the context (e.g., reading a book vs. participating in an intense intellectual game) and personal experience (e.g., the difference between an experienced driver and a novice). Additionally, it's notable that proactive thinking is computationally more demanding.

Uniqueness of Experience, Qualia, Language

The process of thinking – dynamic and interactive – creates a unique experience for each individual. Each new experience not only adds to the previous one but is also formed based on it. This is akin to blockchain technology, where information in a new "layer" of experience is encrypted using data from the previous layer. Consequently, an individual cannot fully understand another without knowing their entire previous experience.

This structure of experience ensures the uniqueness of each person's experiences. However, due to similar patterns in the external environment, these experiences can be partially similar among different people. This partial similarity in representations of the external environment or similar reactions to the same external conditions allows for the creation of language. Language acts as a universal set of signals, bridging the unique world models of different individuals.

It should be noted that language, despite its ability to link the unique world models of different individuals, has its limitations. One key limitation of language is the size of its vocabulary and its ability to define only those concepts that are widely spread and accepted in a given environment. This limitation makes language less effective in describing the unique and deeply personal aspects of experiences, known as qualia.

Thus, while language is a powerful tool for information exchange and mutual understanding, it is not always capable of fully conveying the subjective depth and nuances of individual experience. Unique details and nuances of experiences, which are part of each person's personal experience, often remain beyond the scope of formalized language description. This leads to the understanding that many aspects of human experience can be comprehended or appreciated only at the level of intuitive perception or personal experience, but not fully expressed or conveyed through words.

Language, Speech, Writing

Language as a phenomenon serves to create universal signals, facilitating communication between the unique models of personal experience of different individuals. It forms a new foundational layer of "reality" and acts as an interface for interaction, similar to how our body is for the external environment. This additional interface allows for information exchange without direct interaction with the external environment or immediate experience. Causal relationships established through language become more universally applicable among different individuals and are easier to formalize.

Thus, language serves not only as a means of communication but also as a powerful tool for forming and transmitting complex ideas and concepts, making them accessible to a wide range of people.

In the process of thinking at this level, causal relationships with universal identifiers - words - are used. This explains why we often think in words when thinking in the context of information formalized in words. Modern humans receive most new information through language, transforming it into knowledge.

The universality of language and its close connection to the external environment, from specific words to abstractions defined by language, as well as the fact that text is a direct product of human thinking, make it possible to create large language models like ChatGPT. The vast amount of accumulated digital textual data allows these models to behave similarly to human thinking.

Interaction with the External Environment

The Ideal Mind (IM) interacts with the external environment, including the body, through a specialized interface consisting of receptors and effectors. Initially, IM lacks information about the configuration of these receptors and effectors, but gradually learns about their settings through the process of accumulating interaction experience.

When the body detects states or changes in the external environment using receptors, it transmits corresponding signals to the interface. In response, IM can send volitional signals to the interface, which the body is then expected to execute.

If the body receives a message addressed to an effector, it initiates an action and sends back a message about the activation of that effector. However, if the volitional signal is addressed to a receptor, the body responds with the current value registered by that receptor.

Conditions for Differentiation, Optimization, Additional Information Available with Implementation, Emergent Effects

In the thinking process, some aspects are difficult to separate into ideal principles and their practical implementation. Many of these aspects arise from the need to optimize resources. For example, the grouping of receptors in a single sensory organ can be such an optimization, providing additional information or improving its extraction from data.

Furthermore, there are aspects of human thinking for which no special basic algorithms are provided. It is assumed that these effects are emergent, that is, they manifest as the system scales and interacts with the unique external environment (context, stratum of the external environment). Here are examples of potentially emergent effects in thinking:

- Subjective perception of time and consideration of the speed of processes in the external environment when making decisions;
- Development of speech and other social skills;
- Formation of critical thinking;
- Spatial orientation;
- Preference for certain sensory organs;
- Differentiation of one's body from the rest of the surrounding world.

** Further research and development of the IM concept may necessitate adjustments or expansions of the basic principles if certain effects do not manifest emergently.

Comparison with the Brain

In line with the principle of comprehensive generalization, it is necessary to analyze how the concept encompasses (explains) various phenomena of human thinking and the structure of the brain.

- Why do we observe a "ready-made" branched structure of neurons in the brain, whereas the concept implies its formation from scratch?

The brain is a biological platform. Memory effects can be achieved not only by physically forming connections and nodes but also by removing excess ones.

Research shows that a child's brain contains an enormous number of "redundant" (compared to an adult's brain) connections, which actively reduce during early childhood (the period of active learning).

On the other hand, connections between neurons in the brain change very dynamically, and it can be assumed that the dynamics of changing connections correspond to the dynamics of memorizing new information. When implementing IM on digital technology, data storage in memory also occurs by changing the state of existing memory cells.

- Why doesn't the theory explain the effect of sleep?

The phenomenon of sleep is not so much a strength of the brain but rather a limitation. Likely, during wakefulness, the brain accumulates information in short-term memory that requires further processing, normalization, and transfer to long-term memory. A similar mechanism is implemented in the prototype and is carried out by the Normalizer. The essence of the normalization process is merging two or more cause-and-effect pairs having the same cause and effect into one. Such a process is preferably carried out in a state of reduced (minimal) brain activity.

- Why does the process of memorization and learning require interactivity and sequence?

Conceptually, new information represents knowledge about the existence of certain cause-and-effect relationships. For such a pattern to form, both the cause and the effect must already be defined as separate entities. This means that new information can only be effectively assimilated based on already existing knowledge. In other words, without the presence of specific entities of cause and effect, information about a cause-and-effect relationship cannot be fully assimilated and remains just a set of data. Moreover, the aspect of will (active participation) serves as a kind of catalyst for the thinking process: without it, attention can be diverted to other, more relevant abstractions at the moment.

- Are animals intelligent, and why does their intelligence differ from human intelligence?

Animals and humans may possess minds built on the same basic principles, but their intelligence differs for several interrelated reasons. The differences between animal and human minds can be explained by the following factors:

1. Physical Characteristics: Animals' brains and bodies are well-adapted to their environments. Many animals exceed human capabilities in physical abilities such as speed, strength, and agility. This might reduce the need for a complex brain for survival.
2. Instinctive Behavior: Animals rely more on instincts than humans do. Instinctive behavior allows many animal species to be relatively autonomous soon after birth, but it also limits their ability to change and adapt behavior compared to humans.
3. Lack of Verbal Language: Humans have a unique ability to produce complex sounds and form an extensive "vocabulary" of words. This has enabled the development of language, facilitating the transfer of information and knowledge between individuals and generations, aiding the development of culture and thinking.
4. Dependence on Parents: Unlike animals, human children require prolonged care and upbringing from their parents. This allows for the formation of unique behavior better suited to the complex external environment.

Thus, the differences in intelligence between humans and animals are largely due to differences in physical characteristics, instinctive behavior, communication abilities, and social structures. These differences enable humans to develop in ways that are not accessible to most animals.

- How does this theory of consciousness explain the existence of functionally distinct areas of the brain?

It's a misconception to think that the human brain's sections and those of other animals are exclusively designed for specific functions. Rather, this perspective might be a confusion of cause and effect. Within the presented theory of thinking, the brain's division into sections is due to several factors:

1. Specificity of Modalities: Different information processing modalities have unique patterns requiring complex preliminary signal processing. Before information from one modality can

be integrated with other modalities, it undergoes preprocessing stages, forming specialized brain areas.

2. Heterogeneity of External Environment and Processes: The variety of processes and conditions in the external environment, including different submodels (e.g., spatial or social aspects), contributes to the heterogeneous structure of the brain, where different areas specialize in different aspects of perception and information processing.
3. Individual Behavioral Characteristics: Characteristic individual behaviors, such as professional skills, can contribute to the development or enhancement of certain brain areas. For example, certain types of activities can stimulate the development of those brain areas most actively used in that profession.

Thus, the brain's division into functional areas theoretically results from the complex interaction between the specifics of perception, the external environment, and individual behavioral characteristics.

In addition to objective premises from the perspective of the theory of thinking, the brain's topology at the macro level (biological embodiment) is also subject to evolutionary changes, transmitted through generations at the genetic level. This means that the overall structure and functional division of the brain are not only shaped during personal development but also fixed (through selection) in the genetic code that evolved over many generations. Thus, genetic heritage also plays a key role in determining the basic structure of a biological species' brain, which is then influenced by individual characteristics and interaction with the environment.

- How does this theory characterize vision?

According to the theory, vision, like any interaction, is an interactive and emergent process involving the will in the examination of a scene. This process is unique due to the peculiarities of body structure, the external environment in which the species evolved, and the local context in which it gained life experience. However, all the basic principles of the theory remain unchanged. In adults, the process of vision is mostly automated, built on developed skills.

Low-level vision skills in humans (and other animals) are either innate or acquired at the earliest stages of development (including embryonic). These skills include controlling eye movement and synchronization, lens muscle control to focus on an object (area), and identifying primitive shapes and patterns.

Higher-level skills involve strategies for studying a new scene and maintaining the relevance of the existing one.

The process of vision does not have clear temporal boundaries; it continuously adapts to changes in the environment. In everyday life, a complete change of scene is rare, and it usually changes gradually as one moves and changes the direction of gaze.

In different animal species, primitive vision skills and strategies for exploring scenes vary according to anatomical features and environmental needs. For example, some species are adapted to detect motion, while others are better at perceiving static details.

In terms of visual perception strategy, two main scenarios can be distinguished: the gradual change of a familiar scene and the sudden appearance in a new environment. In the first case, the predictability of changes facilitates adaptation to new information. In the second case, intensive attention and rapid assessment of the new environment are required, which may involve more active participation of the higher cognitive process.

Let's examine how scene analysis occurs at the first glance at a completely new, unexplored space:

1. Initial Overview: The gaze covers the entire scene without specific focus on details.
2. Distance Assessment: An automatic estimation of the overall distance to objects in the scene occurs, with vision focusing approximately on the middle distance.
3. Image Preprocessing: Key elements of the scene - areas, lines, angles, and spatial relationships - are unconsciously highlighted.
4. Context Formation: Based on the preprocessed data, an overall impression of the scene is formed, initiating the process of object recognition.
5. Focus on Key Elements: Objects or areas that do not fit the general context or are of interest/threat automatically invoke a volitional impact. This leads to concentrated attention on these elements, including refocusing of vision and in-depth study.
6. Shifting Focus: After analyzing one area, attention shifts to the next most significant one.
7. Completion of Analysis: This process continues until the entire scene is examined, after which the general focus can expand, encompassing the entire space.
8. Maintaining Relevance: In a familiar and predictable environment, where significant changes are not expected, visual perception maintains an overall overview without the need for constant refocusing.

Now let's consider how the visual process occurs when the scene is already familiar, but changes are taking place, for example, related to movement:

1. Familiarity with the Scene: An already studied scene and objects in it allow us to predict potential changes based on previous experience.

2. Preliminary Prediction: Before the actual perception of changes, our attention forms expectations about how these changes might occur. Thus, predictions become part of our intention and expectation.
3. Confirmation of Expectations: When the expected changes actually occur and are recorded by our perception, this confirms our forecasts. In such cases, conscious attention to the process is usually not required, and all work occurs on an automatic level.
4. Adjustment in Case of Discrepancy: If the expected changes do not occur or the scenario develops unexpectedly, this leads to a disruption in our perception of cause-and-effect relationships. In case of minor discrepancies, automatic adjustment of expectations occurs. However, if the changes are significant and do not match our experience, conscious thinking intervention is required to reassess the situation and make decisions.

Overall, the process of visual perception involves many automated skills and as long as changes in the scene match our previous experience, the details of this process do not require active awareness and are generally not remembered.

The process of vision, like any other forms of perception, is an integral part of the overall thinking process, functioning in harmony with it based on common universal principles. When higher-order Will emerges, for example, a desire to move, it triggers a chain reaction of lower-order will, even affecting individual receptors. These receptors are "obliged" to track changes in the environment that are expectedly and intentionally caused by the initial will. The entire process is complex, dynamic, and interactive, occurring continuously and without clearly defined boundaries of beginning and end.

Thinking and Large Language Models

The transfer of information between two thinking subjects is possible only in the presence of a common context, usually represented by a shared external environment. This environment contains common patterns, serving as a reference for encoding and interpreting the transmitted data. Thus, the common context acts as a universal code or protocol, understandable to both the sender and the receiver.

The absence of a common context makes the transfer of information impossible; subjects can exchange data, but it will not be interpreted as intended.

A primitive protocol for information transfer in the presence of a common context can be agreed upon quite quickly at a basic level (for example, gesture language, intonation is even understandable to animals).

However, to transfer information using language (speech, writing), it first needs to be learned. Learning a language involves linking unique concepts derived from interaction with the external environment to universal symbols (words, gestures, text, images). If the linguistic context significantly differs from an individual's experience, the language cannot be learned due to the lack of corresponding concepts.

Nevertheless, in practice, large language models like ChatGPT can perceive, accurately interpret language, and provide meaningful responses, surpassing the average person. This is because the vast amount of digital text can serve as context. The words used in these texts, intertwined into complex patterns, form their own reality, derivative of the external world, but sufficiently self-contained to serve as a common context for information transfer between machine and human.

The Role of Digital Texts as Context for Large Language Models

The vast amount of digital texts used to train large language models like ChatGPT plays a key role in forming a universal context for information exchange between machines and humans. These texts are not merely a collection of words and phrases; they create a complex network of relationships, patterns, and contexts used by models to understand and generate language.

Creating Virtual Experience: Although machines lack personal experience in interacting with the external world, digital texts provide them with a virtual experience. These texts contain descriptions, narratives, dialogues, and opinions that reflect human experience and knowledge.

Training on Diverse Data: Large language models are trained on texts from various sources, including literature, scientific articles, news publications, and everyday conversations. This enables them to capture and utilize a wide range of linguistic structures and semantic connections.

Understanding Context and Nuances: Using these data, language models learn not just linguistic forms but also the context in which these forms are used. They grasp cultural nuances, figures of speech, and even subtleties of human emotions and irony.

Generating Meaningful Responses: As a result, models can generate responses that are not only grammatically correct but also meaningful, appropriate, and sometimes even creatively innovative. This becomes possible by analyzing and mimicking the complex linguistic and contextual patterns found in training texts.

Self-Learning and Adaptation: Over time, with advanced learning algorithms, models can adapt and improve their skills by processing new data and 'learning' from their interactions with users.

Thus, digital texts serve not just as an information source for machines but as a foundation for creating a sort of virtual experience, enabling language models to effectively interact with human users, sharing a common linguistic and cultural context.

Prototype

This prototype represents an approach to embodying the Ideal Mind (IM) theory using classical computer technologies. The main objectives of the prototype development include:

- Demonstrating how abstract ideas and requirements of the theory can be transformed into a concrete implementation.
- Deepening the understanding of key aspects of the theory.
- Identifying and analyzing implementation issues, including contradictions, technical limitations, and mathematical complexity.
- Exploring optimization opportunities, including the use of innovative algorithms or the application of non-traditional microelectronic architectures.
- Testing the overall concept and its individual elements to obtain quantitative performance data.
- Conducting experiments to study changes in qualitative and quantitative characteristics of the system with modifications to the fundamental principles.
- Practically verifying the applicability of the concept and its implementation.

Key Concepts

The model represents a system that includes an information structure and algorithms for working with it (hereafter referred to as IM).

Information in the system consists of events about signal activations and rule applications.

Interaction between IM and the external environment occurs through an interface representing a set of channels with unique identifiers.

Data are transmitted from the body to IM and vice versa through these channels. Data transmission through a channel is called a signal.

The body generates signals when external forces act on receptors and when receiving a signal, the channel of which is linked to an effector, performs a physical action in the external environment.

At the initial moment of its existence, IM only has the ability to perceive signals from the external environment. During development, IM forms experience about signals from channels and the results of reciprocal action (volitional signals).

IM operates with events that are binary (either the event is present or it is not).

In the physical world, the effect on a body's receptor can have an analog component, such as intensity (as an absolute value) and the change in intensity over time (delta). There are two methods of transmitting information about intensity and its changes: through quantification of the absolute value and through conversion to frequency. The presence or absence of analog characteristics is specific to the body and the external environment and cannot be universal. Therefore, the method of encoding analog information is entrusted to the physical body, not part of the IM.

Rules in the system represent information about patterns linking events into cause-effect pairs.

Since events also arise in relation to rule activation, rules can link events of other rules' activations, forming a multilevel hierarchy (graph). At the base of the hierarchy are events about signals (zero level), and other levels of the hierarchy are formed by rules. Entities of the zero level (events about signals), unlike rules, do not have cause and effect but directly indicate the identifier of the interface channel.

The cause-effect connection embodied by a rule is probabilistic and subjective, as the existence of a pattern is only a hypothesis and can be reinforced or weakened only by subjective experience.

If IM registers a pattern repeatedly (the rule already exists), the rule is activated, generating an event about the rule's activation, and this rule can become a cause and/or effect for other rules. The significance of a rule is determined by the number of rules that use it as a cause and/or effect. In other words, the significance of each rule is determined by the number of references to it as a cause or effect in other rules.

System Architecture of the Prototype

The prototype consists of various blocks, each performing unique functions:

1. **Analyzer**: This algorithm analyzes messages in the queue and creates new rules based on observed sequences of events, looking for patterns where one event follows another.
2. **Forgetter**: This module removes rules with minimal significance, controlling the total number of rules to prevent excessive growth.

3. **Solver**: This algorithm examines messages to determine desirable or undesirable consequences. Based on this analysis, it forms volitional messages referencing existing rules.
4. **Normalizer**: This algorithm merges rules with identical causes and effects. It is active during periods of sleep or rest of the system, helping maintain data integrity and order.
5. **Body**: Not an algorithm, but an interface for interaction with the external environment. It receives signals from the external world and transmits them to the message queue, as well as transmitting volitional messages back to the external environment, acting as a link between the Ideal Mind and the physical world. However, since IM initially lacks information about the configuration of the body's receptors and effectors, a "volitional noise" mechanism was introduced. This random generation of volitional signals allows IM to gradually realize the existence and functioning of effectors.

Analyzer

The Analyzer is responsible for creating rules that serve as markers of potential cause-and-effect relationships between events. All events in the system, including incoming signals, volitional messages, and rule activation, are sent to a data pool (FIFO buffer) where they are ordered by the time of registration. This buffer has a limited size.

The Analyzer continuously works on creating new rules based on events in the buffer. The preceding event is considered the cause and the subsequent one the effect. For instance, if there are three events A, B, and C in the buffer, the analyzer forms rules AB, AC, and BC. When such a sequence of events (A,[...],B,[...],C) repeats, the corresponding rules (AB, AC, BC) are activated. This leads to the registration of new events in the buffer and the creation of higher-level rules such as ABAC, ABBC, ACBC.

By creating rules, the Analyzer forms a directed graph where nodes represent rules and edges represent connections between them. Each node has two "downward" connections (cause and effect) but can be part of multiple other connections at a higher level of hierarchy. The significance of a rule increases depending on how many other rules use it as a cause or effect.

Forgetter

During the operation of the Analyzer, numerous new rules are generated. However, most of these rules are incorrect hypotheses about cause-and-effect relationships and are unlikely to be confirmed in the future. Besides the fact that most created hypotheses are unhelpful, there are also physical limitations for their storage (for example, the maximum density of neurons and axons placement in the human brain) and the problem of combinatorial explosion, where the number of rules at the next hierarchy level increases exponentially.

The Forgetter addresses this issue by selectively deleting rules to limit their growth. In addition to signals and patterns, IM can register information about the time between events and the frequency of their occurrence. In the early stages, IM does not have access to absolute time metrics (e.g., the half-life of cesium-133) but can develop its own subjective timescale based on the number of perceived and generated signals.

Rule deletion occurs only in case of their zero significance, i.e., when they do not act as a cause or effect for other rules, to avoid disrupting data integrity. The priority for deletion is given to lower-level rules that have existed the longest (according to IM's subjective time scale) and are located in the most densely populated areas of the graph.

Solver

The Solver is a high-level thinking algorithm whose task is not only to build a subjective model of the world but also to influence the external world to achieve preferable outcomes from the perspective of IM.

The primary function of the Solver is to determine the most preferable outcome and generate a volitional signal to adjust the course of events. To achieve this, the Solver sequentially analyzes events in the message queue, studying the activation of rules or receiving signals from the external environment. The Solver looks for a rule in the graph created by the Analyzer, which acts as the cause of a given event and has higher significance than the current rule and the maximum among alternatives.

If such a rule is found, it becomes the target. Based on IM's subjective experience, the activation of this rule (with a probability greater than zero) may expand the range of probable outcomes, which is considered a favorable development of events. In this case, a volitional event of the target rule's activation is generated. A new rule marked as a volitional action may need to be created. If the impact proves successful, such a rule can be established as a successful solution and used in the future without the Solver's involvement. In case of failure, it can be established as an unsuccessful solution.

As a result, the "will" sequentially spreads to all levels of causes and effects in the hierarchy of rules.

Normalizer

During the Analyzer's operation, a large number of similar rules (i.e., having the same causes and effects) can be created. Although such redundancy makes sense as they may have arisen at different times and in different contexts, these rules can be generalized. The Normalizer merges two or more rules with the same causes and effects into one. Generally, the more significant rule (more often used as a cause or effect) absorbs the less significant one. Absorption is done by switching references to the more significant rule, and the less significant one is removed.

Body

The Body functions as an algorithm that connects the Intellectual Model (IM) with the external environment. Initially, the body is also part of the external environment for IM, but as knowledge progresses, a division occurs between a more controllable part (the body) and a less controllable part (the external environment beyond the body).

The body can be implemented in any form and exist in different external conditions. The interface between the body and IM is a set of input/output channels, each with its unique identifier (number or address). When the body receives a signal from a receptor, it generates a corresponding event in IM. Conversely, when volitional signals appear in the IM message queue, the body executes them if it's a signal to an effector. If it is a volitional signal to a receptor, it is interpreted as a request for information, and the body reads data from the receptor to transmit to IM.

To inform IM about the existence of effectors (since IM is initially unfamiliar with the body's configuration), the body regularly generates special signals similar to volitional signals, referred to as "volitional noise." This allows IM to gradually become aware of and adapt to various elements and capabilities of the body.

Parameters and Hyperparameters

This prototype is based on principles and algorithms logically derived from the initial postulates. Some aspects remain undefined, and implementation details can vary. The choice of specific parameters should be based on practical research of the model combined with a particular body and external environment.

Some parameters may depend on the performance of the used hardware, while others may dynamically change based on IM's internal state.

Currently, the following parameters/hyperparameters can be identified:

1. Message Buffer Size: A larger buffer allows for correlating time-separated events.
2. Message Buffer Structure: Instead of one buffer for all rule levels, different buffers for various groups or a separate buffer for each level could improve data processing.
3. Separate Buffer for Solver: Allocating a distinct buffer for the Solver can clearly separate automatic and conscious processes.
4. Inter-level Link Parameter: This parameter determines how rules of different levels can be interconnected. It can vary for rules of different levels, allowing the creation of inter-level links.
5. Hardware Performance-related Parameters: This set of parameters defines resource distribution among IM's components, data processing strategies, the aggressiveness of the Forgettor, the operating mode of the Normalizer, and various limits and constraints.

These parameters and approaches are necessary for the effective operation of IM in various conditions and can be adapted depending on the specifics of each task and the surrounding environment.

Distinctive Features of This Consciousness Theory Compared to Other Theories and Implementations

This theory of consciousness stands out from other theories and implementations with several unique features:

- Introduction of the concept of "Will": "Will" is considered an integral part of the intellectual process.
- A dynamic structure that is recursively encoded (formed) by the information itself.
- It does not require and does not have a specific purpose; goals of activity emerge in the process of interaction with data.
- Interactivity: The system's operation process is interactive, including interaction with the environment.
- No need for labeled data: The system is independent of a pre-prepared dataset.
- Recurrent element: The system incorporates a recurrent principle, where new information is considered in the context of the previous one.
- Use of a subjective time scale: It allows for the assessment of event development and decision-making speed based on subjectively perceived time, as well as synchronization with the real-time of target processes in the external environment.

- Combination of automaticity and consciousness: Processes in the system can be both automatic and conscious.
- Phenomenon of attention focus: Attention focusing is explained through active direction of the will towards certain hypotheses of regularities.
- Uniqueness of approach: The development of the theory was not based on existing theories of consciousness, representing an independent perspective on the issue.

Implementation and Testing

I began working on the concept in 2009 and immediately started implementing individual elements in code. By 2011, a C++ implementation was developed. The challenge in testing the system is the difficulty in evaluating its effectiveness without clearly defined tasks due to its inherent universality. Some effects can only be assessed on a large scale and over extended training.

The initial goal of coding was to maximally formalize abstract ideas mathematically, check for internal contradictions, evaluate the system's performance on available hardware, and assess the possibility of scaling it to levels comparable to the human brain.

At this stage, several problems were identified and solutions found. A simple graphical interface was then created to display individual system parameters in real time during operation. The number of "hyperparameters," potential limits, and alternative implementations of certain aspects increased.

Automatic tuning of hyperparameters was impossible due to the lack of objective metrics for evaluating performance.

The main focus at that time was on the Analyzer, Forgetter, and Normalizer. Productivity (the number of processed signals per unit of time), the topology of the emerging rule graph, the speed of reaching the overall limit of rules, the lifespan of rules (maximum, minimum, average), and other quantitative indicators were analyzed. The compressibility of the final model dump (using lossless archivers) was evaluated depending on the initial data provided to the model and various hyperparameters.

After some time, I began to hypothesize that the Resolver plays a key role not only in the "use" of the model but also in its training process. That is, the training process should be interactive, and the volitional component is essential, making it a single, indivisible process in time.

Consequently, further testing required the development of an emulator of the external environment and a body living within it.

At this stage of modeling/testing, the following conclusions were drawn:

1. The theoretical concept is formalized to the level of code, implementable as an algorithm, and does not contain logical contradictions.
2. The implementation can be executed on the current state of technology hardware in real-time and at a scale comparable to the human brain.
3. To obtain valid (plausible, real) results of the work that make sense to analyze, merely labeled (benchmark) data is insufficient. A Body and an external environment (even virtual) are necessary, with which the IM will interact interactively from the very beginning of its development.

Application and Potential

At this stage, the concept and model have become tools for research in the field of consciousness theory and the construction of AGI (Artificial General Intelligence). It provides comprehensive, formally precise assumptions about the phenomenon of intelligence, enabling the exploration of its various aspects in new metrics and spaces.

The implementation of a certain body and the placement of the system (IM + body) in a virtual world will allow for systematic analysis of various hyperparameters and implementation variations.

Determining the dependencies between selected hyperparameters and certain metrics will facilitate more precise assumptions about individual aspects of intelligence. It is also necessary to identify and analyze similar patterns of structures that emerge during operation.

The model can be implemented as an intermediary layer in a classical neural network to investigate its applicability for solving pre-defined tasks.

I hypothesize that this theory opens a new pathway for research in the field of creating full-fledged AGI, or the prototype can be a fundamental technology (a set of algorithms) for AGI creation.

Further Research and Development

I envision the following pathways for the development of the concept and its implementation:

1. Development of a world and body of minimal necessary complexity to enable numerical evaluation of various hyperparameters, followed by systematic research and testing. This will accelerate concept development.
2. Modeling the rule graph and analyzing it at a micro-level to identify minimal recurring structures. If such structures are found and classified, their number and composition could serve as a metric of intelligence. Using these metrics should allow for internal evaluation of the system's functioning and adjustment of hyperparameters to achieve or maintain a specific quantitative composition of these structures.
3. Investigating the possibility of using this concept's implementation as a layer in a classical deep learning neural network.

Areas requiring additional research include:

1. Exploring alternative ways of "descending will" in conjunction with the "boundary" between automatic and conscious (behavior).
2. Seeking and evaluating alternative methods of mapping the "analog" external world to a binary set of causality rules.
3. Researching the need to add concept parameters that influence the entire IM, analogous to hormones and neurotransmitters.

Kyiv, 2024