https://doi.org/10.52326/jes.utm.2026.33(1).03
UDC 004.8:004.94:004.31:57.087

ADAPTIVE COMPUTING STRUCTURES FOR SERVICE-ORIENTED MULTI-AGENT SYSTEMS BASED ON KNOWLEDGE MODELS

Vadim Struna[1*], ORCID: 0000-0001-6579-3054,
Victor Ababii[1], ORCID: 0000-0002-0769-8144,
Viorica Sudacevschi[1], ORCID: 0000-0003-0125-3491,
Silvia Munteanu[1], ORCID: 0000-0003-0749-8457,
Olesea Borozan[1], ORCID: 0000-0003-1091-5506,
Victoria Alexei[1], ORCID: 0000-0003-4560-3131

[1] *Technical University of Moldova, Ştefan cel Mare, 168, Chişinău, MD-2004, Republic of Moldova*
* Corresponding author: Vadim Struna, *vadim.struna@iis.utm.md*

Received: 01. 24. 2026
Accepted: 02. 27. 2026

Abstract. This paper proposes an adaptive computing framework for service-oriented multi-agent systems, based on knowledge models inspired by the hierarchical organization of the human brain. The approach integrates neurophysiological principles of conscious and subconscious processing with rigorous mathematical formalization and hardware-oriented architectural design. The conscious−subconscious interaction is modeled as a two-level computational hierarchy, in which subconscious processing is fast, parallel, adaptive, and high-dimensional, and conscious processing is deliberative, symbolic, and low-dimensional, being responsible for control, planning, and decision-making. An attention-based coupling mechanism controls the flow of information between the two levels, allowing for dynamic adaptation and efficient use of resources. Based on this model, a heterogeneous hardware architecture is proposed that maps subconscious processing to NPU/GPU accelerators, and conscious processing to CPU units. The framework is extended to multi-agent systems, in which each agent implements a conscious−subconscious hierarchy, and the emergent coordination is achieved through a collective conscious level. The approach supports distributed intelligence, scalability, and adaptive service composition.

Keywords: *conscious−subconscious processing, cognitive architectures, distributed intelligence, heterogeneous hardware, hierarchical computing, knowledge models, mathematical modeling, physiology of the human brain.*

1. Introduction

The accelerated evolution of distributed computing systems, driven by the growing complexity of digital environments and rising demands for flexibility, scalability and autonomy, has led to the emergence of new design paradigms based on adaptivity and distributed intelligence. In this context, service-oriented multi-agent systems represent a promising solution for modeling and managing complex, dynamic and heterogeneous processes, characteristic of modern applications in areas such as intelligent infrastructures, cyber-physical systems, the Internet of Things (IoT) and critical situation management [1].

A major challenge in the development of service-oriented multi-agent systems lies in their ability to adapt in real time to rapid and unpredictable changes in the environment, the continuous evolution of functional requirements and the dynamic integration of new resources and services, including AI- and machine-learning-based services. Traditional computational structures, rigid and predefined, can no longer sustain the complexity and dynamism necessary for the emergence of self-organizing intelligent behaviors at the system level. In this context, adaptive computing structures and architectures based on continuous learning become essential, facilitating the dynamic reorganization of components, optimal redistribution of tasks, autonomous adjustment of agent behavior and efficient integration of emerging technologies, depending on context, objectives and performance criteria [2, 8, 9].

A central role in achieving adaptivity is played by knowledge models, which provide formal mechanisms for representing, acquiring, inferring and updating knowledge, both at the agent level and as well at the system level. The integration of knowledge models into service-oriented multi-agent system architectures enables agents to make informed decisions, anticipate environmental developments, and effectively to coordinate the services offered. Thus, knowledge becomes not only an information resource, but also an active factor of control, structural adaptation and continuous learning within intelligent systems [3–7, 10–12].

The present paper aims to develop service-oriented adaptive multi-agent systems, based on hierarchical knowledge models inspired by the conscious–subconscious interaction. It is proposed to develop mathematical models for formalizing the functioning mechanisms inspired by the human brain and for describing the behavior of multi-agent systems, as well as the development of heterogeneous hardware architectures, which combine neural accelerators (NPU/GPU) for subconscious processing with CPU units for deliberative control, supporting adaptive learning and decision-making algorithms. The expansion to multi-agent systems will integrate collective coordination mechanisms and emerging behavior, ensuring scalability, fault tolerance and interoperability in dynamic environments. This approach will enable the development of autonomous cognitive platforms capable of managing complex services and supporting applications in areas such as digital twins, autonomous robotics and smart environments.

2. Physiology of the human brain. The conscious and subconscious

The study of consciousness and subconscious processes is one of the most complex and challenging directions of modern neuroscience. Consciousness is investigated both as a subjective phenomenon and as a result of the coordinated activity of neural networks, while the subconscious is analyzed as a fundamental mechanism of rapid and efficient adaptation to the environment. Modern perspectives in cognitive neuroscience emphasize the interdisciplinary nature of this issue, integrating data from neurophysiology, cognitive psychology and computational modeling. Seth and Bayne [22] provide a synthesis of the theories of consciousness and justify the relevance of the research topic. Additionally, Gazzaniga et al. [18] highlight the perspective of cognitive neuroscience on consciousness and the integration of neuroanatomical structure with cognitive functions.

The human brain is a highly complex biological system, made up of about 86 billion neurons interconnected by dynamic synaptic networks. Its functioning is based on the propagation of action potentials, on chemical and electrical synaptic transmission, as well as on synaptic plasticity mechanisms, considered essential for learning and memory processes [19]. From a functional perspective, cortical structures are involved in higher cognitive processes, while subcortical structures and the brainstem support automatic and vital processes. This organization allows the simultaneous emergence of conscious and subconscious processes in a unitary system [18].

Consciousness (≈10% of human brain activity) represents the upper level of the rational decision-making center, being relatively slow and operating based on predefined patterns. It is responsible for analytical thinking, intentional planning, and processing information that we are immediately aware of. The conscious is associated with mental states accessible to subjective reporting, such as perception, attention and deliberative reasoning. Neuroimaging studies indicate that these processes are correlated with the activation of fronto-parietal networks and with the mechanisms of global information integration. The Global Neuronal Workspace theory holds that information becomes conscious when it is amplified and widely distributed in the cortex [16, 20]. The dynamics of neuronal synchronization and cortical "binding" processes, researched by Baars in [1, 13], are essential for maintaining conscious content.

The subconscious (≈90% of the human brain's activity) represents a vast array of automated, autonomous processes based on long-term experience, habits, and memories. These processes are extremely fast, capable of processing up to 20 million bits of information per second. The subconscious comprises cognitive processes that take place without direct access to consciousness, but that significantly influence behavior and decision-making. These processes are fast, parallel, and energy-efficient, being associated with structures such as the amygdala, hippocampus, and basal ganglia. The role of subconscious emotions in regulating behavior and in decision-making is evidenced by Damasio's studies [15]. Network models of the emotional brain, researched by Pessoa in [21], show that affective and cognitive processes are deeply interconnected, even in the absence of explicit awareness. Also, social and adaptive behaviors are often mediated by subconscious mechanisms demonstrated by Stanley and Adolphs in [23].

The relationship between the conscious and the subconscious is a two-way and dynamic one. Predictive coding models describe the brain as an inference system that generates subconscious predictions and uses conscious processes to correct significant prediction errors [14]. The principle of free energy, researched by Friston in [17], provides a unifying mathematical framework, according to which both conscious and subconscious processes aim to minimize uncertainty and internal errors. Emotions and subconscious somatic markers decisively influence conscious rational processes, confirming the interdependence of the two levels [15]. The conscious monitors and interprets cognitive activity, and delegates routine tasks to the subconscious level, while the subconscious provides decisions based on experience (intuition) and automatically manages processes, freeing up conscious resources.

Modern approaches treat consciousness as an emergent property of information integration into complex neural networks. Integrated Information Theory (IIT), researched by Tononi et al. in [24], proposes a formal measure of the level of consciousness, based on the degree of informational integration. Recent comparative analyses [22], highlight the complementarity between IIT, *the Global Neuronal Workspace* and predictive models. Also, mathematical formulations inspired by Bayesian inference provide a rigorous framework for describing the conscious–subconscious interaction [17].

Understanding the relationship between the conscious and subconscious has major implications in areas such as clinical neuropsychology, cognitive artificial intelligence, and adaptive systems. Models inspired by *Global Workspace* and *predictive coding* have influenced the development of hybrid cognitive architectures in artificial intelligence [13, 14]. Also, the concepts of information integration are theoretically explored in the context of complex artificial systems [24].

Therefore, the conscious and the subconscious are not separate entities, but complementary functional levels of brain activity. The subconscious provides efficiency and quick adaptation, while the conscious provides deliberative control and cognitive flexibility. Neurophysiological data and modern theoretical models support the idea that the interaction between these levels is essential for the coherent functioning of the human brain [19, 22].

3. Mathematical Modeling of Conscious–Subconscious Interaction

The functioning of the human brain is based on a clear functional separation between subconscious, fast, automatic and distributed processes and conscious processes, characterized by global integration, deliberation and decision-making control. This duality constitutes a relevant biological model for the design of adaptive computational structures and service-oriented multi-agent systems.

From a mathematical perspective, the conscious–subconscious interaction can be described by a nonlinear dynamical system, in which the global state is decomposable into two functional subspaces: one associated with local and implicit processing (subconscious), and one associated with global and explicit processing (conscious). Subconscious processes operate on the basis of dynamic attractors and implicit knowledge, ensuring rapid and robust reactions, while conscious processes achieve informational integration, contextual evaluation and decision selection [25].

Consciousness emerges when the level of information integration and synchronization exceeds a critical threshold, allowing global access to relevant representations. The subconscious continuously provides content and preliminary assessments, which are amplified, filtered and stabilized at the conscious level. This bidirectional interaction enables adaptive optimization of system behavior in dynamic environments.

The conscious–subconscious model thus provides a formal framework for defining the mechanisms of hierarchical control and decision-making emergence in adaptive systems, constituting a conceptual bridge between cognitive physiology and advanced computational architectures based on knowledge models.

The neurophysiological state space represents a multidimensional space of neuronal and physiological variables, in which each point corresponds to a distinct functional state of the brain. We will consider the human brain as a complex, distributed and nonlinear dynamic system, described by a state space (1):

$$B = \langle \mathrm{N}, S, D, F \rangle, \tag{1}$$

where:

$\mathrm{N} = \{ n_i, i = \overline{1,N} \}$ - is the total number of neurons;

$S \subset \mathbb{R}^N$ - it is the space of neural states;

D - represents the temporal dynamics of neurons;

F - are the functions of information processing.

The overall state of the brain at time t is defined as (2):

$$X(t)=\left[x_1(t),x_2(t),...,x_N(t)\right]\subset S, \tag{2}$$

where $x_i(t)$ represents the activation potential of the neuron i.

Fundamental neuronal dynamics represent the totality of biophysical and mathematical mechanisms that determine the temporal evolution of neuronal activity, constituting the basis of all cognitive and behavioral processes. Therefore, the general dynamics of the neural system can be expressed by a system of nonlinear differential equations (3):

$$\frac{dX(t)}{dt}=F\left(X(t),u(t),\eta(t)\right), \tag{3}$$

where:

$u(t)$ - are the set of external stimuli (sensory inputs for perceiving the environment);

$\eta(t)$ - is neurophysiological noise;

F - is the nonlinear operator of synaptic interaction.

Conscious–subconscious functional decomposition represents the conceptual separation of cognitive processes into explicit and implicit components, used for the analysis and modeling of cognitive and neurophysiological dynamics. The state space is decomposed into two functional subspaces (4):

$$S=S_C\oplus S_{SC}, \tag{4}$$

where:

S_C - is the subspace of the conscious;

S_{SC} - is the subspace of the subconscious.

The global condition is expressed as (5):

$$X(t)=X_C(t)+X_{SC}(t). \tag{5}$$

Conscious modeling is the process of formalizing explicit cognitive mechanisms and their dynamics, in order to analyze, simulate and implement them in natural or artificial cognitive systems. The conscious is associated with integrative, global processes that are accessible to cognitive reporting. From a mathematical point of view, it can be expressed by the expression (6):

$$X_C(t)=P_C\left(X(t)\right), \tag{6}$$

where P_C is a space projection operator $X(t)$ on the conscious subspace $X_C(t)$.

The dynamics of conscious processes are defined by the system of differential equations (7):

$$\frac{dX_C(t)}{dt}=F_C\left(X_C(t),X_{SC}(t)\right). \tag{7}$$

The conscious manifests itself as a cognitive regime in which information is integrated at a global level, allowing the coherent coordination of perceptual, memory and decision-making processes.

Subconscious modeling is the formalization of implicit cognitive processes and their dynamics, in order to analyze their interaction with the conscious level and their role in adaptive behavior. The subconscious, associated with automated, fast, local processes that are poorly accessible to introspection, is defined by the expression (8):

$$\mathrm{X}_{SC}(t)=\mathrm{P}_{SC}\left(\mathrm{X}(t)\right), \tag{8}$$

where P_{SC} is a space projection operator $\mathrm{X}(t)$ on the subconscious subspace $\mathrm{X}_{SC}(t)$

The dynamics of subconscious processes are defined by the system of differential equations (9):

$$\frac{d\mathrm{X}_{SC}(t)}{dt}=F_{SC}\left(\mathrm{X}_{SC}(t),u(t)\right). \tag{9}$$

Subconscious processes are governed by attractor-type dynamics, corresponding to neuronal and behavioral patterns stabilized by learning, which determine automatic and adaptive responses.

The mechanism of emergence of consciousness is the process by which the dynamic interactions of neural networks lead to the emergence of a global cognitive regime, characterized by informational integration and access to the conscious level. Consciousness appears as an emergent phenomenon when the degree of synchronization and integration exceeds a critical threshold defined by the expression (10):

$$C(t)=\begin{cases}1, & if\ \Phi(t)\geq\Phi_{crit}, \\ 0, & after.\end{cases} \tag{10}$$

where:

$\Phi(t)$ - is the measure of informational integration;

Φ_{crit} - is the threshold of awareness.

The conscious–subconscious interaction is a two-way process through which the conscious and subconscious levels cooperate and influence each other in the functioning of the cognitive system. This interaction constitutes the functional coupling between explicit and implicit cognitive processes, through which information is generated, selected and used in adaptive behavior. The bidirectional coupling is modeled and formalized by the system of differential equations (11):

$$\begin{cases}\dfrac{d\mathrm{X}_C}{dt}=F_C\left(\mathrm{X}_C,\mathrm{X}_{SC}\right), \\ \dfrac{d\mathrm{X}_{SC}}{dt}=F_{SC}\left(\mathrm{X}_{SC},\mathrm{X}_C\right).\end{cases} \tag{11}$$

The contribution of the subconscious to the conscious decision is calculated on the basis of the expression (12):

$$D(t)=\alpha\mathrm{X}_C(t)+(1-\alpha)\mathrm{X}_{SC}(t),\quad \forall\alpha\in\left[0,1\right]. \tag{12}$$

Physiological and cognitive interpretation highlights the fact that the functioning of the brain is based on a hierarchical and complementary architecture between the subconscious and the conscious. The subconscious provides parallel, fast and metabolically efficient processing, automatically managing most sensory and behavioral information, while the conscious achieves the global, slow and energy-intensive integration necessary for voluntary control, decision-making and flexible adaptation. The interaction between these two levels allows the brain to function as a self-organizing adaptive system, capable of continuously optimizing the relationship between energy efficiency and cognitive complexity in relation to the environment.

4. Hardware Architecture of the Computing System Based on Hierarchical Conscious–Subconscious processing

The synthesis of the hardware architecture for artificial intelligence systems with hierarchical processing is based on the mathematical formalization of the conscious–subconscious interaction. The architecture is designed as a heterogeneous system, in which the subconscious level performs local, parallel and fast processing, using neuromorphic processing units, TPU/NPU accelerators and associative memory for distributed inference and reactive control. The conscious layer operates on CPUs and many-core, integrating global data, knowledge bases, and planning and decision-making mechanisms. The interaction between levels is governed by an informational threshold module, which evaluates the integration measure Φ and enables global processing, according to the expression (10) $\Phi \geq \Phi_{crit}$. This structure allows the emergence of global decision-making, the maintenance of information coherence and the adaptive optimization of behavior in service-oriented multi-agent systems. Hardware synthesis thus integrates the principles of cognitive hierarchy and parallel processing with the requirements of performance, scalability and energy efficiency, providing an implementable framework for cognitive and autonomous AI systems. The hardware architecture of the computing system based on conscious–subconscious hierarchical processing is shown in Figure 1 (*generated with the support of AI ChatGPT-5.2*).

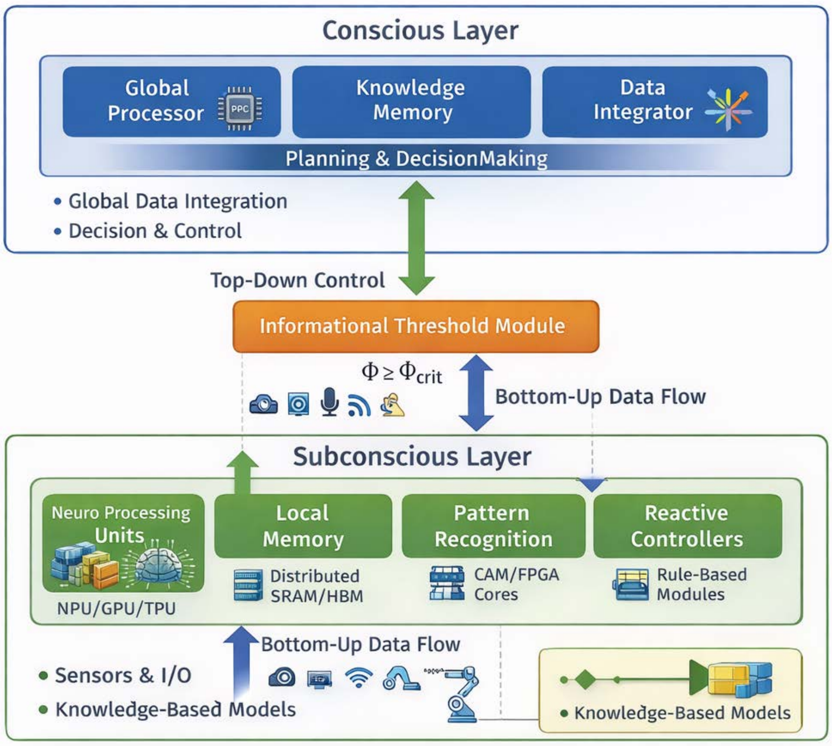

Figure 1. Hardware architecture of the computing system.

Functional specification of hardware architecture:

Conscious Layer – the upper level of architecture; responsible for the global integration of information, planning, decision-making and top-down control over the subconscious level;

Global Processor (CPU / Many-Core / PPC architectures) – responsible for global decision processing and data integration; executes symbolic reasoning and strategic planning algorithms, receives bottom-up input from the subconscious module and generates top-down commands to the subconscious module;

Knowledge Memory – responsible for storing explicit information, rules, behavior patterns, and historical data; provides quick access to data for contextual assessments and integrated inferences, and serves as a reference for top-down control and consolidation of emerging decisions;

Data Integrator – ensures the aggregation of multiple data streams from the subconscious level and from external sensors; achieves the reconciliation and correlation of information, reduces redundancy and inconsistencies and prepares the input for decision-making at the conscious level;

Top-Down Control – flow of commands and constraints transmitted to the subconscious level; activates the conscious for global decision and planning and receives parameters and objectives from the conscious layer, adjusting local processing and task prioritization;

Bottom-Up Data Flow – flow of information preliminarily processed in the subconscious and sent for integration and global evaluation at the conscious level; collects data and patterns from the Subconscious Layer and transmits them to the threshold and conscious module for global decision;

Informational Threshold Module – applies the computational model in expression (10) to assess entropy and correlation between data streams; filters and aggregates inputs from the subconscious level and determines when information is propagated to the conscious level;

Subconscious Layer – the lower level of the architecture, responsible for fast processing, local inference, pattern recognition and reactive control; works as a parallel and distributed system with very low latency;

Neuro Processing Units (NPU/GPU/TPU architectures) – process data in parallel for pattern recognition and automatic inference, handle continuous inputs from sensors and I/O, and execute fast prediction and associative inference algorithms;

Local Memory – includes distributed SRAM/HBM and content-addressable memories (CAM) for quick access to information; stores default knowledge and recurring patterns and enables local reactions without immediately involving the conscious level;

Pattern Recognition – detects patterns, anomalies and complex relationships in input data; uses dedicated FPGAs and accelerators for parallel inference and prepares data for bottom-up propagation to the informational threshold module;

Reactive Controllers – include rule-based modules and predefined logic for immediate response to stimuli and ensure stability and rapid reaction in distributed systems;

Sensors & I/O (External Interface) – functions for collecting data from the physical environment or other systems; provide bottom-up input to the Subconscious Layer and collect feedback for adaptation, ensuring connectivity with external systems;

Knowledge-Based Models – include modules integrated at both the subconscious and conscious levels; they represent explicit (conscious) and implicit (subconscious) knowledge, and serve for inference, planning, and adaptive control.

The proposed hardware architecture implements a hierarchical conscious–subconscious processing, combining a fast and parallel subconscious level with an integrative and deliberative conscious level. The informational threshold module coordinates the two-way interaction, allowing global decisions to emerge only when informational integration exceeds a critical threshold $\Phi \geq \Phi_{crit}$. This structure ensures adaptability, information coherence, and energy efficiency, providing a scalable framework for service-oriented cognitive and multi-agent AI systems.

5. A Mathematical Model for Multi-Agent Systems Based on Conscious–Subconscious Hierarchical Computation

Multi-Agent Systems (MAS) are an essential framework for modeling and implementing complex distributed systems, characterized by autonomy, interaction and emergent behavior. As modern applications become increasingly dynamic and unpredictable, classical approaches, based exclusively on local rules or global optimization, prove insufficient to explain the adaptability and efficiency of natural systems.

In this context, a mathematical model of a Multi-Agent System based on a conscious–subconscious hierarchical calculation is proposed, inspired by the cognitive architecture of the human brain. The model distinguishes between a subconscious level, responsible for fast, parallel, and reactive processing, and a conscious level, dedicated to global integration, coordination, and deliberative decision-making.

By formally defining agents, internal states, and mechanisms of interaction, this framework provides a rigorous basis for the analysis of emerging behavior. It also allows the evaluation of the adaptability and efficiency of complex multi-agent systems.

We will consider a Multi-Agent system represented by the expression (13):

$$MAS = \langle A, E, I, T \rangle, \tag{13}$$

where:

$A = \{A_1, A_2, \ldots, A_M\}$ - is the set of agents;

$E \subset \mathbb{R}^M$ - represents the environment of the Multi-Agent system;

I - represents the interactions between agents;

T - is the temporal dynamics.

Each agent $A_j, \forall j = \overline{1, M}$ is defined as a hierarchical cognitive agent defined by the expression (14):

$$A_j = \langle S_j, C_j, U_j, \Pi_j \rangle, \tag{14}$$

where:

S_j - is the subconscious subsystem;

C_j - is the conscious subsystem;

U_j - is the learning mechanism;

Π_j - is the decision policy.

The hierarchical state space represents the organization of the set of states of a system on interdependent functional levels, allowing the description of multi-scalar dynamics and control and emergence mechanisms. The agent state $A_j, \forall j = \overline{1, M}$ is composed of two subspaces defined by the expression (15):

$$x_j(t) = \left(x_j^{sub}(t), x_j^{con}(t)\right), \tag{15}$$

where:

$x_j^{sub}(t) \in \mathbb{R}^m$ - is the subconscious state (implicit, continuous);

$x_j^{con}(t) \in \mathbb{R}^\kappa$ - is the conscious state (explicit, discrete);

provided that $m \gg \kappa$.

The dynamics of the subconscious correspond to a level of fast and parallel processing. These ensure short response time, prediction and pattern recognition, being characterized by stable local developments and automatic reactions that provide informational support for conscious processes. The dynamics of the subconscious are modeled as a continuous nonlinear dynamic system defined by the expression (16):

$$\dot{x}_j^{sub} = f_{sub}\left(x_j^{sub}(t), u_j(t), e_j(t)\right), \tag{16}$$

where:

$u_j(t)$ - the set of local stimuli;

$e_j(t) \in E$ - is the activity environment for the agent $A_j, \forall j = \overline{1,M}$.

The dynamics of consciousness correspond to a slow and deliberative level of processing, characterized by global integration of information, executive control and intentional decision-making. The conscious is modeled as a discrete system with a reduced state, based on the expression (17):

$$x_j^{con}(k+1) = f_{con}\left(x_j^{con}(k), y_j^{sub}(k)\right), \tag{17}$$

where:

$$y_j^{sub}(k) = P\left(x_j^{sub}(t_k)\right), \tag{18}$$

is the informational projection of the subconscious to the conscious (attention / selection).

The attention mechanism and cognitive threshold describe the processes by which information is selected and amplified, and only representations that exceed a certain level of activation or relevance become accessible to conscious processing. Subconscious–conscious transfer occurs only if condition (19) is met:

$$\left\| y_j^{sub}(k) \right\| > \theta_j, \tag{19}$$

where θ_j is the cognitive threshold of the agent, which formalizes the appearance of intuition, alerts or cognitive conflicts.

The hierarchical decision-making policy is a mechanism for selecting actions, structured on functional levels, which allows the coordination of quick local decisions with global deliberative decisions. The agent's decision is the result of a hierarchical merger, defined by the expression (20):

$$a_j(k) = \Pi_j\left(x_j^{con}(k), x_j^{sub}(t)\right), \tag{20}$$

where the subconscious $x_j^{sub}(t)$ provides quick suggestions and the conscious $x_j^{con}(k)$ strategically validates decisions.

Formally, the hierarchical decision-making policy can be calculated based on the expression (21):

$$\Pi_j = \alpha \Pi_{sub} + (1-\alpha)\Pi_{con}, \tag{21}$$

where $\alpha \in [0.7,...,0.95]$.

Conscious–subconscious learning and transfer is the process by which explicit cognitive strategies are internalized as implicit mechanisms, leading to behavior automation and increased adaptive efficiency. The learning process is modelled as a gradual consolidation and takes place according to the expression (22):

$$U_j:\ x_j^{sub}(t+1) = x_j^{sub}(t) = \eta \nabla J_j, \tag{22}$$

where:

J_j - is the performance function;

η - is the learning rate.

It has been shown that repeated conscious rules are internalized at the subconscious level and can be expressed by the expression (23):

$$x_j^{sub} \leftarrow \lim_{k\to\infty}\left(x_j^{con}(k)\right). \tag{23}$$

The interaction between agents represents the set of mechanisms through which autonomous agents exchange information and coordinate their actions, leading to the emergence of collective behavior. The interaction between agents takes place predominantly at the subconscious level and is shaped by the expression (24):

$$x_j^{sub}(t+1) = f_{sub}\left(x_j^{sub}(t), x_i^{sub}(t)\right). \tag{24}$$

At the same time, the coordination of interaction between agents at a global level is managed at a conscious level, based on the model in the expression (25):

$$x_j^{con}(k+1) = g\left(x_j^{con}(k), M_{ij}(k)\right), \tag{25}$$

where M_{ij} are the symbolic messages sent between agents.

The emergence of collective behavior is the process by which local interactions between autonomous agents lead to the emergence of coherent global patterns, without being explicitly imposed by centralized control. The emerging behavior of the Multi-Agent system is evaluated based on the expression (26):

$$B_{global} = \lim_{K\to\infty}\sum_{i=1}^{K}\Pi_i\ . \tag{26}$$

The emerging behavior of the Multi-Agent System results from local subconscious interactions and global strategic conscious interactions.

6. Hardware Architecture of the Multi-Agent Computing System

A complete hardware architecture is proposed for a service-oriented Multi-Agent Computing System, designed directly on the basis of the conscious–subconscious hierarchical mathematical model. This architecture explicitly translates the functional structure of agents and their interactions into hardware and software components, providing a robust framework for implementing distributed cognitive services.

Each agent integrates two levels of processing:

Subconscious level – dedicated to fast, parallel and energy-efficient processing. This layer is responsible for executing automatic reactions, learned patterns, and local inference, being implemented through massive parallel computing units such as GPUs, NPUs, or dedicated AI accelerators;

Conscious level – responsible for the global integration of information, deliberative decision-making and coordination of services at the system level. It is achieved through CPU/TPU control cores or specialized processors, which manage attention, cognitive thresholds, and hierarchical decision policies.

Communication between levels and agents is achieved through data buses, shared memory and asynchronous messaging, allowing dynamic coupling and the emergence of collective behavior without the need for rigid centralized control. Each agent is capable of acting autonomously, but integrated into an ecosystem of distributed services, favoring the scalability, adaptability and robustness of the system.

This approach allows the direct mapping of the conscious–subconscious mathematical model on heterogeneous infrastructures, facilitating the development of service-oriented, adaptive, scalable and energy-efficient multi-agent systems, capable of supporting complex cognitive services in distributed environments.

The hardware architecture of the service-oriented Multi-Agent Computing System is shown in Figure 2 (*generated with the support of AI ChatGPT-5.2*), where the following are mentioned:

Intelligent agent – is the basic unit of the multi-agent system. Each agent is autonomous, capable of perceiving the environment, making decisions and acting, integrating two distinct cognitive levels: conscious and subconscious. The Agents interact with each other to support distributed services and emerging behaviors at the system level;

Conscious level – intended for slow, sequential and energy-intensive processing. It functions as a central system, solving planning and deliberative adjustment tasks;

Deliberative decision – evaluates complex options, analyzes risks and plans strategic actions;

Planning and control – coordinates the agent's behavior in the medium and long term, integrating local and global objectives;

Global integration – synchronizes information from the subconscious level and from other agents, allowing coherent and adaptive decision-making;

Subconscious level – intended for massive, parallel, low-latency processing. It supports the automation of behaviors and the transfer of knowledge from the conscious level;

Fast processing – executes calculations and automatic reactions for immediate responses to stimuli;

GPU/NPU/AI accelerators – the subconscious level uses specialized computing units for parallel, energy-efficient processing;

Local inference/automatic reactions – implementation of learned behavioral patterns and reactive decisions, without deliberative involvement of the conscious level;

Communication and interaction – carries out the exchange of information between agents and between the conscious and subconscious levels, ensuring the emergence of collective behavior for local control and global coordination of services;

Distributed messaging – asynchronous communication between agents and distributed services;

Data bus – fast and structured transfer of data between hardware and software components;

Shared memory – allows shared access to information and synchronization of statuses between agents;

Hardware and platforms – allow the flexible implementation of the multi-agent architecture, adapted to the requirements of performance, latency and energy consumption;

Integrated SoC systems – compact, integrated systems for edge and intelligent IoT applications;

Edge Cluster – a group of distributed, fast-processing, and scalable servers located close to the data source;

HPC / TPC systems – high-performance computing infrastructure for large-scale intensive simulations and processing.

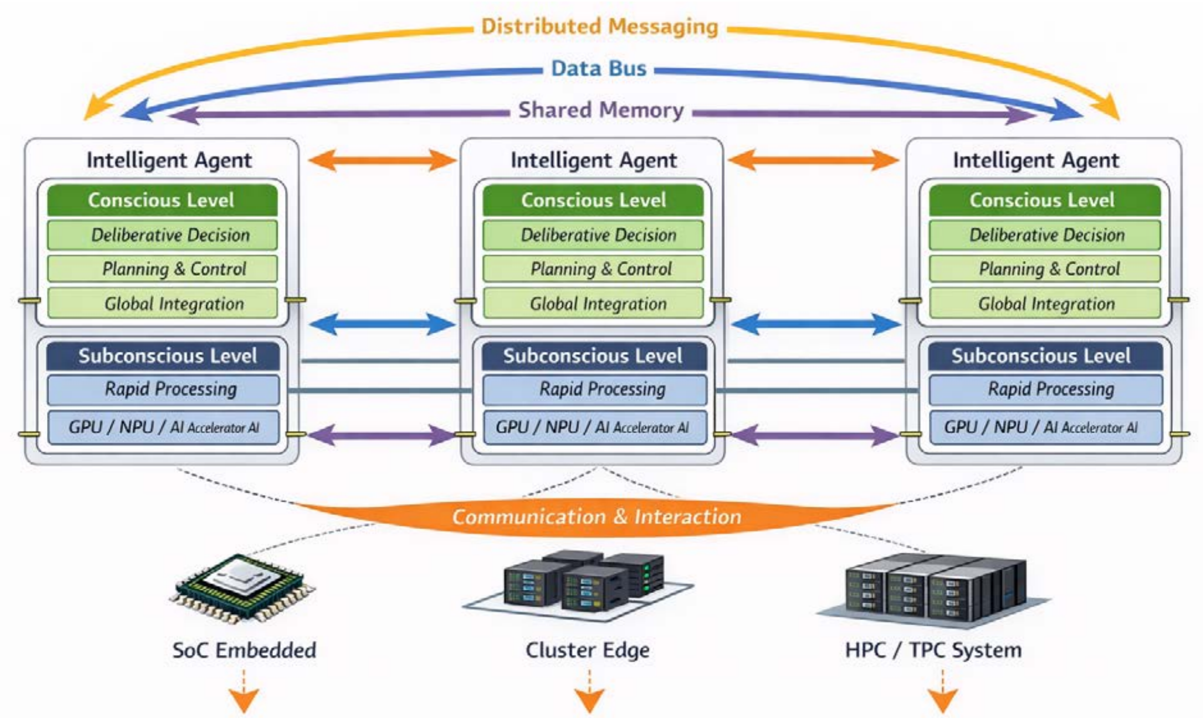

Figure 2. Hardware architecture of the Multi-Agent Computing System.

The proposed hardware architecture implements a Multi-Agent Computing System based on conscious–subconscious hierarchical processing, in which each agent is realized as a cognitive SoC, and global coordination is ensured by a collective conscious level.

7. Conclusions

This paper presented a comprehensive interdisciplinary framework for adaptive computational structures in service-oriented multi-agent systems, grounded in knowledge models inspired by the hierarchical organization of the human brain. By systematically integrating neurophysiological principles, mathematical formalization, and hardware-oriented architectural design, the proposed approach lays a unified foundation for next-generation intelligent systems.

Starting from the physiology of the human brain, the study highlights the functional asymmetry between conscious and subconscious processes, emphasizing the rapid, parallel and implicit processing at the subconscious level, in parallel with the strategic, deliberative and symbolic role of conscious processing. This biological perspective provides a scientifically justified paradigm for hierarchical information processing in artificial systems.

The mathematical modeling of conscious–subconscious interaction formalizes this paradigm through a two-level computational framework, in which subconscious dynamics are represented by continuous, adaptive and high-dimensional processes, while conscious processing functions as a discrete, low-dimensional level dedicated to control and decision-making. The introduction of attention and coupling operators allows for a dynamic information flow and adaptive tuning between these levels, ensuring both computational efficiency and interpretability.

Based on this formalization, a hardware architecture for hierarchical conscious–subconscious processing is proposed, demonstrating how heterogeneous computational elements—such as NPUs, GPUs, and CPUs—can be structurally and functionally aligned with distinct cognitive roles. This architecture supports real-time adaptive behaviors, scalability, and energy efficiency, reflecting essential properties of biological cognitive systems.

Extending the model from the individual level, a mathematical framework for multi-agent systems is developed, in which each agent incorporates a local conscious–subconscious hierarchy, and coordination at the higher level emerges through collective conscious mechanisms. This formulation allows for distributed decision-making, cooperative adaptation and knowledge sharing in service-oriented environments, without introducing centralized control blockages.

Finally, the hardware architecture of the Multi-Agent Computing System demonstrates the feasibility of practical implementation of the proposed model in real computational infrastructures. By combining autonomous cognitive agents with a level of collective coordination, the architecture supports modularity, fault tolerance and dynamic service composition, essential properties for modern distributed and intelligent systems.

In conclusion, the proposed adaptive computing structures represent a significant step towards cognitive, knowledge-oriented multi-agent systems, in which neurobiological inspired hierarchy, rigorous mathematical modeling and conscious hardware architecture design converge in a coherent framework. Future research directions are aimed at experimental validation, analysis of the dynamics of learning processes at the multi-agent level, as well as the integration of ethical and explainability constraints in the conscious decision layer.

Acknowledgments: This research was carried out within the 020404 Institutional Project, with the technical and technological support provided by the "Artificial Intelligence and Multi-Agent Systems" Laboratory, within the Department of Informatics and Systems Engineering, Technical University of Moldova.

During the preparation of this study, the authors used ChatGPT-5.2 for grammar checking, correct wording of text, and evaluation of the correctness of scientific results. The authors have reviewed the result and take full responsibility for the content of this publication.

Contribution of authors:

Vadim Struna: Conceptualization, Methodology, Resources, Writing – original draft.

Victor Ababii: Conceptualization, Formal analysis, Methodology, Resources, Writing – original draft.
Viorica Sudacevschi: Methodology, Project administration, Writing – review & editing.
Silvia Munteanu: Formal analysis, Resources.
Olesea Borozan: Formal analysis, Resources, Writing – review & editing.
Victoria Alexei: Validation, Investigation, Writing – review & editing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Borangiu, T.; Trentesaux, D.; Leitão, P.; Cardin, O.; & Lamouri, S. (Eds.). (2021). *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future: Proceedings of SOHOMA 2020* (Vol. 952). Springer Nature, 544 p., https://doi.org/10.1007/978-3-030-69373-2.
2. Li, A.; Xie, Y.; Li, S.; Tsung, F.; Ding, B.; & Li, Y. (2024). Agent-oriented planning in multi-agent systems. *arXiv preprint arXiv:2410.02189*, https://doi.org/10.48550/arXiv.2410.02189.
3. Wold, H. (1982). Models for knowledge. *The making of statisticians*, pp. 189-212, Springer, New York, NY. https://doi.org/10.1007/978-1-4613-8171-6_14.
4. Antonelli, C. (2005). Models of knowledge and systems of governance. *Journal of institutional economics*, *1(1)*, pp. 51-73, https://doi.org/10.1017/S1744137405000044.
5. Munir, K.; & Anjum, M. S. (2018). The use of ontologies for effective knowledge modelling and information retrieval. *Applied computing and informatics, 14(2)*, pp. 116-126, https://doi.org/10.1016/j.aci.2017.07.003.
6. Haslinda, A.; Sarinah, A. (2009). A review of knowledge management models. *Journal of international social research, 2(9)*, pp. 187, ISSN 1307-9581.
7. Noy, N.F.; Fergerson, R.W.; Musen, M.A. (2000). The knowledge model of Protégé-2000: Combining interoperability and flexibility. *In International conference on knowledge engineering and knowledge management*, pp. 17-32, Berlin, Heidelberg: Springer Berlin Heidelberg, https://doi.org/10.1007/3-540-39967-4_2.
8. Ababii, V.; Sudacevschi, V.; Munteanu, S.; Carbune, V.; Borozan, O. (2025) Multi-Coalition Multi-Agent decision making system synthesis. *In International Journal of Computing, 24(3)*, pp. 513-519, ISSN 1727-6209, https://doi.org/10.47839/ijc.24.3.4188.
9. Ababii, V.; Cărbune, V.; Sudacevschi, V.; Marusic, G.; Braniște, R.; Drumea, N. (2025) Sistem Multi-Agent pentru monitorizarea și predicția proceselor de mediu. *Akademos, 1(76)*, pp. 22-30, https://doi.org/10.52673/18570461.25.1-76.01.
10. Struna, V.; Borozan, O.; Carauș, A. (2023) Knowledge models for evolutionary systems with artificial intelligence. *In Proceedings of the Twelfth Conference "Informatics and Computer Technologies Problems" (ICTP – 2023)*, pp. 38-41, Chernivtsi, Ukraine, 10−12 November, 2023.
11. Ababii, V.; Strună, V.; Sudacevschi, V.; Borozan, O.; Munteanu, S. (2024) Method for knowledge acquisition based on image processing for decision-making systems. *In Electronics, Communications and Computing (IC ECCO-2024): The conference program and abstract book: 13th intern. conf.,* pp. 165-166, Chișinău: Tehnica-UTM, 17-18 Oct. 2024. Technical University of Moldova, ISBN: 978-9975-64-480-8 (PDF).
12. Struna, V.; Ursu A.; Kapusteanski, M. (2024) Analiza modelelor de cunoștințe pentru sisteme cu inteligență artificială. *In Technical Scientific Conference of Undergraduate, Master and PhD Students*, Universitatea Tehnică a Moldovei, 27-29 Martie 2024. Chișinău, 2024, vol. 1, pp. 392-394. ISBN 978-9975-64-458-7, ISBN 978-9975-64-459-4.
13. Baars, B. J.; Franklin, S.; Ramsoy, T. Z. (2013) Global workspace dynamics: Cortical "binding and propagation" enables conscious contents. *Frontiers in Psychology*, 4, 200. https://doi.org/10.3389/fpsyg.2013.00200.
14. Clark, A. (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001.
15. Damasio, A. (2018) *The Strange Order of Things: Life, Feeling, and the Making of Cultures*. NY: Pantheon Books, 336p.
16. Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking Press, 363p., ISBN: 978-0-698-15140-6.

17. Friston, K. (2019) A free energy principle for a particular physics. *arXiv preprint*, arXiv:1906.10184. https://doi.org/10.48550/arXiv.1906.10184.
18. Gazzaniga, M. S.; Ivry, R. B.; Mangun, G. R. (2019) *Cognitive Neuroscience: The Biology of the Mind* (5th ed.). W. W. Norton & Company, ISBN: 978-0-393-60317-0.
19. Kandel, E. R.; Koester, J. D.; Mack, S. H.; Siegelbaum, S. A. (2021) *Principles of Neural Science* (6th ed.). McGraw-Hill Education, 108p., ISBN: 978-1-25-964224-1.
20. Mashour, G. A.; Roelfsema, P.; Changeux, J.-P.; Dehaene, S. (2020) Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5), pp. 776-798. https://doi.org/10.1016/j.neuron.2020.01.026.
21. Pessoa, L. (2017) A network model of the emotional brain. *Trends in Cognitive Sciences*, 21(5), pp. 357–371. https://doi.org/10.1016/j.tics.2017.03.002.
22. Seth, A. K.; Bayne, T. (2022) Theories of consciousness. *Nature Reviews Neuroscience*, 23, pp. 439–452. https://doi.org/10.1038/s41583-022-00587-4.
23. Stanley, D. A.; Adolphs, R. (2013) Toward a neural basis for social behavior. *Neuron*, 80(3), pp. 816–826. https://doi.org/10.1016/j.neuron.2013.10.038.
24. Tononi, G.; Boly, M.; Massimini, M.; Koch, C. (2016) Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), pp. 450–461. https://doi.org/10.1038/nrn.2016.44.
25. Friston, K. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), pp. 127–138. https://doi.org/10.1038/nrn2787.

Citation: Struna, V., Ababii, V., Sudacevschi, V., Munteanu, S., Borozan, O. and Alexei, V. (2026). Adaptive computing structures for service-oriented multi-agent systems based on knowledge models. *Journal of Engineering Science*. 2026, 33 (1), pp. 37-52. https://doi.org/10.52326/jes.utm.2026.33(1).03.

Submission of manuscripts: jes@meridian.utm.md