ARC Journal of Neuroscience
Volume 8, Issue 1 2025, PP 23-38
ISSN 2456-057X
https://doi.org/10.20431/2456-057X.0801004
www.arcjournals.org

How the Brain Works – Insights for AI Systems Analysts

Andrzej Brodziak[1], Filip Romaniuk[2], Danuta Abram[1]

[1]*University of Applied Sciences, Nysa, Poland*

[2]*Delft University of Technology, Architecture and the Built Environment Department, Netherlands.*

Received: 03 July 2025 Accepted: 18 July 2025 Published: 24 July 2025

Corresponding Author:** ***Andrzej Brodziak, *University of Applied Sciences, Nysa, Poland*

Abstract

Given the author's decades of publications in neurophysiology, this text proposes a brief conceptual synthesis intended to be useful for educational purposes. The following article is not a research paper; rather, it serves as a commentary on the current state of knowledge. The article offers a concise synthesis of fundamental neurophysiological concepts, aiming to bridge neuroscience and artificial intelligence for educational use. Instead of presenting original research, it provides an intuitive overview of how the human brain performs perception, mental imagery, problem-solving, and self-awareness — presented in a manner accessible to both medical professionals and computer scientists. Key models are highlighted, including neurons as "integrate-and-fire" units; hierarchical neural circuits underlying perception and memory; dual memory loops (hippocampal-cortical and limbic); and the distributed function of the brain's speech areas. The author also reviews theories positing the brain's endogenous electromagnetic field and the controversial hypothesis of quantum processing in neuronal microtubules (Orch OR) as possible substrates of consciousness. The paper draws explicit parallels between human cognitive processes and contemporary AI systems. For example, certain AI models can now generate images from language in ways that functionally resemble the brain's translation of words into mental imagery—though, crucially, AI lacks consciousness and embodiment. The author proposes that comparisons between brain function and AI systems should focus on four priority processes: language and speech, problem-solving mechanisms, the formation of mental images, and AI-*driven image generation from linguistic prompts. This synthesis provides a practical, conceptual framework for understanding and comparing cognition in brains and machines, advocating for interdisciplinary clarity and intuitive understanding over excessive technical detail.*

1. Introduction

Thousands of studies have been published explaining the structure and function of the brain [1-5]. However, in the context of the rapid development of so-called artificial intelligence, there is now a need for a concise, illustrative, and intuitively understandable overview — one that would enable individuals with limited background in neuroscience to engage meaningfully with the growing body of discussions comparing the nature of brain function to the operation of AI systems. Such a need arises especially among people with medical education, who are familiar with the basic structure and functioning of the brain, and who are now also users of artificial intelligence systems.

Understanding "how the brain works" is central also to neuroscience, cognitive science, and medicine. A clear, intuitive model is particularly necessary for education and interdisciplinary research.

This article draws from a series of previous works by the authors, providing a coherent and visually supported explanation of the neural underpinnings of memory, perception, mental imagery, functioning of the brain's language areas**,** problem solving and self-awareness [6-12].

The article is not a research paper, nor is it even a comprehensive review. Its purpose is merely to compile and juxtapose fairly general descriptions of how the human brain and AI systems carry out selected, important, analogous cognitive processes.

The aim of this comparison is to stimulate inquiry into the extent to which these mechanisms may lead to comparable results. Such inquiries should be helpful both for neuroscientists and computer scientists.

2. Neurons as Integrate-and-Fire Elements

All neurons maintain a resting membrane potential via ionic pumps. Incoming synaptic inputs modulate this potential; when the threshold is reached, an **action potential** (a ~1 ms voltage spike) is generated and propagated along the axon to downstream synapses.This **“integrate-and-fire”** mechanism underlies all neural signaling. Repeated high-frequency firing (up to ~500 Hz) encodes stimulus intensity, while synaptic “weights” are adjusted via plasticity, forming the basis of learning.

Contemporary knowledge about information transmission during this “firing” is extensive, but for the educational purposes of this article, a symbolic, simplified illustration of this process using Figure 1 will suffice. Such a simple model of the neuron derives from the original concepts proposed by M.S. McCulloch and W. Pitts [13].

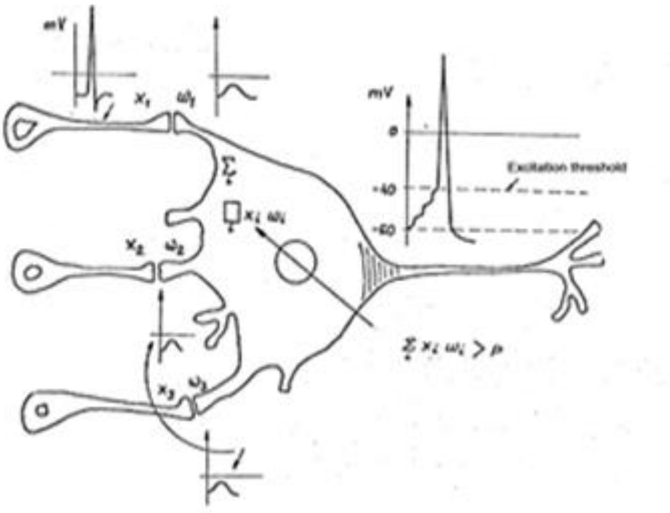

Figure 1. *The resting potential of a neuron is altered by impulses arriving through synapses from lower-level neurons. The release of neurotransmitter by the presynaptic synapse causes a certain increase (or decrease) in the resting potential of the receiving neuron's interior. The neuron continuously “sums up” these changes occurring across its entire surface and over a certain time interval. If the sum of these changes exceeds the excitation threshold, an action potential is generated in the region of the axon hillock and then propagated along the axon and its branches.*

3. Hierarchical Neural Circuits and “Object Neurons”

Perception and memory in the brain are realized via hierarchical networks. The experimental foundations for the findings concerning these hierarchical structures tuned to detect specific patterns within the receptive field were established as early as the 1960s [14].

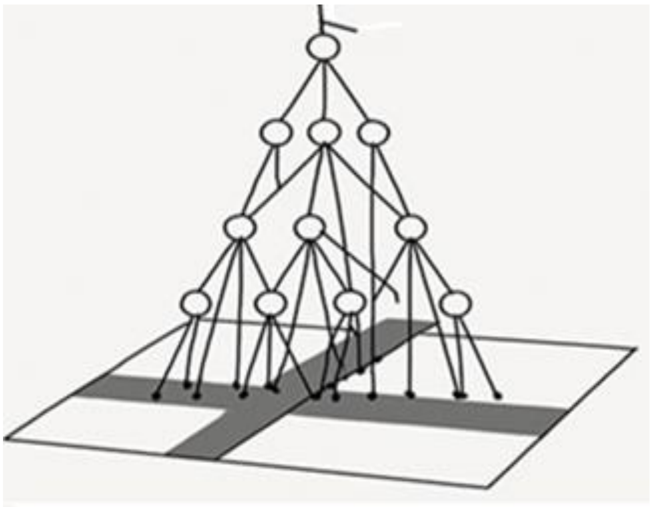

Figure 2. *A neuron connected via active receptive synapses to two “neurons” sensitive to the image of a line will itself become sensitive to the image of a “crossing of lines.” Experimental work on cats by David Hubel and Torsten Wiesel demonstrated that neurons in the striate cortex of the visual pathway — at the third level of the hierarchical structure — process information in this manner.*

Sensory information, for example visual input, ascends from primary sensory cortices (such as the occipital lobe for vision) to higher-order areas in the temporal and parietal lobes. At the apex of this hierarchy reside so-called "object neurons" (also referred to as "concept neurons") — specialized neurons that encode complex representations, such as familiar faces or objects [15-18].

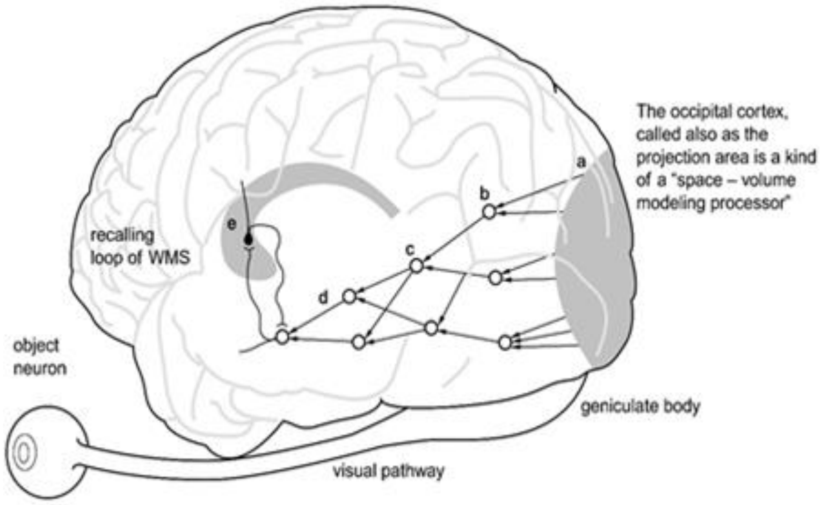
The occipital cortex, called also as the projection area is a kind of a "space – volume modeling processor"
a
b
c
d
e
recalling loop of WMS
object neuron
geniculate body
visual pathway

Figure 3. *The visual pathway proceeds from the retina to the occipital cortex, then continues to the anterior temporal lobe where object neurons are found. These neurons are interconnected with memory loops, allowing both recognition and recall.*

Importantly, these high-level neurons are not single units; multiple functionally equivalent neurons may represent the same object, enabling parallel processing of various features such as shape, color, or texture.

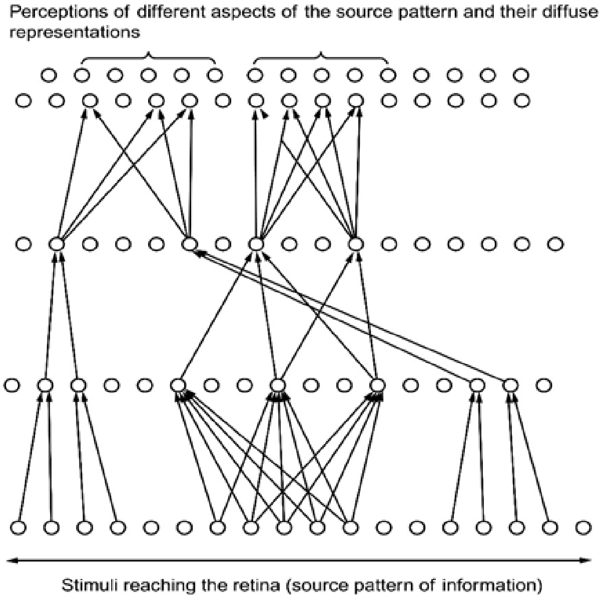
Perceptions of different aspects of the source pattern and their diffuse representations
Stimuli reaching the retina (source pattern of information)

Figure 4. *Parallel processing of object features occurs. Visual information is processed in parallel streams — shape, color, texture, and size. At the top of this hierarchy, multiple "concept neurons" encode different aspects of the same object, supporting robust recognition and flexible recall.*

4. The Dual-Loop Model: Memory Consolidation and Mental Imagery

Memory and imagery depend on two interacting neural circuits:

Loop 1: Hippocampal-Cortical Loop

Essential for the temporary recall of mental images (working memory, episodic memory). Activation of object neurons triggers oscillatory, recurrent excitation with the hippocampus, sustaining imagery and problem-solving.

Loop 2: Amygdaloid/Limbic Loop

Involved in consolidation of long-term memory, especially when emotions are engaged. Pathways from the hypothalamus and amygdala facilitate the strengthening of synaptic connections.

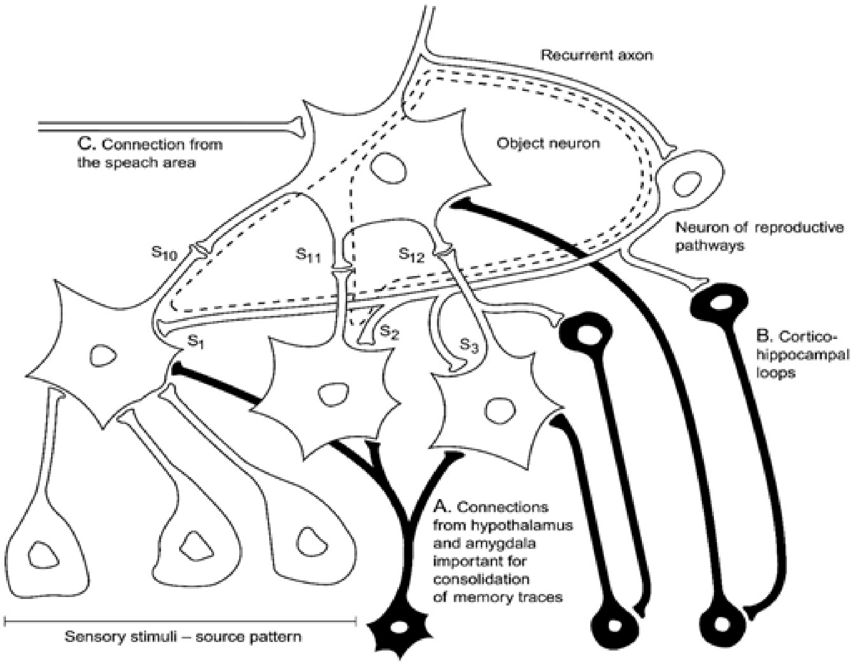
Recurrent axon
C. Connection from the speach area
Object neuron
Neuron of reproductive pathways
S10
S11
S12
S1
S2
S3
B. Cortico-hippocampal loops
A. Connections from hypothalamus and amygdala important for consolidation of memory traces
Sensory stimuli – source pattern

Figure 5. *The illustration how two loops connect object neurons with both the hippocampus (for recall and working memory) and limbic structures (for emotional encoding and long-term memory consolidation). Three layers of the hierarchical structure of neurons integrating sensory information, are discerned. The traces of long term memory are consolidated under the influence of connections from hypothalamus and amygdale (A). There are consolidated through two mappings: weights of receiving, ascending synapses (s10, s11, s12) and synaptic weights of reproducing connections (s1, s2, s3). The second, cortico - hippocampal loop (B) is necessary for temporary recalling of the mental image. The activation of the object neuron by this indexing loop causes the recurrent reactivation of neurons in lower layers. The dotted line indicate the pathways of repetitive circulations of stimuli in the upper layers of the hierarchical structure, what is the essence of mental imaginary. The object neuron can be activated also from the side of speech area (C). Mental imagery is the essence of episodic memory, short term memory and is used for working memory activity.*

Top-down pathways from the prefrontal cortex and other higher cortical regions enable the brain to activate lower levels of the hierarchical structures, which is essential for mental imagery [19–24].

5. The Neural Basis of Mental Imagery

Neuroimaging and electrophysiological studies confirm that mental imagery and actual perception share largely the same neural substrates. When an object neuron is activated (e.g., by hearing the name of an object), impulses circulate up and down the hierarchy, reactivating the perceptual "trace" of that object. The oscillations (circulation) through top-down pathways are the physical substrate of imagination [19-24]. Key evidence for such a model comes from fMRI, EEG, and MEG studies, which show overlapping in neural processing during perception and imagery, particularly in visual, parietal, and frontal cortex [25]. The vividness of mental imagery depends on the excitability of early visual areas. The maintenance of mental images (even in the absence of stimuli) relies on the cortico-hippocampal loops. Emotionally salient memories engage the amygdaloid loop, consolidating information for long-term retention.

6. Complex Scenes and Self-Imagination

Mental imagery is not limited to simple objects — it also encompasses complex scenarios, autobiographical memories, and future event simulation. Neural correlates of these processes include the hippocampus, prefrontal cortex, posterior cingulate cortex, angular gyrus, and other association areas. Construction of future scenarios is mediated especially by the right anterior hippocampus. It should already be noted here that the concept of "imagining a particular object or a complex situation" can be generalized to include "imagining oneself." Such self-imagination requires the activation of brain regions responsible for storing autobiographical data [25, 27].

7. The Speech Area in the Human Brain

The speech area consists of three structures of the cerebral cortex located in the dominant hemisphere (most often the left): (1) the Broca's area (Brodmann 44/45, posterior part of the inferior frontal gyrus), (2) the Wernicke's area (Brodmann 22, posterior part of the superior temporal gyrus), and (3) the arcuate fasciculus [28, 29].

Broca's area is responsible for speech production — it plans, coordinates, and sends neural impulses to the muscles of the speech apparatus (the tongue, lips, and larynx). Wernicke's area, on the other hand, is responsible for speech comprehension — it analyzes sounds, decodes words, and assigns them meaning. Damage to Wernicke's area causes a type of aphasia in which speech remains fluent but loses sense and is accompanied by difficulties in understanding language. The third structure, the arcuate fasciculus, connects both centers, enabling repetition and the smooth exchange of information between "understood" and "produced" speech. Both main centers also cooperate with other brain regions, such as those responsible for memory, emotions, and motor functions.

8. The Function of Speech Area Structures in Light of Current Knowledge

The structures of the speech area (mainly Broca's and Wernicke's areas) are not repositories for word patterns — these patterns, both written and phonetic, are stored in other specialized regions of the brain, such as the temporal cortex and occipito-temporal areas [28].

The speech area instead serves as an **integrator and coordinator** — it is responsible for analyzing, processing, and assembling perceived language elements into meaningful, grammatically correct utterances. In particular, Broca's area is crucial for recognizing and constructing the grammar of sentences, both during comprehension and production of speech. Wernicke's area analyzes the meaning of words within the context of the sentence.

In summary, **the speech area does not store words or phrases**, but integrates received patterns and ensures that both the comprehension and production of sentences comply with the rules of natural language grammar [28,29]. To a large extent the meaning of words and sentences in natural language is determined by the mental images evoked by their perception, but not exclusively.

Contemporary cognitive science and neuroscience increasingly indicate that the meaning of words and sentences is linked to perceptual and experiential imagery. For example the word "apple" activates not only language-processing areas in the brain, but also those responsible for vision, taste, touch, and even movement - like reaching for a fruit. What we understand as "meaning" is, in fact, a network of sensory, motor, and emotional associations — triggered when perceiving a word or sentence.

When you receive a sentence, your brain "simulates" the situation it describes. If you hear, "The cat lay down on my bed" not only your language centers are activated, but also your imagery – related areas – you can "see" this scene in your mind's eye. The meaning of sentences or some words can also be purely abstract (for example, "democracy" or "number"), and in such cases, the imagery may be less clear, though still often based on prior experience or metaphor. For people who are blind from birth, the meaning of words like "to see" or "brightness" is built on experiences other than vision. "To see" may, for them, mean "to understand something," "to get to know," or "to experience" through touch, hearing, or even conversation. "Brightness" may be associated with something pleasant, light, warm, or with a feeling evoked by a voice or a sound—for instance, a "bright" sound produced by a musical instrument.

It can be said that the meaning of words and sentences in the brain is largely determined by imagery and a network of sensorimotor associations that are activated when these words are perceived. This explains why language comprehension is so deeply rooted in our experience

9. The Role of the Brain's Endogenous Electromagnetic Field

There is growing recognition of the theories of consciousness related to an electromagnetic field [30-35]. Recently Johnjoe McFadden published in a trustworthy journal the convincing justification for his theory, developed over twenty years, which he denotes as "the conscious electromagnetic information field theory (cemi)" [31].

An endogenous magnetic field is created within the brain, which is recorded during magnetoencephalography. There are closed electrical circuits in brain tissue. They are considered when we take into account the pathways running backwards to the lower levels.

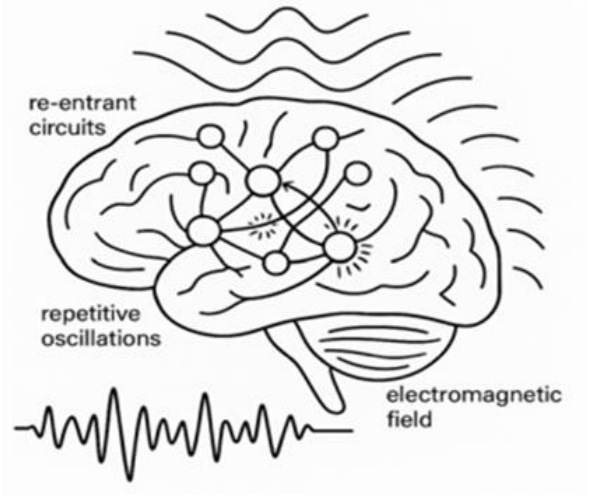

Figure 6. *Schematic symbolic illustration highlighting the existence of the brain's endogenous electromagnetic field.*

The authors of theories of consciousness related to an electromagnetic field assume that the basic process evoking this phenomenon should consist not only in actions occurring over time but also causing the formation of a certain spatial wholeness [31].

Johnjoe McFadden points out that neural networks realize step by step time-progressive process. He argues that when looking for a physical medium supporting something that has the feature of a spatial structure, one can appeal to the endogenous brain's electromagnetic field [31].

The electric phenomena are routinely recorded by electroencephalography (EEG) and the EM field is also routinely detected, assessed and recorded by magnetoencephalography (MEG). The authors of "the conscious electromagnetic information field theory" indicating the role of the brain's electromagnetic field emphasizes also that there is a massive synchronization of neuronal activity and that the repetitive oscillations in neuronal circuits occur. They emphasizes the significance of the "re-entrant circuits, essentially closed loops of neuronal activity whereby neuronal outputs are fed back into input neurons" [31, 12].

10. Microtubule Quantum Processing - Hameroff and Penrose's 'Orch OR' Theory

Stewart Hameroff and Roger Penrose promote their "Orchestrated Objective Reduction (Orch OR) Theory", which is based on the assumption of the existence of quantum processes within neuronal microtubules — tiny structures composed of tubulin protein dimers capable of forming quantum superpositions (qubits) [36-45].

In Orch OR, it is proposed that these microtubules undergo cycles of quantum coherence ("orchestration"), interrupted by abrupt "objective reductions" (OR), which correspond to discrete moments of conscious experience (so-called "conscious moments"). These events are thought to occur at rates of roughly 24–90 Hz, aligning with gamma oscillations and other brain rhythms observed in EEG.

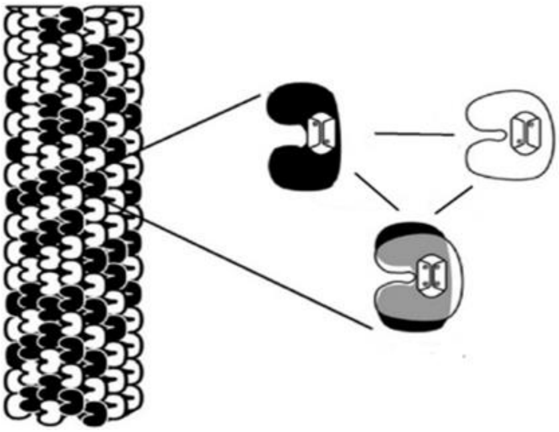

Figure 7. *Symbolic presentation of a microtubule, composed of tubulin dimers existing in three different states. The tubulin dimers due to the positions of delocalized electrons can be in two different states and additionally in a quantum superposition state. On the left a symbolic presentation of a microtubule composed of tubulin dimers existing in three different states.*

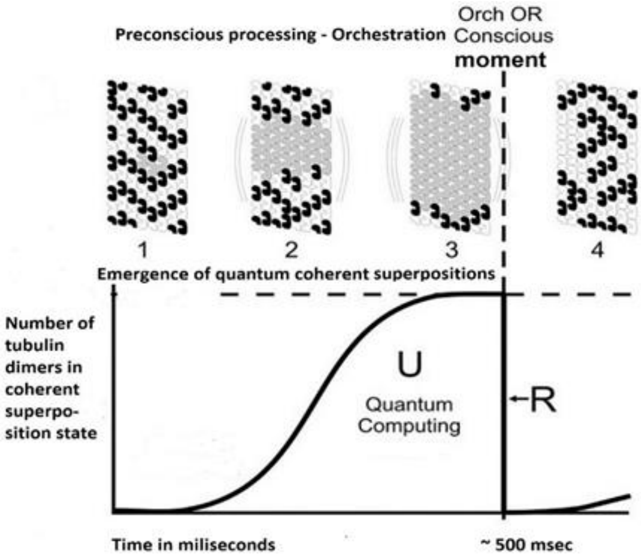
Preconscious processing - Orchestration
Orch OR Conscious moment
1
2
3
4
Emergence of quantum coherent superpositions
Number of tubulin dimers in coherent superpo-sition state
U
Quantum Computing
←R
Time in miliseconds
~ 500 msec

Figure 8. *Symbolic illustration of the essence of Roger Penrose and Stuart Harmeroff's 'Orch OR' theory. The authors assume that in the microtubules of neurons repetitive, cyclic quantum information processing is realized, which consists in increasing the quantum coherence of the tubulin dimers, interrupted by the OR operation, causing the conscious moment. In larger synchronized sets of neurons, the termination of orchestration by OR moment occurs after approx. 300 to 500 ms.*

A unique aspect of the theory is the claim that quantum entanglement and tunneling may synchronize activity across large regions of the brain, facilitated by structures such as gap junctions. This quantum synchronization is proposed to be critical for the emergence of unified consciousness and the generation of qualia – the vivid sensory qualities of experience. The theory contends that these non-computable quantum effects are necessary for consciousness and cannot be replicated by classical computational systems. Critics often highlight the difficulty of sustaining quantum coherence in the "warm, wet, and noisy" environment of the brain; however, the authors point to evidence of quantum effects in biological systems (e.g., photosynthesis, avian navigation) and experimental data showing microtubular resonance.

Ultimately, Orch OR posits that consciousness arises not just from classical neural computation, but from orchestrated quantum events within microtubules, leading to the subjective richness of conscious life and possibly accounting for self-awareness.

All general discussions of brain function should take into account – and attempt to evaluate — the validity of Hameroff and Penrose's 'Orch OR' Theory, due to the necessity of addressing the so-called Chalmers' hard problem. Chalmers' "hard problem" of consciousness refers to the question of how and why subjective experiences — feelings, sensations, and qualia (for example, colors, tastes, and smells) – arise from physical processes in the brain.

While neuroscience can explain brain functions and behaviors ("easy problems"), it struggles to explain why these processes are accompanied by vivid "colorful" subjective experiences.

11. Emergence of Self-Awareness and Consciousness

The feeling of identity, also the understanding of the word "I" is based on autobiographical memory [26, 27, 46]. The comprehension of the word "I" requires to recall "who I am". This, nonetheless, is established by the memorized biography.

The proposal of the nature of the feeling of one's self can be derived from the presented above theoretical model of imagery. We described a structure forming the mental image of a certain object. Probably the feeling of one's self appears when there is a perception of one's own body and environment. In order to accurately imagine yourself, it is additionally necessary to imagine one's past and recall images of one's environment, which also allow a person to imagine his foreseeable future.

On the other hand, Rodolfo Llinas and Georg Northoff convinced most contemporary neuroscientists that one of the basic mechanisms of consciousness consists in the self-excitation of a neuronal network composed of thalamocortical pathways. They describe this activity as re-entrant processing [47, 48]. Probably the existence of neural circuits which trigger themselves is a very basic rule of nervous systems. It should also be noted here that McFadden, the author of the cemi theory, emphasized that the key “signature of consciousness” is re-entrant processing in the parietal and frontal areas [31, 35]. Metaphorically speaking, one could propose that the equivalent of "object neurons" for the word "I" are structures located "high up" - that is, in the parietal, prefrontal and frontal lobes.

Thus, the integration of the theories mentioned facilitates a conclusive answer to the questions about the nature of self-awareness. Only multi-level hierarchical data processing, embracing the areas placed in the parietal and prefrontal lobes makes it possible to pass the mirror test successfully. It appears that a persuasive explanation of the phenomenon requires, apart from an indication of neural substrates, also a description of the role of a spatial field, namely of an electromagnetic field, which carries energy and has a particular shape.

Moreover, the attempt to explain what self-awareness is leads to the conclusion that first it is indispensable to explain how mental images are realized. Next, it involves the necessity for an explanation of what the ‘image of oneself’ is.

12. The Process of Searching for Solutions to Problems that are Perceived or Presented in Interactions

The processes of searching for solutions of perceived or externally presented problems involve the coordinated activation of several key brain areas [1, 2, 3, 6, 49-51]. The most important is here the prefrontal cortex, especially the dorsolateral prefrontal cortex, responsible for planning, abstract thinking, hypothesis testing, and working memory. The default mode network becomes engaged during introspection, analysis from various perspectives, and creative idea generation. Other regions – including the parietal cortex (for information integration), the hippocampus (for recalling memories and experiences), and the anterior cingulate cortex (for attention control and strategy selection) – also participate.

Throughout problem solving, oscillatory loops between higher-order association cortices and lower brain areas, mediated by top-down pathways, enable the generation and testing of mental solutions. The moment of sudden insight is linked to increased activation in the right temporal lobe and transient synchronization across multiple brain networks.

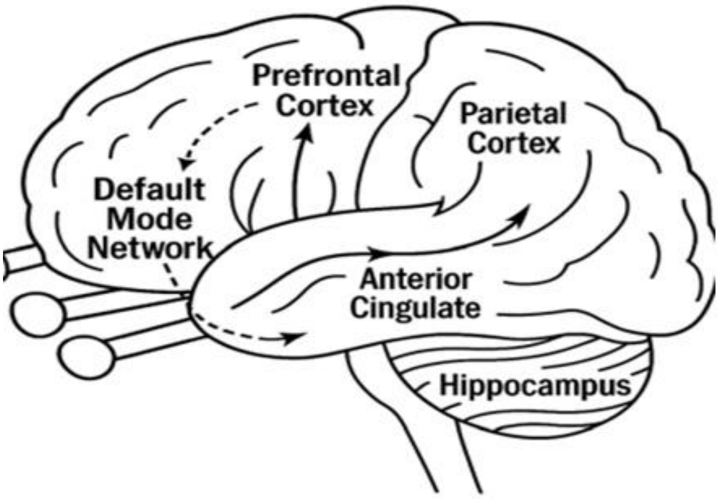

Figure 9. *Schematic symbolic illustration highlighting the sophistication of the process of searching for solutions*

In summary, problem solving relies on dynamic cooperation among distributed brain areas, especially the prefrontal cortex and associative networks, using neural oscillations and recurrent circuits to create, test, and select solutions. The essence of data processing in the course of problem-solving can be understood as searching for a “pathway” from the perceived current situation (reinforced by its mental image) to the mental representation (mental image) of a desired situation, which constitutes the sought-after solution to the problem [6]. Typically, such a “transition” from the current situation to the desired situation is not achieved through a single action. Usually, it is necessary to reach intermediate stages that establish a so-called "transition path," which is implemented through a sequence of actions bringing one closer to the desired situation. This is illustrated metaphorically in Figures 10 and 11.

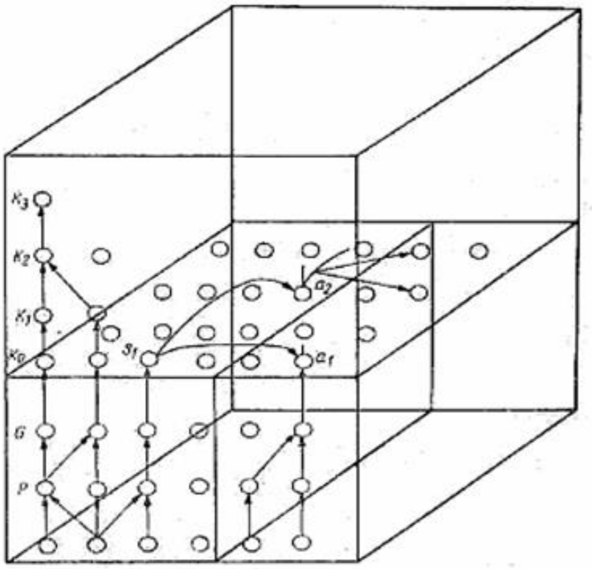

Figure 10. *There are connections between object neurons or neurons of the highest-level structures representing perception and mental imagery of different situations*

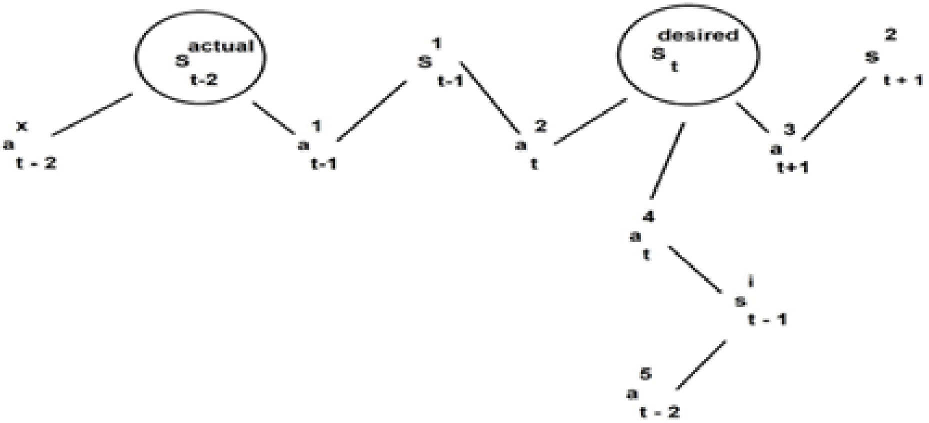

Figure 11. *The diagram will help clarify the presented inference, provided the reader engages in an act of imagination. First, it is necessary to recall the key message of Figure 2, which illustrates how the image of a particular object is perceived within a hierarchical neural structure. Next, one should refer to Figure 10 to understand that there are "horizontal" connections between the highest-level neurons of these hierarchical structures, linking them to other top-level neurons within the cortex. Only then, by examining Figure 11, can we understand that it is possible to distinguish structures representing both the current situation (S actual) and the desired situation (S desired). The mental process of problem-solving consists in searching for actions (a1, a2...) that gradually transform the current situation into imagined states (S) that increasing lresemble the desired situation.*

13. Examples of Ongoing Contemporary Reflections on the Cognitive Processes Occurring in the Human Brain and AI Systems

In the scientific literature – whether in the natural and medical sciences or in technical and computer science fields – there are not many works comparing the mechanisms underlying cognitive processes in the human brain with those occurring in AI systems. Therefore, even to such a fundamental and crucial question as whether the essence of human brain function, which involves operations on mental imaginations first activated by verbalizing them through appropriate words, phrases, or sentences in natural language, also takes place in AI systems – we do not find a definitive answer.

Explanations encountered on this subject usually go as follows: When a person hears or reads a sentence, their brain not only decodes the grammatical structure, but almost automatically activates a network of mental images, sensory associations, and emotions. For example, the sentence "A child is eating ice cream on the beach" evokes in the mind images of taste, texture, smell, the sound of the sea, and even memories of personal experiences. Thus, meaning is deeply embodied — it is based on one's own experiences, imagination, context, and emotions. Mental images can be very vivid: the word "to run" may evoke in the imagination the image of one's own legs in motion, the rustling of grass, or even a slight acceleration of heartbeat.

The phrase "heated discussion" will be understood not only as a conversation, but also as a situation of tension, emotion, sometimes even a "charged" atmosphere. An analogous process in artificial intelligence systems works differently. A so-called Large Language Model (LLM), when it receives a prompt or sentence, does not activate imagination, but form its vector representation and analyzes the statistical correlations between similar representations learned during training on huge text corpora. Here, "understanding" is indirect – it derives solely from the analysis of language patterns, without any physical grounding for objects. The essence of data processing may be somewhat different in graphical, image-generating artificial intelligence systems. The undersigned, the author of this article, perhaps as one of the first, puts forward the possibility of considering whether those AI systems that generate images based on linguistically expressed prompts are not, at least in an embryonic form, performing a process analogous to the formation of mental imagery. The present attempt — made within this article — to address this question begins from providing a description of the "data processing operations" that occur during such image generation. In general, one can arrive at the following description of this operation:

14. Image generation by AI systems based on linguistic descriptions

The process of generating images from textual prompts in contemporary AI systems involves several key stages [52-55]:

1. Text encoding

The user provides a description (prompt) in natural language (e.g., "a woman reading a book in a park on a sunny day"). This prompt is processed by a language encoder, often based on so called "transformer" architecture, to extract semantic features and encode them into high-dimensional vectors in a shared latent space.

Transformers process entire sentences or paragraphs at once (rather than word by word), allowing them to capture the relationships and context between words effectively. Semantic features are elements of meaning extracted from text. When a model analyzes a sentence, it identifies the core ideas, topics, objects, relationships, and attributes – essentially, what the text is "about". Latent space is a high-dimensional mathematical space where complex information (like images or text) is represented as vectors (arrays of numbers). In this space, similar concepts are positioned close together. By mapping both language and images into the same latent space, an AI can "translate" between them — for example, turning a sentence into an image that matches its meaning.

2. Conditioning in Latent Space

These text-derived embeddings guide the generative model, aligning semantic concepts from text and images within the model's latent space. This step ensures that similar ideas are located near one another, making it possible for the model to "translate" language into visuals.

3. Image Generation Via Some Kinds of Artificial Networks

3.1. Diffusion Models: State-of-the-art systems like DALL-E 2/3 and Stable Diffusion use diffusion models, which start from random noise and iteratively "denoise" the image, each step guided by the text embedding.

Earlier models used Generative Adversarial Networks **(GANs),** where a generator and discriminator compete; however, GANs are now less common for text-to-image generation. Cross-attention layers enable the model to directly link textual phrases to corresponding image regions, improving compositionality and fidelity.

4. *Output and User Feedback:* The final image is produced and presented to the user. If necessary, the user may adjust the prompt or select among multiple outputs. In deployed systems, post-processing steps (such as upscaling or content moderation) may be applied.

One May Therefore Ask: Do LLM systems that cooperate with image-generating AI produce something analogous to the mental imagery created in the human brain? This is a very timely and important question, touching the core of modern AI research and the boundaries between "simulated imagination" and the genuine mental imagery known from the human mind. The current answer seems to be as follows:

Modern large language models (LLMs), such as GPT-4.1, Gemini, or Claude, are capable of generating highly detailed scene descriptions and instructions, which are then passed on to image-generating systems (like DALL-E, Midjourney, or Stable Diffusion).

Based on these instructions, a graphic is produced that corresponds to the prompt's content. This process does resemble, to some extent, the way the human brain creates an image based on language.

15. There are Notable Similarities to Human Imagination

When transforming language into images, the LLM analyzes the prompt, breaks it down into meaningful elements (objects, relationships, attributes), and passes these as a set of instructions to the image model. Similarly, the human brain can generate a mental image from a sentence. Additionally, AI systems are increasingly integrating text, image, and even sound processing — creating a form of multimodal representation, much like the human brain does when imagining a scene described in language.

16. However, it is Essential to Emphasize the Differences

AI has no consciousness or introspection — it does not "see" these images nor does it experience them. It creates a symbolic (vector) representation, which "the image generating system" converts then into a picture, but there is no "mental image" – "seen" as it happens in the human mind.

Furthermore, AI lacks embodiment: for humans, imagination is grounded in sensory, emotional, and motor experience. AI operates only on statistical patterns memorized in a high-dimensional mathematical space. It does not feel color, movement, taste, or the atmosphere of a scene.

In the human brain, imagery involves activation of perceptual areas, such as the visual cortex; so the neuronal structures the same which participated during perceptions. In AI "image generation systems" it is simply the processing of data in an abstract feature space.

One could say however that AI creates a "functional equivalent" of imagination: a complex, internal representation of the description that can be transformed into an image – even though this process is symbolic and statistical only, not phenomenal — there is no experience of "seeing."

So, LLMs combined with image-generating AI create a functional equivalent of imagination, but do not experience it as the human brain does — they lack subjectivity and embodiment. Nevertheless, their ability to translate language into images is increasingly approaching the human capacity for "visualizing" word content.

17. The Essence of Problem-Solving Operations in Contemporary AI Systems

Data processing in AI systems during the search for solutions to presented problems is described in numerous scientific articles [56-58]. This process typically consists of several stages and can be generally described as follows:

1. AI systems first analyze the tasks they are given by a textual description, command, question, or a combination of inputs (e.g., text and image). The formulation of the task is referred to as the "prompt." Advanced language models take into account the context, intent, and constraints of the problem.
2. Modern AI models leverage vast knowledge bases. They process the prompt by mapping it into semantic representations (embeddings), which enable comparison, classification, and link various aspects of the problem.
3. AI systems use mechanisms of sequential reasoning, such as chain-of-thought, scratchpad, generating intermediate reasoning steps. They may use planning tools and act step by step and use internal "agents" that decompose the problem into smaller parts.
4. Based on the inferred solution pathway, the model generates a response – this could be text, image, code, an action plan, etc. In multimodal systems, different modalities (text, image, audio, code) can be combined.
5. The most advanced systems can iteratively assess their own solutions through self-consistency checking, reflection, RLAIF (Reinforcement Learning from AI Feedback) and improve them until a satisfactory result is achieved.
6. **Learning from mistakes and fine-tuning is often used. The systems use** mechanisms such as reinforcement learning from human feedback (RLHF) and fine-tuning on specific tasks are employed, allowing the AI to solve increasingly difficult or specialized problems.

The way AI systems search for solutions especially highlights the ongoing efforts by computer scientists to make the representation of objects and entire situations more similar to analogous phenomena occurring in the human brain. These improved methods of implementing representations are often referred to as the neuro-symbolic approach [59].

However, it appears that, at present, there are still no efficient AI systems based on neuro-symbolic representation that would closely resemble human-like mental imagery. Some authors believe that these newly proposed alternative systems for representing objects and situations are, in essence, very similar to the vector-based representation in latent space [60]

18. Self-Awareness and the Potential Consciousness of AI Systems

The issue indicated in the title of this chapter has been so extensively discussed in both popular and scientific articles that we will not devote much space to it here, instead referring readers to numerous works by other authors. In particular, important are the studies that consider the two main competing theories known as the "Integrated Information Theory (IIT)" and the "Global Neuronal Workspace Theory (GNWT)." [61, 62].

We have addressed the phenomenon of human consciousness in depth in one of our recent papers [11]. This paper was well received and read by several thousand people. In it, we present the reasons why a theory that successfully explains the essence of consciousness should integrate three distinct contemporary theories. Such integration of all three is necessary in order to propose an explanation for the so-called Chalmers' "hard problem" and to answer the question of "who perceives" the images generated during mental imagery. Here, we will focus on two specific aspects of this topic that we consider particularly important. First, authors frequently ponder how to determine whether a particular entity (being or system) is self-aware (63). Since classic tests such as the mirror test are difficult to apply to AI systems, we propose a simpler and more natural approach. Namely, we suggest that one should start by recalling the dictionary definition of the concept and then consider whether the features listed in such a definition are present. If we refer to such a description, we generally find the following explanation:

"Self-awareness" is the ability for introspection – that is, the capacity to recognize and understand one's own thoughts, emotions, motivations, and behaviors. It encompasses:

1. **Recognition of one's own internal states:** The ability to notice what is happening in one's mind and body, forming the basis for consciously regulating emotions and reactions.
2. **Awareness of identity:** The sense of oneself as a distinct entity, which enables evaluation of one's actions and the impact they have on the environment.
3. **A basis for personal development:** Self-awareness is a key element of emotional intelligence, allowing for reflection on one's own behavior, identification of strengths, and areas for improvement.

In the psychological literature, "self-awareness" is analyzed both in the context of everyday self-reflection and in studies of brain function and mechanisms of conscious thought. Scientific works in this field often indicate that a high level of self-awareness promotes better emotional management and more effective decision-making.

Thus, the consideration of whether a particular AI system is self-aware comes down to determining the presence of the features listed above. Such an evaluation is, in fact, an example of the comparative procedure for analyzing similarities and differences between the processes under consideration, as proposed in this work. Within the realm of possible or problematic self-awareness in AI systems, we believe it is also crucial to clarify one further point. It is always necessary to consider to what extent the AI system in question operates only locally, or whether it receives information and has access to a wide environment. It is possible, after all, to imagine non-local AI systems [64]. If we assume that the reasoning component of a large, futuristic AI system has access not only to images, photos, and films, but also to cameras installed on the planet's surface and astronomical telescopes, then we could say that such an AI system is perceiving its environment. Such a system would, by its very nature, also have insight into the essence of all existing software. On that basis, such a system would be able to construct an "image of itself"—the equivalent of a human's self-concept, that is, the equivalent of the concept of "I".

19. Conclusion

1. A description of how the human brain performs perception, imagery, problem-

solving, and the acquisition of self-awareness – in such a way that the presented theoretical model could be considered by computer scientists constructing artificial intelligence systems, as well as their users – ***must be based on selectively chosen descriptions of findings*** made by neuroscientists that are not overly detailed and allow for intuitive understanding.

2. ***The key sequence of cognitive phenomena that must be explained*** to enable an intuitive understanding of how the brain functions includes the following theoretical models: Models of neurons as "integrate and fire" elements; Hierarchical structures and neuronal circuits responsible for perception; Mechanisms of memory consolidation and the distinction between two types of connections of object neurons with the thalamus and the hippocampus; The way mental imagery is generated; The role of the language areas in the human brain; Cognitive mechanisms involved in problem-solving; The significance of theories concerning the brain's endogenous electromagnetic field, as well as m icrotubule quantum processing – specifically Hameroff and Penrose's "Orch OR" theory – for explaining the emergence of self-awareness.
3. It seems that the following cognitive processes, carried out both by the human brain and by A.I. systems, ***should be*** thoroughly understood and ***compared*** with each other as a priority, so that their essence is clear to medical professionals, natural scientists, and computer scientists alike: the function of the brain's speech areas, problem-solving mechanisms in the brain and in A.I. systems, the formation and manipulation of mental images (imagination), and image generation by A.I. systems based on linguistic descriptions.

REFERENCES

[1] Munn BR, Müller EJ, Favre-Bulle I, Scott E, Lizier JT, Breakspear M, Shine JM. Multiscale organization of neuronal activity unifies scale-dependent theories of brain function. Cell. 2024; 187(25):7303-7313.e15, https://doi.org/10.1016/j.cell.2024.10.004.

[2] Goel V, Navarrete G, Noveck IA, Prado J. Editorial: The reasoning brain: The interplay between cognitive neuroscience and theories of reasoning. Front Hum Neurosci. 2017;10:673, https://doi.org/10.3389/fnhum.2016.00673

[3] Heit E. Brain imaging, forward inference, and theories of reasoning. Front Hum Neurosci. 2015; 8: 1056. https://doi.org/10.3389/fnhum.2014.01056

[4] Guo Y, Li Y, Wei HL, Zhao Y. Editorial: New theories, models, and AI methods of brain dynamics, brain decoding and neuromodulation. Front Neurosci. 2023; 17: 1302505, https://doi.org/10.3389/fnins.2023.1302505

[5] Whittington JCR, Bogacz R. Theories of error back-propagation in the brain. Trends Cogn Sci. 2019;23(3):235-250, https://doi.org/10.1016/j.tics.2018.12.005

[6] Brodziak A. [Psychonics. Theory of structures and information processes of human central nervous system and its application in computer sciences] – in Polish - Published by Medical Section of Polish Academy of Sciences, Katowice, 1974 https://tezeusz.pl/psychonika-brodziak-andrzej-5702900?srsltid=AfmBOoof5h397I1f_kDpgQ6HPECIzZdxpBoI3lZey-HEmoKNMqiUKmPl https://katalog.nukat.edu.pl/cgi-bin/koha/opac-detail.pl?biblionumber=3222665

[7] Brodziak A. Neurophysiology of the mental image. Med Sci Monit. 2001;7(3):534-8 https://www.medscimonit.com/abstract/index/idArt/510460

[8] Brodziak A, Brewczyński A, Bajor G. Clinical significance of knowledge about the structure, function, and impairments of working memory. Med Sci Monit. 2013; 19:327-38, https://doi.org/10.12659/MSM.883900.

[9] Brodziak A. A current model of neural circuitry active in forming mental images. Med Sci Monit 2013; 19:1146-1158. https://doi.org/10.12659/MSM.889587

[10] Brodziak A, Różyk-Myrta A. The simple, intuitive model of neural circuits, which memorize data, recognize an image and recall it as imagination. ARC Journal of Neuroscience. 2018;3(3):1-5 https://www.arcjournals.org/pdfs/ajns/v3-i3/1.pdf

[11] Różyk-Myrta A, Brodziak A, Muc-Wierzgoń M. Neural circuits, microtubule processing, brain's electromagnetic field-components of self-awareness. Brain Sci. 2021; 11(8):984. https://doi.org/10.3390/brainsci11080984

[12] Brodziak A, Romaniuk F. The functional unit of neural circuits and its relations to eventual sentience of artificial intelligence systems. Qeios, 2023, http://doi.org/10.32388/82VRPG

[13] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943; 5(4):115-133. https://doi.org/10.1007/bf02478259

[14] Hubel, D.H.; Wiesel, T.N. Receptive fields of single neuron in the cat's striate cortex. J. Physiol. 1959; 148(3): 574–591 http://doi.org/10.1113/jphysiol.1959.sp006308

[15] Gross, C.G.; Rocha-Miranda, C.E.; Bender, D.B. Visual properties of neuron in inferotemporal cortex of the macaque. J. Neurophysiol. 1972; 35: 96–111. https://doi.org/10.1152/jn.1972.35.1.96

[16] Mishkin, M. A memory system in the monkey. Philos. Trans. R. Soc. Lond. B Biol. Sci. 1982; 298 (1089): 83–95. https://doi.org/10.1098/rstb.1982.0074

[17] Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. Nature, 2005; 435(7045):1102–1107 https://doi.org/10.1038/nature03687

[18] Quiroga RQ, Fried I, Koch C. Brain cells for grandmother. Sci. Am. 2013; 308(2): 30–35. https://doi.org/10.1038/scientificamerican0213-30

[19] Pearson J, Naselaris T, Holmes EA, Kosslyn SM. Mental imagery: Functional mechanisms and clinical applications. Trends Cogn. Sci. 2015; 19(10): 590–602. https://doi.org/10.1016/j.tics.2015.08.003

[20] Dijkstra N, Bosch SE, van Gerven MAJ. Shared Neural Mechanisms of Visual Perception and Imagery. Trends Cogn. Sci. 2019; 23(5): 423–434. https://doi.org/10.1016/j.tics.2019.02.004

[21] Keogh R, Bergmann J, Pearson J. Cortical excitability controls the strength of mental imagery. Elife. 2020; 9: e50232. https://doi.org/10.7554/eLife.50232

[22] Lou HC, Joensson M, Biermann-Ruben K, Schnitzler A, Østergaard , Kjaer TW, Gross J. Recurrent activity in higher order, modality non-specific brain regions: A Granger causality analysis of autobiographic memory retrieval. PLoS ONE 2011, 6, e22286. https://doi.org/10.1371/journal.pone.0022286

[23] Gilbert CD, Sigman M. Brain states: Top-down influences in sensory processing. Neuron. 2007;54(5):677-96 https://doi.org/10.1016/j.neuron.2007.05.019

[24] Bressler SL, Richter CG. Interareal oscillatory synchronization in top-down neocortical processing. Curr. Opin. Neurobiol. 2015;31:62-6 https://doi.org/10.1016/j.conb.2014.08.010

[25] Sitaram R, Braun C. Involvement of top-down networks in the perception of facial emotions: A magnetoencephalographic investigation. Neuroimage 2020; 222: 117075. https://doi.org/10.1016/j.neuroimage.2020.117075

[26] Svoboda E, McKinnon MC, Levine B. The functional neuroanatomy of autobiographical memory: a meta-analysis. Neuropsychologia. 2006; 44(12):2189-208.

https://doi.org/10.1016/j.neuropsychologia.2006.05.023.

[27] Sulpizio, V., Teghil, A., Ruffo, I. et al. Unveiling the neural network involved in mentally projecting the self through episodic autobiographical memories. Sci Rep 2025; 15: 12781 https://doi.org/10.1038/s41598-025-97515-0

[28] Fujii M, Maesawa S, Ishiai S, Iwami K, Futamura M, Saito K. Neural basis of language: an overview of an evolving model. Neurol Med Chir (Tokyo). 2016; 56(7):379-86. https://doi.org/10.2176/nmc.ra.2016-0014

[29] Friederici AD. The brain basis of language processing: From structure to function. Physiol. Rev., 2011; 91(4): 1357-1392 https://doi.org/10.1152/physrev.00006.2011

[30] McFadden, J. Synchronous firing and its influence on the brain's electromagnetic field: Evidence for an electromagnetic theory of consciousness. J. Conscious. Stud. 2002, b; 9, 23–50. https://www.ingentaconnect.com/content/imp/jcs/2002/00000009/00000004/1267

[31] McFadden, J. Integrating information in the brain's EM field: The cemi field theory of consciousness. Neurosci. Conscious. 2020, 2020(1):niaa016. https://doi.org/10.1093/nc/niaa016

[32] Hales, C.G.The origins of the brain's endogenous electromagnetic field and its relationship to provision of consciousness. J. Integr. Neurosci. 2014; 13 (2):313-61. https://doi.org/10.1142/S0219635214400056

[33] Liboff AR. Magnetic correlates in electromagnetic consciousness. Electromagn. Biol. Med. 2016; 35(3): 228–236. https://doi.org/10.3109/15368378.2015.1057641

[34] Hales, C.G.; Pockett, S. The relationship between local field potentials (LFPs) and the electromagnetic fields that give rise to them. Front. Syst. Neurosci. 2014; 8: 233. https://doi.org/10.3389/fnsys.2014.00233

[35] Hunt T, Jones M, McFadden J, Delorme A, Hales CG, Ericson M, Schooler J. Editorial: Electromagnetic field theories of consciousness: opportunities and obstacles. Front. Hum. Neurosci. 2024; 17. https://doi.org/10.3389/fnhum.2023.1342634

[36] Hameroff S, Penrose R. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. Math. Comput. Simul. 1996, 40, 453–480.

https://www.sciencedirect.com/science/article/pii/0378475496804769?via%3Dihub

[37] Penrose R. The Large, the Small and the Human Mind; Cambridge University Press: Cambridge, UK, 1997.

https://books.google.pl/books?hl=pl&lr=&id=jWHqlijAjyMC&oi=fnd&pg=PR7&ots=bfUHg7L7M9&sig=hfophcAM7unNx4IIq8vQvGAGZzs&redir_esc=y#v=onepage&q&f=false

[38] Penrose R. Consciousness, the brain, and spacetime geometry: An addendum. Some new developments on the Orch OR model for

consciousness. Ann. N. Y. Acad. Sci. 2001, 929, 105–110. https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1111/j.1749-6632.2001.tb05710.x

[39] Hameroff S, Nip A, Porter M, Tuszynski J. Conduction pathways in microtubules, biological quantum computation, and consciousness. Biosystems 2002; 64: 149–168. https://www.sciencedirect.com/science/article/pii/S0303264701001836?via%3Dihub

[40] Hameroff S. Consciousness, the brain and spacetime geometry. Ann. N. Y. Acad. Sci. 2001; 929:74-104. https://doi.org/10.1111/j.1749-6632.2001.tb05709.x.

[41] Hameroff S. Quantum mathematical cognition requires quantum brain biology: The “Orch OR” theory. Behav. Brain Sci. 2013; 36: 287–290. https://doi.org/10.1017/S0140525X1200297X

[42] Hameroff SR, Craddock TJ, Tuszynski JA. Quantum effects in the understanding of consciousness. J. Integr. Neurosci. 2014; 13(2): 229–252. https://doi.org/10.1142/S0219635214400093

[43] Hameroff S. Quantum walks in brain microtubules—A biomolecular basis for quantum cognition? Top. Cogn. Sci. 2014; 6: 91–97. https://onlinelibrary.wiley.com/doi/10.1111/tops.12068

[44] Hameroff S, Penrose R. Consciousness in the universe: A review of the ‘Orch OR’ theory. Phys. Life Rev. 2014; 11 (1):39-78. https://doi.org/10.1016/j.plrev.2013.08.002

[45] Wiest MC. A quantum microtubule substrate of consciousness is experimentally supported and solves the binding and epiphenomenalism problems. Neurosci Conscious. 2025; 2025 (1):niaf011. https://doi.org/10.1093/nc/niaf011

[46] Daselaar SM, Rice HJ, Greenberg DL, Cabeza R, LaBar KS, Rubin DC. The spatiotemporal dynamics of autobiographical memory: Neural correlates of recall, emotional intensity, and reliving. Cereb. Cortex. 2008; 18(1): 217–229. https://doi.org/10.1093/cercor/bhm048

[47] Llinas R, Ribary U, Contreras D, Pedroarena C. The neuronal basis for consciousness. Philos. Trans. R. Soc. Lond. B Biol. Sci. 1998, 353(1377), 1841–1849. https://doi.org/10.1098/rstb.1998.0336

[48] Northoff G. What the brain’s intrinsic activity can tell us about consciousness? A tri-dimensional view. Neurosci. Biobehav. Rev. 2013; 37(4): 726–738. https://doi.org/10.1016/j.neubiorev.2012.12.004 https://www.sciencedirect.com/science/article/pii/S0149763412002138

[49] Szilágyi A, Zachar I, Fedor A, de Vladar HP, Szathmáry E. Breeding novel solutions in the brain: a model of Darwinian neurodynamics. F1000Res. 2016; 5: 2416. https://doi.org/10.12688/f1000research.9630.2

[50] Khona M, Fiete IR. Attractor and integrator networks in the brain. Nat Rev Neurosci. 2022; 23(12):744-766. https://doi.org/10.1038/s41583-022-00642-0

[51] Fedor A, Zachar I, Szilágyi A, Öllinger M, de Vladar HP, Szathmáry E. Cognitive architecture with evolutionary dynamics solves insight problem. Front Psychol. 2017; 8:427. https://doi.org/10.3389/fpsyg.2017.00427

[52] Chitwan Saharia, William Chan, Saurabh Saxena, et al. Photorealistic text-to-image diffusion models with deep language understanding. arXiv:2205.11487v1 , 2022 https://arxiv.org/pdf/2205.11487

[53] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, Junmo Kim. Text-to-image diffusion models in generative AI: A survey. arXiv:2303.07909, 2023 https://arxiv.org/pdf/2303.07909

[54] Mingyang Yi, Aoxue Li, Yi Xin, Zhenguo Li. Towards understanding the working mechanism of text-to-image diffusion model. arXiv: 2405.15330 , 2024 https://openreview.net/pdf?id=zTu0QEpvtZ

[55] Joynt, V., Cooper, J., Bhargava, N. et al. A comparative analysis of text-to-image generative AI models in scientific contexts: a case study on nuclear power. Sci Rep 2024;14: 30377 https://doi.org/10.1038/s41598-024-79705-4

[56] Gizzi E, Nair L, Shernova S, Sinapov J. Creative Problem Solving in Artificially Intelligent Agents: A Survey and Framework. arXiv:2204.10358v1, 2022, https://arxiv.org/pdf/2204.10358

[57] Kebin Jin, Hankz Hankui Zhuo. Integrating AI planning with natural language processing: A combination of explicit and tacit knowledge. arXiv:2204.10358v1, 2022, https://arxiv.org/pdf/2202.07138

[58] Ajith Vallath Prabhakar. Latent Reasoning: The next evolution in AI for scalable, adaptive, and efficient problem-solving. Ajith's AI Pulse, 2025, https://ajithp.com/2025/02/14/latent-reasoning-the-next-evolution-in-ai-for-scalable-adaptive-and-efficient-problem-solving/

[59] Nawaz U, Anees-ur-Rahaman M, Saeed Z. A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems. Intelligent Systems with Applications. 2025; 26: 2667-3053. https://doi.org/10.1016/j.iswa.2025.200541

[60] Drexler E. LLMs and Beyond: All Roads Lead to Latent Space. AI Prospects: Toward Global Goal Alignment. 2025 https://aiprospects.substack.com/p/llms-and-beyond-all-roads-lead-to

[61] Zeman A, Coebergh JA. The nature of consciousness. Handb Clin Neurol. 2013; 118:373-407. https://doi.org/10.1016/B978-0-444-53501-6.00031-7

[62] Parshall A. Where does consciousness come from? Two neuroscience theories go head-to-head. Scientific American, 2025 https://www.scientificamerican.com/article/where-does-consciousness-come-from-two-neuroscience-theories-go-head-to-head/

[63] Biyu J. He; Integrating consciousness science with cognitive neuroscience: An introduction to the special focus. J Cogn Neurosci 2024; 36 (8): 1541–1545. https://doi.org/10.1162/jocn_a_02193

[64] Brodziak A, Abram D, Różyk-Myrta A. Planetary consciousness incites probably transcendent feelings and deepens the polarization of worldviews. Qeios, 2023, https://doi.org/10.32388/BU24PQ

Citation: *Andrzej Brodziak, et al. "How the Brain Works – Insights for AI Systems Analysts". ARC Journal of Neuroscience. 2025; 8(1): 23-38. DOI:* https://doi.org/*10.20431/2456-057X.0801004*