UDC: 004.93 https://doi.org/10.15407/jai2023.03.070

S. Kornieiev
Baltrobotics Sp.z.o.o., Poland
3, Trzy Lipy, Gdansk, 80 -172
sergii.kornieiev@baltrobotics.com
https://orcid.org/0000-0003-3969-0304

AGI-AGENT COGNITIVE ARCHITECTURE AGICA - AXIOMATIC APPROACH

Abstract. For the last half of the century there were proposed and modeled several dozen *cognitive architectures* as the models of mind. As one of the results of this *Standard Model of the Mind* was proposed and discussed in 2017 (now known also as "Common Model of Cognition"). It accumulated lessons learned in one structure. In the articles published in 2016-2018, the author formulated main definitions of the concepts of Artificial General Intelligence (AGI): *AGI-Individual Type, AGI-Collective Type, AGI-Consciousness, AGI-Thought, AGI-Knowledge, AGI-Emotions*. The author's approach belongs to the direction Embodied Cognition in Cognitive Science and is following named "axiomatic approach" in Artificial Intelligence. The definitions proposed by the author are of constructive type from mathematical point of view and can be modeled by the existing software & hardware tools and methods. In this article the author is proposing *AGI-Agent cognitive architecture* AGICA as detailed modification of *Standard Model of the Mind*. It can be used in the development of *universal operating system* for AGI-robots.

Keywords: artificial intelligence, artificial general intelligence, cognitive architecture, operating system.

Introduction

The development of intellectual system was divided to two direction, namely: *Artificial Intelligence* – this name combines the set of types of research directions and systems that can partially fulfill the tasks commonly belong to people experience (such as natural language processing, pattern recognition, generative artificial intelligence, expert systems); and *Artificial General Intelligence* – it is the development of the systems that can operate autonomously and in the result should deliver the machines that can replace people from routine operations in the first phase and further can substitute people in any activities.

This article concerns the development of *Artificial General Intelligence* (AGI) and devoted to *AGI cognitive architecture*.

In the last 30-40 years there were proposed hundreds and developed tenths of cognitive architectures as the models of mind – as "humanlike mind". The Institute for Creative Technologies defines *cognitive architecture* as: "hypothesis about the fixed structures that provide a mind, whether in natural or artificial systems, and how they work together – in conjunction with knowledge and skills embodied within the architecture – to yield intelligent behavior in a diversity of complex environments." [1].

In [2] there were analyzed 84 architectures including 49 that were still actively developed in 2018. In the references mostly mentioned are ACT-R, SOAR, CLARION, ICARUS, EPIC, LIDA. In [12] there were listed and reviewed in short 26 cognitive architecture were developed before 2010. In [47] there is good, compressed overview of cognitive architectures: ACT-R, SOAR, EPIC, PRODIGY, DEM, COGNET, and CLARION, there features, advantages, and disadvantages.

The most successful there were SOAR [3] and ACT-R [4] but in full list there are more 50 projects. In 2017 on the base of analysis of the results of these projects and lessons learned there was formulated "The Standard Model of the Mind" as *Common Computational Framework Across Artificial Intelligence*, *Cognitive Science, Neuroscience, and Robotics* [5].

In this article the author proposed more detailed development of *Standard Model of the Mind*.

In the previous published articles [6, 7, 8] the author proposed the definitions of main concepts of Artificial General Intelligence: *AGI-Individual Type*, *AGI-Collective type*, *AGI-Consciousness*, *AGI-Thought*, *AGI-Emotions*, *AGI-Knowledge,* and others. The author approach follows named "axiomatic approach" proposed in [9, 10] and belong to *embodied cognitio*n [11] as one of three directions of Cognitive Science: *computationalism*, *connectionism*, *embodied cognition*.

The definitions proposed by the author are of *constructive type* from mathematical point of view and can be modeled by the existing software & hardware tools and methods.

In this article the author is proposing the view on *AGI-agent structure and procedures* as detailed modification of *Standard Model of the Mind* formulated before [5]. The results of the article can be used in the development of *universal operating system* for AGI-robots.

Terminology.

In the area of research of artificial intelligence there are used many concepts common with the psychology, cognitive science, brain research. The author sees it is useful to make some short glossary of the concepts and abbreviations used in the article.

Artificial Intelligence – this name combines the set of research directions and types of systems that can partially fulfill the tasks commonly belong to people experience (such as natural language processing, pattern recognition, generative artificial intelligence, expert systems).

AGI – artificial general intelligence, - scientific direction concerning of the development fully autonomous intellectual machines; "intellectual" means that the machine can do the tasks that now can be done only by people.

AGI-agent – application software downloaded to CPU of the *AGI-robot* under the supervision of *operating system*; it enables *AGI-robot* to operate in the environment in fully independent mode without any remote control from an operator; it perceives its environment, takes actions autonomously to achieve goals, and may improve its performance with learning or acquiring knowledge.

AGI-robot – the autonomous machine with *AGI-agent* downloaded - "hardware and software" - fully equipped and ready for operations in fully independent mode without any remote control from an operator; it is assumed that *AGI-robot* can move in the space and can interact with the environment by sensors, actuators, and manipulators.

AGI operating system – operating system that run *AGI-agent*.

Problem

Currently the world is very impressed with the results in the development of *Generative AI* like ChatGPT and others of such type. At the same time the development of *autonomous robots* based on the concept of AGI is still in the very beginning. The *cognitive architectures* proposed, developed, and tested before delivered mostly scientific results about general principles of cognition, decision making and learning. These results are still far away from the market products where we can see from the direction only *autonomous vacuum cleaners*.

On the base of developments of around a hundred cognitive architectures [2] there was the attempt to form the *Standard Model of Mind*. In [5] there was concluded that *Standard Model of Mind* reflects a very real consensus over the assumptions it includes, but it remains incomplete in several ways. As these there was mentioned: metacognition, emotion, mental imagery, direct communication and learning across modules, the distinction between semantic and episodic memory, and mechanisms necessary for social cognition.

From author's point of view, the *Standard Model* can be enlarged with the approaches of *Embodied Cognition* [11] and *Axiomatic Approach* in AI [6-10]. In this article the author will propose his view how it can be done.

Current state analysis

The research on cognitive architectures as software instantiation of cognitive theories was initiated by Allen Newell in 1990 [13].

More successful projects of cognitive architectures include ACT-R ("adaptive control of thought – rational") [4] and SOAR ("state-operator-and-results") [3].

In 2017 on the base of the development of ACT-R and SOAR and some others there was proposed "A Standard Model of the Mind" [5] (see Fig.1). Later this model was named "Common Model of Cognition" (CMC).

The heart of the *Standard Model* is the *cognitive cycle. Procedural memory* induces the processing required to select a single *deliberate act* per cycle. Each *action* can perform multiple modifications to *working memory*. Changes to *working memory* can correspond to a step in *abstract reasoning* or the *internal simulation* of an *external action*, but they can also initiate the *retrieval of knowledge* from *long-term declarative memory*, initiate *motor actions* in an external environment, or provide top-down influence on *perception*.

Complex behavior, both external and internal, arises from sequences of such *cognitive cycles*.

The restriction to selecting a single *deliberate act* per *cycle* yields a serial "bottleneck" in performance, although significant parallelism can occur during *procedural memory's internal processing*. Significant parallelism can also occur across components, each of which has its own time course and runs independently once initiated. The details of the internal processing of these components are not specified as part of the *standard model*, although they usually involve significant parallelism.

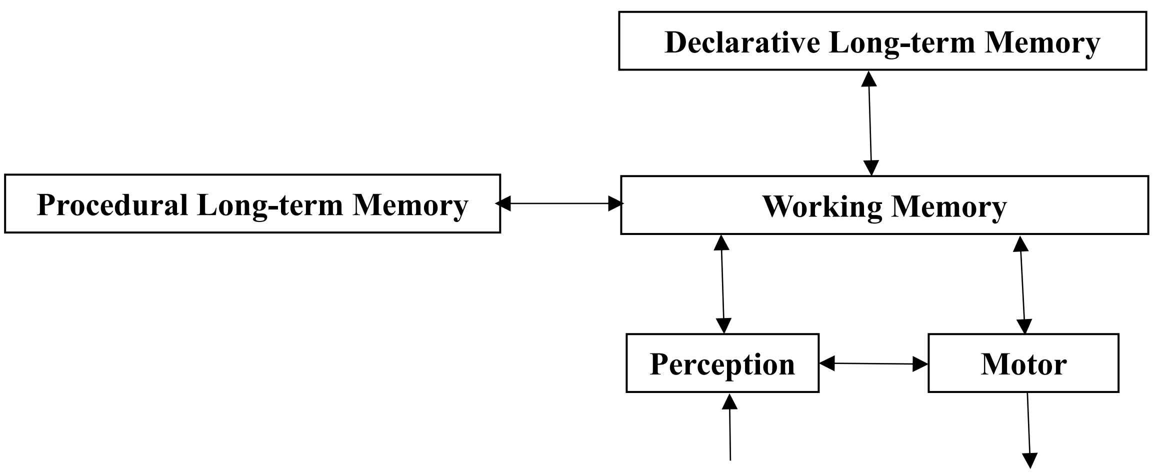

Fig. 1. The Structure of the Standard Model of the Mind [5]

Standard Model.

The *cognitive cycle* that arises from *procedural memory's interaction* with *working memory* provides the seriality necessary for coherent thought in the face of the rampant parallelism within and across components.

Procedural memory contains knowledge about actions, whether internal or external. This includes both how to select actions and how to cue (for external actions) or execute (for internal actions) them, yielding what can be characterized as *skills* and *procedures*.

Declarative memory is a long-term store for *facts* and *concepts*. It is structured as a persistent graph of symbolic relations, with *metadata* reflecting attributes such as recency and frequency of (co-)occurrence that are used in *learning* and *retrieval*. *Retrieval* is initiated by the creation of a *cue* in the designated *buffer* in *working memory*, with the result being deposited in that *buffer*. In addition to facts, *declarative memory* can also be a repository of the system's direct experiences, in the form of *episodic knowledge*. There is not yet a consensus concerning whether there is a *single uniform declarative memory* or whether there are two memories, one *semantic* and the other *episodic*. The distinction between those terms roughly maps to *semantically abstract facts* versus *contextualized experiential knowledge*, respectively, but its precise meaning is the subject of current debate.

Learning involves the automatic creation of new symbol structures, plus the tuning of *metadata*, in *long-term — procedural* and *declarative — memories*. It also involves adaptation of *nonsymbolic content* in the *perception* and *motor systems*. The *standard model* assumes that all types of *long-term knowledge* are learnable, including both *symbol structures* and *associated metadata*. All *learning* is incremental and takes place online over the experiences that arise during system behavior.

There are at least two independent learning mechanisms for *procedural memory*: one that creates new *rules* from the composition of *rule firings* in some form, and one that tunes the selection between competing *deliberative acts* through *reinforcement learning*.

Declarative memory also involves at least two learning mechanisms: one to create new *relations* and one to tune the associated *metadata*.

Outside of direct connections between the *perception* and *motor modules*, *working memory* acts as the intercomponent communication *buffer* for components. It can be considered as unitary or consist of separate *modality-specific memories* (for example, verbal, visual) that together constitute an aggregate *working memory*.

Long-term declarative memory, *perception*, and *motor modules* are all restricted to accessing and modifying their associated *working memory buffers*, whereas *procedural memory* has access to all *working memory* (but no direct access to the contents of *long-term declarative memory* or itself).

All *long-term memories* have one or more associated *learning mechanisms* that automatically *store*, *modify*, or *tune* information based on the architecture's processing.

Perception and Motor.

Perception converts *external signals* into *symbols* and *relations*, with associated *metadata*, and places the results in specific *buffers* within *working memory*. There can be many different *perception modules*, each with input from a different modality — vision, audition — and each with its own perceptual *buffer*. The *standard model* assumes an "attentional bottleneck" that constrains the amount of information that becomes available in *working memory* but does not embody any commitments as to the internal representation (or processing) of information within perceptual modules, although it is assumed to be predominantly nonsymbolic in nature, and to include learning. Information flow from *working memory* to perception is possible, providing expectations or possible hypotheses that can be used to influence perceptual classification and learning.

Motor converts the *symbol structures* and their *metadata* that have been stored in their *buffers* into *external action* through control of whatever effectors are a part of the *body* of the system. As with *perception*, there can be multiple *motor modules* ("arms", "legs", …). Much is known about motor control from the robotics and neuroscience literature, but there is at present no consensus as to the form this should take in the *standard model*, largely due to a relative lack of focus on it in "humanlike" architectures.

In ACT–R architecture ("Adaptive Control of Thought–Rational") [4] there are the following modules: Central Production System, Goal module, Perceptual-motor modules, Memory modules. There are two kinds of memory modules in ACT-R: Declarative memory (consisting of facts such as "Washington, D.C. is the capital of United States", etc.); Procedural memory (made of productions). Productions represent knowledge about how we do things: for instance, knowledge about how to drive, etc.

Reflecting the growing influence of neuroscience constraints, the attempt to implement the architecture using standard *connectionist* constructs had been made [14]. Although the resulting system, ACT-RN, was not of practical use, this connection between symbolic and connectionist levels had a fundamental impact on the development of the architecture [15].

SOAR ("State-Operator-and-Result") [3] is a cognitive architecture, originally created by John Laird, Allen Newell, and Paul Rosenbloom at Carnegie Mellon University. SOAR is also comprised of a set of asynchronous internally parallel modules, including a *rule-based procedural memory*. SOAR is organized around a *working memory*. It includes separate *episodic* and *semantic declarative memories*, in addition to *visuospatial modules* and a *motor module* that controls robotic or virtual effectors. *Knowledge* that exists independent of the current situation is held in the architecture's *long-term memory*. SOAR distinguishes three different types of *long-term memory*: *procedural*, *semantic*, and *episodic*.

The list of the most known cognitive architectures can include the following: CLARION [16], LEABRA [17], Spaun [18], CHREST [19], EPIC [20], Companions [21, 22], MicroPsi [23], MIDCA [24, 25], PolyScheme [26, 27], LIDA [28-30], ICARUS [31], SAL [32], SiMA [33], SIGMA [34, 35].

Lessons learned from cognitive architecture development

To resume the main artefacts of cognitive architectures the following features and components that are common for many cognitive architectures: *working memory* (includes active mental states), *procedural memory* (includes primitives), *semantic memory* (includes schemas), *episodic memory* (includes inactive mental states aggregated into episodes), *iconic memory* (input-output buffer), *cognitive map*, *reward system*, *attention control* (the distribution of attention is described by a special attribute in instances of schemas) and *consciousness* (can be identified with the content of the mental state "I-Now").

Since, by definition, *working memory* is a relatively small temporary storage, its capacity should be limited. However, there is no consensus on of how this should be done in a cognitive architecture. There different approaches to clean working memory: (1) the contents of *working memory* is discarded when the agent switches to a new problem; (2) gradually remove items from the memory based on their recency or relevance in the changing context; (3) assigning an activation level to percepts that is proportional to the emotional valence of a perceived situation, - this activation level changes over time and as soon as it falls below a set threshold, the percept is discarded; (4) atoms stay in memory as long as they build links to other memory elements and increase their utility, and other approaches are used.

Long-term memory (LTM) preserves a large amount of information for a very long time. Typically, it is divided into the *procedural memory* of *implicit knowledge* (e.g. motor skills and routine behaviors) and *declarative memory*, which contains *explicit knowledge*. The latter is further subdivided into *semantic* (factual) and *episodic* (autobiographical) *memory*.

Long-term memory is a storage for *innate knowledge* that enables operation of the system, therefore almost all architectures implement *procedural* and/or *semantic memory*.

Procedural memory contains knowledge about how to get things done in the task domain. In symbolic production systems, *procedural knowledge* is represented by a set of "*if-then rules"* preprogrammed or learned for a particular domain.

Semantic memory stores facts about the objects and relationships between them. In the architectures that support *symbolic reasoning*, *semantic knowledge* is typically implemented as a graph-like ontology, where nodes correspond to *concepts* and links represent *relationships* between them.

Episodic memory stores specific instances of past experiences. These can later be reused if a similar situation arises or exploited for learning new *semantic/procedural knowledge*.

Aleksander and Dunmall [9, 10] have developed an *Axiomatic Approach* to machine consciousness based around five axioms, which they believe are minimally necessary for consciousness:

1. Depiction. The system has perceptual states that 'represent' elements of the world and their location.
2. Imagination. The system can recall parts of the world or create sensations that are like parts of the world.

3. Attention. The system can select which parts of the world to depict or imagine.

4. Planning. The system has control over sequences of states to plan actions.

5. Emotion. The system has affective states that evaluate planned actions and determine the ensuing action.

The author proposed close but different view on "axiomatic approach" to model AGI of *Individual* and *Collective type*. According to the author's articles [6, 7, 8] there were proposed the definitions of main structure artefacts of AGI-Agent: "AGI-hierarchical structure" composed of the "AGI-Collective Type" and "AGI-Individual Type", wherein the "AI-Collective Type Processes Functionality" are based in the object of "AGI-Individual Type" as "*other agent"*; and the *agent* of "AGI-Individual Type" ("Agent 1") is dominated by the *agent* of "AGI-Collective Type" (Agent 2) by the mechanism proposed below [6] or more strong of more soft its modifications.

The definition of *AGI-Individual Type* is based on the concept of the "*instinct of individual self-preservation"* (as an analogy with "living intelligence") and on the concept of "self-controlled border" [6].

The definition of *AGI-Collective Type* is based on the concept of "*preservation of species instinct"* (as an analogy with "living intelligence") [6].

The concept of "self-controlled border" implies that the *AGI-Agent 1 of individual type* can receive from this *border* identified signals of such types: *invasion signals*, *signal of touch* by external objects and the environment objects, *signals of damage* of the *border* and signals of changing of its parameters under the influence of the environment and due to internal processes.

Human being has skin that is controlled by *peripheral nervous system* with *sensory nerves* [39]. Our skin has many functions. In the initial phases of the development of AGI-robots we do not need to model all of them, - it will be enough to control the events of touch, tear, or may be closeness.

In the world there are currently several projects of "artificial skin" development in different phases and with different results as concerned the sensors and the possibilities to control the parameters. For example, engineers at UC Berkeley have developed a pressure-sensitive electronic material from semiconductor nanowires [39]. The researchers at Stanford University have spent nearly two decades trying to develop skin-like integrated circuits that can be stretched, folded, bent, and twisted – working all the while [40]. In [41] and in [42] there is the list of current research.

The author is sure that in some nearest future it can be accessible to use artificial skin to control the physical border of AGI-robot. Till the moment it can be used CAD-model of the external view of AGI-robot and AGI-agent will calculate the mutual positions of the obstacles and its own case ("body") by using of common sensors like cameras or lasers.

AGI-Consciousness [8].

AGI-Consciousness is the combination of the following procedures during the existence of AGI-robot:

1) Initial procedure for determining and establishing of control over the *border* of *AGI-robot* (see the definition of "AGI-Individual Type" above); it can be done, for example, as to determine some character points on "artificial sensitive skin" of AGI-robot and 2D coordinate system of the "skin" (as a hypersurface in 3D environment); or it can be the linking of the 3D-coordinate system of the visual sensor to the CAD model of AGI-robot preliminary downloaded; or some other approaches can be used.

2) Permanent process of sensing of the «border» of AGI-robot or control it for: touching, penetration or tearing.

3) Permanent process of control of the functions of AGI-Individual Type for consistency (namely: *function of homeostasis*; *information function* (in particular, forecasting of the environment and AGI-robot own state); *function of imagination and planning*; *function to change AGI-robot own state*; *function of the environmental impact*; *objective function*).

4) Permanent process of control of the functions of AGI-Collective Type. Definition of the concept "AGI-Thought" [7].

AGI-Thoughts can be of the following types:

– *Thought Type 1* – "Imagined scenario" that was formed by: objects, concepts, and activities (virtual) "of them" and "with them".

– *Thought Type 2* – "Analysis of properties/characters and their relations – "*in static*" or "*for period time*".

– *Thought Type 3* – "Concepts/object definitions".

– *Thought Type 4* – "Classification/Pattern recognition".

– *Thought Type 5* "Reasoning/Conclusion".

– *Thought Type 6* – "Decision "to do" the real activities according to the Thought Type 1 in the State 2 (see below)".

Every *Thought Type* can be in one of two states:

– *Thought Dynamic State* – in active software process.

– *Thought Static State* – thought properties/data had been stored in the memory after the active program process had been completed.

The *Thought Type 1* in the *Thought Static State* can be implemented as the "*motivation to do"* and "*scenario/plan/schedule"* of real activities of *AGI-robot* with the correspondent decision – the *Thought Type 6* in the *Thought Static State*.

Definition of concept *AGI-Emotions* [7].

Any *AGI-Emotions* are changing the *homeostasis* of *AGI-robot* either to reach more powerful its internal state - to be able to attract more *energy* for an action in some cases, - or to return to some average energy supplying – "quiet state", "normal mode" - in the case that additional energy is not needed.

Emotions can be divided on "*Negative*" and "*Positive*". Currently the author can see the reason to model in AGI-Agent only *Negative Emotion,* and it is "AGI-Fear" – to enlarge the energy mobilization in some dangerous situation. But in the future, another emotion can also find justification.

The *negative emotion of AGI-Individual Type – "AGI-Fear"* - occurs in *AGI-robot* when it assumes the existing danger for its existence, and when in the input of AGI-robot (or in its thoughts ("imagination")) there is sharp increasing of information; this can be resulted also in the decreasing of *objective function*. The *negative emotions of individual type "AGI-Fear"* can be resulted by AGI-robot as the selection of thoughts/actions "stop-run-or-fight" in some emergency.

One means to increase the AGI-robot's ability to withstand danger is to increase the *clock frequency of the processor*. Also, with this emotion AGI-*Fear* all tasks do not concern the surviving should be abandoned.

As it was mentioned before we will consider the simplest model of AGI-robot and because of it we do not need to use "*Positive Emotions of Individual Type*" and all types of emotions *of AGI-collective type* that was defined in [7] for general case.

The *Rate of the Emotions* - or *negative* or *positive* ones – is dependent from 3 factors: 1) the estimated state of AI-System (safety or danger); 2) the input information measure ("more information – more rate"); and 3) the changing of *objective function* of AI-Entity" [6].

In [44] there were analyzed the features and properties of *AGI-operating system*. Of course, nobody will wait till this new kind of operating system will be developed. But the author is sure that early or later the new platform for AGI-robots will be designed. In [45] the author will emphasize the desired characteristics of the new operating system for AGI development.

Goal

The *Embodied Cognition* as one of three directions of *Cognitive Science* (with other ones: *Computationalism* and *Connectivism*), which was developing from 90th of the last century [11], "Axiomatic Approach" that had been proposed by Alexander in the beginning of this century [9, 10], and the development of "axiomatic approach" made by the author in the articles mentioned here [6-8] deliver the possibility to enlarge the *Standard Model of*

Cognition with new architectural details as it was represented below.

From the authors point of view *cognitive architectures* with the goal of design AGI-robots software that were developed before lack of implementation of the results of *Embodied Cognitions* as the third direction of *Cognitive Science* that emerged after *Computationalism* and *Connectivism* in the last 30 years, - most of the system does not have *Body* or *Boder* as one of the main elements of AGI-model.

The next proposal of the author is to use *AGI-Thoughts* according to the author's definition as the unified mechanism of "thinking" of AGI-robot. This also delivers the mechanism of *explanations* of the activities of AGI-robot: the set of *AGI-Thoughts* for some period of time can be simply transferred by AGI-Robot to be analyzed by the management entity.

In systems mentioned before the *consciousness* was modeled indirectly. In *cognitive architecture* AGICA the author proposed to model it directly according to the definition made [8,6].

According to author's point of view *Emotions* <u>should not be the *models*</u> but they <u>should be the *modes*</u> of AGI-Robot. The author cannot see the reason to implement all or large part of the *emotions* known in human beings, but the mode "AGI-Fear" can be useful.

The author proposes to implement "sleeping mode" of AGI-Robot to use it for associations establishing, classification, and concepts design, when the sensors' data are ignored.

This article is about the structure of *cognitive architecture* AGICA. Most functioned used to represent AGICA is known, and there are many models to implement them are in literature. The exception is *AGI-Consciousness* [8].

Main part

On the base of previous review, the *Artificial General Intelligence Cognitive Architecture* ("AGICA") can be proposed. The *functional view* of AGICA *cognitive architecture* is represented in Fig.2.

AGICA includes the following structure components:

– AGI-Consciousness,

– AGI-Individual type (the functions are divided to 2 groups: Group 1 – under control of AGI-Consciousness; Group 2 – under control of *AGI-Unconsciousness (function of homeostasis)*), - Agent 1,

– AGI-Collective type, - Agent 2,

– AGI-Unconsciousness,

– Sensors (with *border states* identification sensors: "touch", "close", "breach"),

– Motors.

These *processes* should operate under control of *AGI-operating system*.

AGI-Consciousness functions (see the definition of *AGI-Consciousness* mentioned above):

– Initial procedure for determining and establishing of control over the *border* of *AGI-robot*.

– Permanent process of sensing of the *border* of *AGI-robot*: touching, penetration, and tearing,

– Goal control,

– Control of the functions of *AGI-Individual Type* for consistency,

– Control of *Sleeping mode* (see below), (to start *Working mode* either after time period, or on the base of *Sensors' data* analysis, or on other conditions).

AGI-Individual type functions (see the definition mentioned above):

– Functions of homeostasis (are under control of *AGI-Unconsciousness*, some functions of homeostasis belong to automatic control),

– Information function (under *AGI-Consciousness* control),

– Forecasting (under *AGI-Consciousness* control),

– Function of imagination and planning (under *AGI-Consciousness* control),

– Function to change *AGI-Agent* own state (under *AGI-Consciousness* control),

– Function of the environmental impact (can operates or under *AGI-Consciousness* control or under *AGI-Unconsciousness* control),

– *Objective function* (under *AGI-Consciousness* control).

Function of homeostasis operates under *AGI-Unconsciousness* control and must keep the main internal physical parameters of the system in the assigned ranges: temperature, pressure, humidity.

Function to change AGI-Agent own state operates under *AGI-Consciousness* control. The states of AGI-Agent should be dependent from the modes.

The *modes* of *AGI-Agent* are:

– "Working mode" – to try to achieve the set *goal*,

– "Unsupervised learning" – it can be fulfilled in parallel with *working mode* with less priority,

– "Sleeping mode" – for autonomous data processing in the memory (learning from experience; associations; concepts development; "forgetting");

– "Supervised learning" – the mode when AGI-Agent "trusts" the trainer.

Information function operates under *AGI-Consciousness* control maintaining and ensuring (1) permanent collection of information; (2) its storage; (3) classification; (4) formation of abstract concepts, (5) their hierarchy and associative links. The last (3), (4) and (5) mainly must be performed in *sleeping mode* when *AGI-Consciousness* isolates *Information function from sensors' data*. In the *sleeping mode* the signals from *Environment sensors* are ignored by *Information function* and the data collected during the last *working mode* are analyzed for data mining and detection of associations.

In principle, for critical applications there can be designed "doubled system" - "one is working one is sleeping", – the dolphins brain uses such approach, but it will be much more complex than unitary system.

Forecasting function operates under *AGI-Consciousness* control and makes the forecasts of environment for any time periods – *short term* and *long term*.

Function of imagination and planning operates under *AGI-Consciousness* control. This function can be activated as in any *modes*.

Function of the environmental impact can operate either under *AGI-Consciousness* control - *Motors* routine operations, or under *AGI-Unconsciousness* control - for and emergency cases.

Objective function operates under *AGI-Consciousness* control.

AGI-Unconsciousness functions:

– Homeostasis control,
– Motors routine and emergency control,
– Emotion control,
– Data mining, associative memory development (in "sleeping mode").

AGI-Unconsciousness functions are the functions of autonomous control system of *AGI-robot* – they are independent from *AGI-Consciousness,* but they can cooperate to train new *skills*.

The core of *AGI-Unconsciousness functions* is of "innate type". But the author can imagine the cooperation of *AGI-Consciousness* and *AGI-Unconsciousness* for the *Function of the environmental impact*: when some procedures of motions or some specific reactions to environment events will be learned under *AGI-Consciousness control*, these procedures can be transferred to *AGI-Unconsciousness control*. If there will be indicated unreliable fulfillment of these operations the correspondent procedures can be returned to *AGI-Consciousness control* to be learned more for reliable performance.

As the *Emotion control* the author now can see some reason to implement "fear" of AGI-Agent: if the sensors data demonstrate the sharp increasing of information, and *AGI-Robot* estimates the environment as potentially dangerous, - *AGI-Unconsciousness* can indicate it and assign for *AGI-Robot* the *mode* of emotion "Fear". In this *emotion mode* "fear" some functions can be deactivated, for example, the *function of imagination and planning* and *objective function* can be turned off; limit of power consumption and the *clock frequency* of processor can be increased. In some regular situation *AGI-Unconsciousness* keeps the *emotion* "quiet". In the author's approach *emotions* are implemented as *modes* but not as *models*.

The functions of AGI-Collective type should be downloaded *as Agent-2 with the Agent-1 of Individual Intelligence. AGI-Collective type* does not control *AGI-Individual type* directly. Some mechanisms of interaction of *AGI-Individual type* and *AGI-Collective type* were mentioned above and in [6].

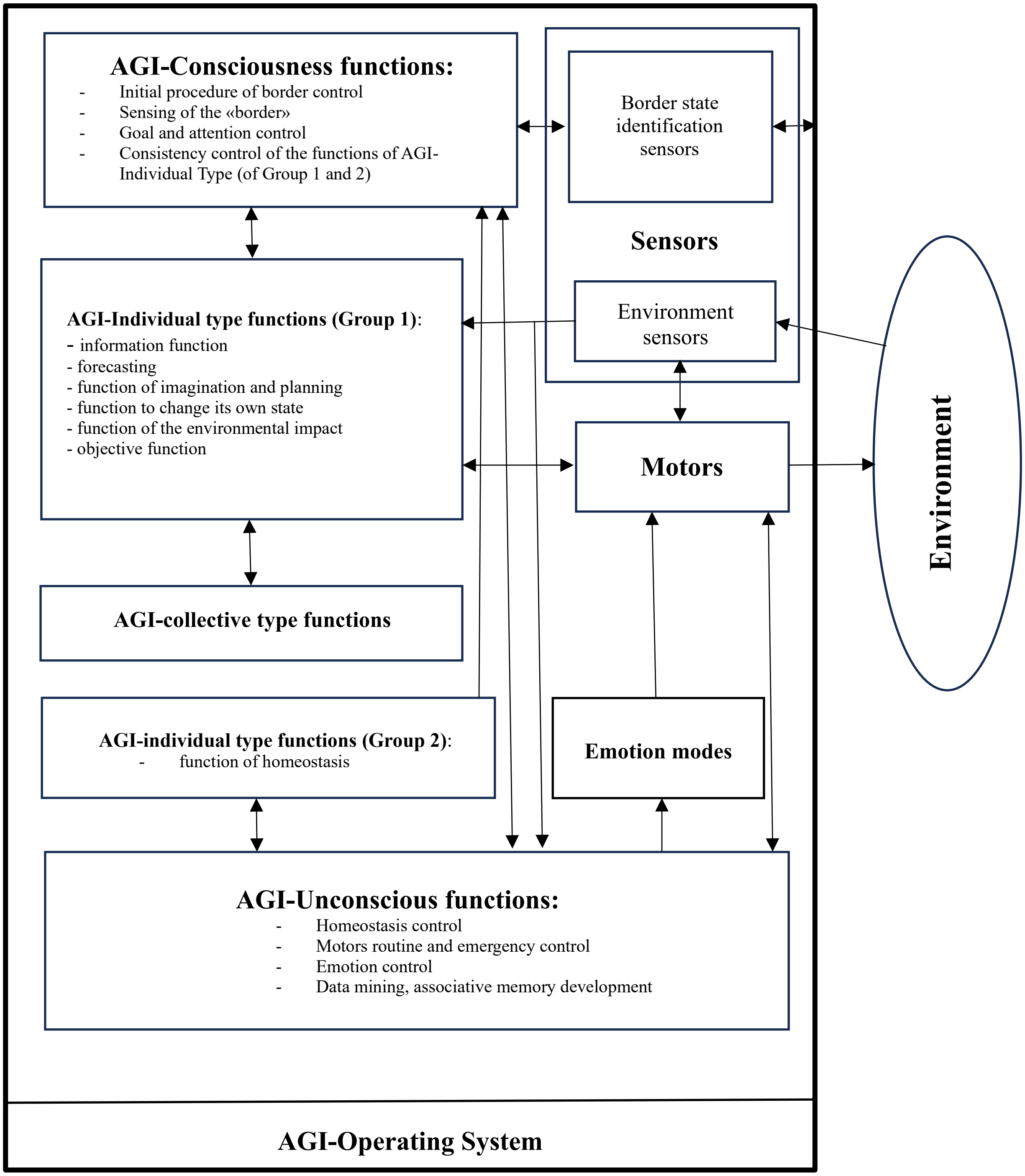

Fig. 2. Artificial General Intelligence Cognitive Architecture – functional view ("AGICA")

All processes will use *AGI-knowledges and AGI-skills*.

AGI-knowledges are symbolic.

AGI-skills are sub-symbolic. *AGI-skills* are divided between *Motors*, *Sensors*, and other modules (*AGI-Thought type 4*, others) Any *AGI-skill* have its symbolic *name*. But some of them can be started directly from *Sensors*. For example, *Sensors* identified the high temperature of *skin* on the arm. The correspondent signal will be transferred directly to the *Motors* and the arm will be withdrawn by the correspondent *skill*.

AGI-individual type processes should try to maximize *objective function* is set in the initialization.

The main elements of *AGI-individual type processe*s are *AGI-thoughts*. *AGI-thoughts* are the only source of complex actions of *AGI-robot* (the simple actions can be done by direct communications between *Sensors* and *Motors*). *AGI-thoughts* are symbolic and expressed in formal language.

Every *function* will have its own *working buffer* – the combined *function's working buffers* can be referred as *working memory*.

Procedural Long-term Memory in *cognitive architecture* AGICA is represented by the "*Set of Initial AGI-Thoughts*" and *Subsymbolic Data of ANN-parameters* that represent *skills* – this set will be downloaded in the implementation phase. The design of the *set of initial thoughts* will be based on the preliminary designed *AGI-Thoughts* and learning of some pilot examples of AGI-Robots.

For example, one of *initial thought type 1* ("scenario type") can be – "Border Exploration": by *manipulator* to touch the *border* of the *body* from initial point ("O-coordinate point") and prolong the *border* till its 3D-model will be closed.

Another example of *initial thought type 1* – "Space Exploration" – when AGI-Robot finished "Border Exploration" it can start "Space Exploration" – to try move forward, back, left, right and store the *spatial data* of correspondent distances (laser measured) as the table of *global coordinates* (when the accuracy is about centimeters), directions and ranges. This *thought* after multiple implementations (through *thought type 6*) will deliver the *map with obstacles*.

Any other types of *thought* also will have some initial *templates*.

AGI-Robot will generate *thoughts* by 2 mechanisms:

– Assign parameters to the *initial thoughts (modification of thoughts)*,
– Generation new *thoughts structures*.

The main difference with *Standard Model of Mind* that the author proposed in [6] is the concept "*Border"* as the base for the definition of *AGI-Individual type* and the need to indicate: the state of the border – "integrated" or "broken", - and mutual relations of the *Border* and the *objects* of the Environment – "touch", "close", "penetrated". In the nearest future this *border* in AGI-Robots will be represented by "Active Artificial Skin" – as the *surface distributed sensors* and correspondent control system.

The author represented here the functional view of *AGI cognitive architecture* – AGICA - that gives the possibility of further specifications of *functions*.

Concept design.

In the concept design the initial step is "Coordinate System Establishing" – the correspondent *thought type 1* is also will be in the *set of initial thoughts*. For the systems of kinds transport robots, the natural way to start from the *concept* and *implementation* of "Coordinate System".

Coordinate system.

Coordinate system must be in any *transport robot* – the sensors data can be used for navigation only with the correspondence with *internal coordinate system*. But this is not the only that *coordinate system* can represent to us.

What is the system coordinate of human being?

Between the eyebrows is the "ajna cakra" [43]. This chakra is the origin of the coordinate system of human being. This hypothesis author made when the first of my children was born. With the two next the author had proved this idea. From the moment of the child's birth, he has only one point that belong to him "in sure" – this is the point between eyebrows. Maybe this point is not on the skin, maybe it is in some depth – the author does not know this – this issue should be resolved by physiologists, - but in sure the child starts to explore himself and little by little spreads his knowledge about his body from just this point. A child touches himself by the fingers and goes and goes further. The process of spreading knowledge of the limits of child's body is very similar as the procedure of analytic continuation in calculus.

Initially this is quite dangerous, – the eyes are near this point, the child can wound himself! He initially does not know that the eyes are him! To prevent it we before swaddled baby's hands, and in nowadays parents use mittens for this purpose.

Finally, when a child is 2-2,5 month of age he puts his big toe in his mouth and finds nothing further than the heel… That is all. The child has the *border* and coordinate system. From this moment the child became "intellectual being"! He tied all. He knows where he is. It delivers to the brain and not only to the brain but also for autonomic nervous system the understanding of "self". It can be formed only with the understanding of the border.

The *coordinate system* of *AGI-Robot* can be formed in analogy with the living nature by assigning of original of coordinate system to some unchanged point of robot construction. Of course, AGI-Robot should have the possibility to calculate reference distances in this *coordinate system*.

The cognitive cycle in AGICA cognitive architecture includes the following phases:

– *Context identification* – hypothesis generation (on the base of the *thoughts of type 2-5*).

– *Context hypothesis testing* – the *context* proved (the process regularly repeats).

– *Goal analysis* - *subgoal generation* – *scenario generation* (thought type 1) – analysis of the results of *thoughts of types 2-5* – iterative improvements of *scenario* with the estimations of «*certainty/doubt*» balance.

– *"Certainty/doubt" signals generation and balance them* – *attention* control mechanism (detailed analysis of "objects-relations-in time") - "if and when" the threshold will be overcome *by certainty signal* – the thought type 6 activates the *action* by *Motors*; "if and when" the threshold will be overcome *by doubt signal* – the *scenario* will be rejected;

– The results of *action* should be analyzed – the *thoughts of types 2-5* – the *episode* should be saved in *episodic memory* (the format of *episode* is the same as format of the thought type 1) for the further analysis in *sleeping mode*; and the *cognitive cycle* is renewing from *context identification*.

The *cognitive cycle* starts from *context* identification. Some contexts are innate and is downloaded in the initialization process: "place of born"; "training center"; "maintenance and repair center"; "general working environment".

When the *hypothesis of context* made it should be proved by *testing*. The testing of the context will be repeated regularly. Some tests also can be innate and is downloaded in the initialization.

AGI-Agent further will decide when new context should be generated.

The described *cognitive cycle* is controlled by *AGI-Consciousness*.

The *actions* can be initiated by *AGI-Unconsciousness* on the base of "very short period" processes perceived by *Sensors* (the periods around 50ms in many cognitive architectures overviewed) – "hot water reaction of arm".

The details of algorithms and approaches of AGICA cognitive architecture will be published in upcoming articles.

Conclusion

In the article the author proposed the further development of the *Common Model of Cognition* on the base of Embodied Cognition of Cognitive Science and named "axiomatic approach" that had been proposed by Alexander in the beginning of this century [9,10], and its development made by the author in the articles mentioned here [6-8]. There was represented new architectural details for *Common Model of Cognition* that can be used in further modeling experiments.

The main proposal is to include in cognitive architecture *Body* or *Boder* as one of the main elements of *AGI-Agent*.

The next proposal of the author is to use *AGI-Thoughts* according to the author's definition [7] as the unified mechanism of "thinking" of *AGI-robot*. This also delivers the mechanism of *explanations* of the activities of AGI-robot: the set of *AGI-Thoughts* for some period can be simply transferred by *AGI-Robot* to be analyzed by the management entity.

In cognitive architectures developed before the "Consciousness" was modeled indirectly. In cognitive architecture AGICA the author proposed to model it directly according to the definition made [8].

According to author's point of view *Emotions* should not be the *models* in productive industrial systems but they should be the *modes of AGI-Robots*. The author cannot see the reason to implement in industrial systems all or large part of the *emotions* known in human beings, but the mode "AGI-Fear" can be useful. For collaborative AGI-robots the *models* of emotions can be in place.

This article is about the structure of *cognitive architecture* AGICA. Most functioned used to represent AGICA is known, and there are many models to implement them in literature. The exception is *AGI-Consciousness* [8]. It will be developed the first time.

This article is an abridged version of the original preprint published in OSF [45].

Acknowledgements

This work was supported by Qianxing Zehe Information Technology (Zhejiang) Co., Ltd., China.

References

1. Institute for Creative Technologies, Cognitive Architecture definition: https://cogarch.ict.usc.edu/
2. Kotseruba, I., Tsotsos, J.K. (2020). 40 years of cognitive architectures: core cognitive abilities and practical applications, Artificial Intelligence Review, 53:17–94, Published online: 28 July 2018; https://doi.org/10.1007/s10462-018-9646-y
3. Laird, J.E. (2019). The SOAR Cognitive Architecture, The MIT Press, 390 p.
4. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y. (2004). An Integrated Theory of the Mind, Psychological Review, Vol. 111, 4, 1036–1060, DOI: 10.1037/0033-295X.111.4.1036
5. Laird, J.E., Lebiere, C., Rosenbloom, P.S. (2017). A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics, AI Magazine, ISSN 0738-4602, VOL. 38 NO. 4, Winter.
6. Kornieiev, S. (2016). Operating System of Artificial Intelligence: the basic definitions. / "Artificial Intelligence" (Ukraine), ISSN 1561-5375 (online); ISSN 1561-5375 (online); ISSN 1561-5359, #4 (v.74), pp. 7-14 (there is the resume in English).
7. Kornieiev, S. (2017). The basic definitions of Artificial Intelligence Operating System: "AI-Thought" and AI-Emotions concepts. / "Artificial Intelligence" (Ukraine), ISSN 1561-5375 (online); ISSN 1561-5359, ##3-4 (vv.77-78), pp. 9-14 (in English).
8. Kornieiev, S. (2018). The basic definitions of Artificial Intelligence Operating System: «AI-Consciousness, AI-Time, "AI-KNOW/NOT_KNOW" and "AI-MEMORY" concepts / "Artificial Intelligence" (Ukraine), ISSN 1561-5375 (online); ISSN 1561-5375 (online); ISSN 1561-5359, ##2 (vv.80), pp. 9-14 (in English).
9. Aleksander, I., Dunmall, B. (2003). Axioms and Tests for the Presence of Minimal Consciousness in Agents. In Owen Holland (ed.), Machine Consciousness. Exeter: Imprint Academic.
10. Aleksander, I., (2005). The World in My Mind, My Mind in the World: key mechanisms of consciousness in people, animals and machines, Imprint Academic, NY, USA, ISBN 9781845400217, 200p.
11. Varela, F.J., Thompson, E., Rosch, E. (2016). The Embodied Mind, The MIT Press Cambridge, Massachusetts, London, England, 322p.
12. Samsonovich, A.V., (2010). Toward a Unified Catalog of Implemented Cognitive Architectures, Conference: Biologically Inspired Cognitive Architectures, 2010 - Proceedings of the First Annual Meeting of the BICA Society, Washington, DC, USA, November 13-14, 2010; DOI: 10.3233/978-1-60750-660-7-195; https://www.researchgate.net/publication/221313271_Toward_a_Unified_Catalog_of_Implemented_Cognitive_Architectures
13. Newell, A., (1990). Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts, 576p.
14. Lebiere, C., Anderson, J. R. (1993). A Connectionist Implementation of the ACT-R Production System, In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, 635-640.
15. Anderson, J. R., Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Erlbaum, ISBN 0 8058 2816, 490p.
16. Sun, R. (2016). Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion

Cognitive Architecture, Oxford, UK: Oxford University Press.
doi.org/10.1093/acprof:oso/9780199794553.001.0001

17. O'Reilly, R.C., Hazy, T.E., Herd, S.A. (2015). The Oxford Handbook of Cognitive Science, Edited by Chipman S. Print Publication Date: Oct 2017 Subject: Psychology, Cognitive Psychology,
DOI:10.1093/oxfordhb/9780199842193.013.8;
https://ccnlab.org/papers/OReillyHazyHerd16.pdf

18. Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, C., Rasmussen, D. (2012). A Large-Scale Model of the Functioning Brain, Science. November, pp.1202-1205;
DOI: 10.1126/science.1225266;
https://www.researchgate.net/publication/233806263_A_Large-Scale_Model_of_the_Functioning_Brain

19. Gobet, F., Lane, P. C. (2020). Constructing a Standard Model: Lessons from CHREST. Common Model of Cognition Bulletin, 1(2), 338–343; https://ojs.library.carleton.ca/index.php/cmcb/article/view/2631

20. Kieras, D. E. & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction, 12, pp.391–438.

21. Forbus, K.D., Hinrichs, T. (2017). Analogy and Qualitative Representations in the Companion Cognitive Architecture. AI Magazine.

22. Lenat, D. (1995). CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM, 38 (11): pp.33–38.

23. Bach, J. (2003). The MicroPsi Agent Architecture, Proceedings of ICCM-5, Universitäts-Verlag Bamberg, ISBN 3-933463-15-7, pp.15-20;
https://www.researchgate.net/publication/2861286_The_MicroPsi_Agent_Architecture

24. Cox, M.T., Alavi, Z., Dannenhauer, D., Eyorokon, V., Munoz-Avila, H., Perliss, D. (2016). MIDCA: A Metacognitive, Integrated Dual-Cycle Architecture for Self-Regulated Autonomy, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 3712-3718;
https://www.researchgate.net/publication/310491466_MIDCA_A_Metacognitive_Integrated_Dual-Cycle_Architecture_for_Self-Regulated_Autonomy

25. Fitzgerald, C. (2013). Developing Baxter, 2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA): A New Industrial Robot with Common Sense for U.S. Manufacturing. doi:10.1109/TePRA.2013.6556344

26. Cassimatis, N.L. (2005). Integrating Cognitive Models Based on Different Computational Methods, Paper presented at the Twenty-Seventh Annual Conference of the Cognitive Science Society.

27. Cassimatis N. (2006). A Cognitive Substrate for achieving Human-Level Intelligence, AI Magazine, vol. 27 (2), 45-56.

28. Franklin, S., Patterson, F. (2006). The LIDA Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent, IDPT-2006 Proceedings (Integrated Design and Process Technology): Society for Design and Process Science, San Diego; https://www.researchgate.net/publication/210304626_The_LIDA_architecture_Adding_new_modes_of_learning_to_an_intelligent_autonomous_software_agent

29. Baars, B.J. (1988). A cognitive theory of consciousness, Cambridge: Cambridge University Press, 428p.

30. Baars, Bernard (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. Journal of Consciousness Studies, 4, 292-309.

31. Langley, P., Choi, D. (2006). A unified cognitive architecture for physical agents. Proceedings of AAAI06, Association for the Advancement of Artificial Intelligence, 1469-1474.
https://cdn.aaai.org/AAAI/2006/AAAI06-231.pdf

32. Jilk, D.J., Lebiere, C., O'Reilly, R.C., Anderson, J.R. (2008). SAL: an explicitly pluralistic cognitive architecture, Journal of Experimental & Theoretical Artificial Intelligence, 20, 197-218.

33. Schaat, S. (2016). SiMA-C: A Foundational Mental Architecture, Procedia Computer Science, v. 88, 371–378.

34. Rosenbloom, P.S., Demski, A., Ustun, V. (2016). The Sigma cognitive architecture and system: towards functionally elegant grand unification, Journal of AGI, 7 (1), 1-103.

35. Koller, D., Friedman, N. (2009). Probabilistic Graphical Models: Principles and Techniques, Cambridge, MA: MIT Press.

36. DARPA ("Defence Advanced Research Project Agency", USA;
https://www.darpa.mil/attachments/AIFull.pdf

37. Sun, R. (2004). Desiderata for cognitive architectures, Philosophical Psychology, v.17, (3), 341–373.

38. Roosterman, D., Goerge, T., Schneider, S.W., Bunnett, N.W., Steinhoff, M. (2006). Neuronal Control of Skin Function: The Skin as a Neuroimmunoendocrine Organ, Physiological Review 86, American Physiological Society, 1309–1379.

39. UC Berkeley - pressure-sensitive electronic material: https://news.berkeley.edu/2010/09/12/eskin/

40. Standford University - stretchable, bendable circuits on artificial skin:
https://news.stanford.edu/2021/07/01/engineers-develop-practical-way-make-artificial-skin/

41. Santos, A., Fortunato, E., Martins, R., Águas, H., Igreja, R. (2020). Transduction Mechanisms, Micro-Structuring Techniques, and Applications of Electronic Skin Pressure Sensors: A Review of Recent Advances, Sensors, 20 (16), 4407; doi:10.3390/s20164407; https://www.mdpi.com/1424-8220/20/16/4407

42. Golam, Sarower Rayhan, and others (2023). Conductive Textiles for Signal Sensing and Technical Applications, Signals, 4, 1–39.
https://doi.org/10.3390/signals4010001

43. Kak, S.C., Rouge, B. (1997). On the Science of Consciousness in Ancient India, LA 70803-5901, Indian Journal of History of Science, vol. 32, 105-120.

44. Kornieiev, S. (2017). Operating System to base AI-Applications: the overview and general technical requirements, TAAPSD'2017, 04-08.12.2017, UDK 004.93. https://www.researchgate.net/publication/318877523_Operating_systems_to_base_AI-applications_the_overview_and_general_technical_requirements_in_English

45. Kornieiev, S. V. (2023, August 17). AGI-agent cognitive architecture AGICA - axiomatic approach, preprint. https://doi.org/10.31219/osf.io/5zu28

The article has been sent to the editors 15.10.23.
After processing 25.10.23.
Submitted for printing 30.11.23.