СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ, СТАТИСТИКА

УДК 004.89 Научная статья
DOI: 10.35330/1991-6639-2025-27-1-152-170
EDN: ZXPNUH

Основные принципы нейрокогнитивного моделирования сознания агента универсального искусственного интеллекта

З. В. Нагоев

Институт информатики и проблем регионального управления –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, г. Нальчик, ул. И. Арманд, 37-а

Аннотация. Целью работы является создание теоретических оснований имитационного моделирования агентов универсального искусственного интеллекта на основе метафоры проектирования мультиагентных нейрокогнитивных архитектур. Задача исследования состоит в разработке основных принципов имитационного моделирования сознания интеллектуальных агентов под управлением нейрокогнитивных архитектур. Дано формальное определение агента универсального искусственного интеллекта. Предложена гипотеза о структурно-функциональной организации сознания универсального искусственного интеллекта на основе метафоры проектирования многокомпонентной мультиагентной нейрокогнитивной архитектуры. Разработаны некоторые принципы имитационного моделирования сознания агентов универсального искусственного интеллекта на основе контекстно-детерминированного развития управляющей нейрокогнитивной архитектуры в коммуникативной социальной реальной среде.

Ключевые слова: универсальный искусственный интеллект, имитационное моделирование сознания, нейрокогнитивные архитектуры, мультиагентные системы

Поступила 30.01.2025, одобрена после рецензирования 10.02.2025, принята к публикации 11.02.2025

Для цитирования. Нагоев З. В. Основные принципы нейрокогнитивного моделирования сознания агента универсального искусственного интеллекта // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 1. С. 152–170. DOI: 10.35330/1991-6639-2025-27-1-152-170

MSC: 68T42 Original article

Basic principles of neurocognitive modeling of consciousness of an agent of universal artificial intelligence

Z.V. Nagoev

Institute of Computer Science and Problems of Regional Management –
branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
360000, Russia, Nalchik, 37-a I. Armand street

Abstract. The aim of the work is to create theoretical foundations for simulation modeling of agents of universal artificial intelligence based on the metaphor of designing multi-agent neurocognitive architectures. The task of the study is to develop basic principles of simulation modeling of consciousness of intelligent agents under the control of neurocognitive architectures. A formal definition of an agent of universal artificial intelligence is given. A hypothesis is proposed about the structural and functional organization of consciousness of universal artificial intelligence based on the

© Нагоев З. В., 2025

metaphor of designing a multi-component multi-agent neurocognitive architecture. Some principles of simulation modeling of the consciousness of agents of universal artificial intelligence are developed based on the context-deterministic development of the control neurocognitive architecture in a communicative social real environment.

***Keywords**:* universal artificial intelligence, simulation modeling of consciousness, neurocognitive architectures, multi-agent systems

Submitted 30.01.2025, approved after reviewing 10.02.2025, accepted for publication 11.02.2025

For citation. Nagoev Z.V. Basic principles of neurocognitive modeling of consciousness of an agent of universal artificial intelligence. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2025. Vol. 27. No. 1. Pp. 152–170. DOI: 10.35330/1991-6639-2025-27-1-152-170

Введение

Имитационное моделирование сознания, концептуально и исторически являющееся одной из ключевых теоретических проблем искусственного интеллекта как области исследования возможностей функционального моделирования разума [1, 2], для случая *универсального интеллектуального агента* является проблемой в первую очередь прикладной. Как показано в настоящей работе, необходимость моделирования сознания интеллектуального агента определяется, прежде всего, задачей *управления его поведением* в *условиях реальной среды*.

Эти условия, среди прочего, характеризуются *эпизодичностью и частичностью наблюдений, стохастичностью событий, динамикой, физической корректностью, агентностью взаимодействий, неопределенностью, априорной неструктурированностью и многомодальностью данных*, существенно затрудняющими решение задач интеллектуального управления.

В соответствии с гипотезой, на базе которой был разработан подход к созданию систем искусственного интеллекта с помощью метафоры проектирования *мультиагентных нейрокогнитивных архитектур*, нашедший отражение в многолетнем цикле научных работ [3, 4, 5, 6, 7], интеллектуальные системы управления поведением природных высших биологических организмов («естественный» интеллект) стали основным средством их приспособления к таким условиям реальной среды, а главным инструментом обеспечения эффективности адаптации интеллекта стала нейропластичность головного мозга, реализующая динамическую аппроксимацию законов управления к таким условиям.

В частности, в вышеприведенных работах показано, что самоорганизация *агентов-нейронов* (*агнейронов*) в составе функциональных узлов (*когнитонов*) управляющих нейрокогнитивных архитектур (*интеллектонов*) интеллектуальных агентов, детерминированная процессами обмена энергии и информации между агнейронами с помощью динамического формирования аксо-дендрональных связей и отправки нейромедиаторов, направленная на максимизацию локальных целевых функций этих агнейронов, обеспечивает синтез так называемых *специальных функциональных систем* (по П. К. Анохину) *онтологизации, идентификации и решения проблем универсального спектра* в системе «интеллектуальный агент – реальная среда».

Под функциональной системой в данном случае понимается множество агнейронов из различных когнитонов, связанных т.н. контрактами на обмен энергии и информации, скоординированный диалог на основе передачи сообщений между которыми обеспечивает реализацию т.н. *нейрокогнитивной пьесы* (*нк-пьеса*) – *нейрокогнитивного мультиагентного алгоритма*, реализующего функционал онтологизации (первичного определения) некоторой проблемы, либо ее идентификации (повторного распознавания), либо синтеза ее решения.

Выдвинутая ранее гипотеза о наличии в головном мозге высших организмов *метафункциональных систем*, способных в зависимости от контекста генерировать специальные функциональные системы онтологизации, идентификации и решения различных проблем в системе «интеллектуальный агент – среда», была реализована в метафоре проектирования управляющих нейрокогнитивных архитектур в виде концепции так называемых *нейронных фабрик* (*нейрофабрики*), реализующих *функцию нейрогенеза* – ситуативно обусловленного динамического порождения новых агнейронов в различных когнитонах таких нейрокогнитивных архитектур [5].

Вкупе с введенными в [4, 8] элементами формального описания контрактных взаимодействий между агнейронами в составе интеллектона (*н-функции*) концепция динамического синтеза специальных функциональных систем в составе управляющей нейрокогнитивной архитектуры интеллектуального агента формирует теоретический базис для обоснования способности агентов универсального интеллекта автономно онтологизировать, идентифицировать и решать проблемы универсального спектра, возникающие в системе «интеллектуальный агент – реальная среда».

Если упрощенно под *проблемой* понимать такой отрезок траектории интеллектуального агента в реальной среде, который оканчивается состоянием, в котором *целевая функция агента* под управлением нейрокогнитивной архитектуры претерпевает значительные изменения, то можно утверждать, что метафункциональные системы, реализуемые этой нейрокогнитивной архитектурой, динамически, по требованию, обусловленному ситуацией отсутствия в нейрокогнитивной архитектуре нейрокогнитивного «обработчика» этой проблемы, порождая в ней дополнительные специальные функциональные системы онтологизации, идентификации и решения, обеспечивают базу для единообразного решения любых возможных в данной системе «интеллектуальный агент – реальная среда» проблем на основе скоординированного выполнения нейрокогнитивных пьес, реализуемых этими метафункциональными и функциональными системами, т.е. на основе синтеза поведения интеллектуального агента в реальной среде.

В соответствии с ранее выдвинутой гипотезой [9] сознание рассматривается как функциональная надстройка над управляющей нейрокогнитивной архитектурой нижнего уровня (*подсознание интеллектуального агента*), сама представляющая собой нейрокогнитивную архитектуру, объектом управления которой является подсознание интеллектуального агента.

Так как интеллектуальный агент синтезирует свое поведение в реальной среде, ему, как правило, приходится одновременно онтологизировать, идентифицировать и решать несколько проблем, события в составе которых растянуты во времени и происходят в различных частях пространства. Поэтому одной из основных функций сознания является, в соответствии с нашей гипотезой, управление распределением монопольных ресурсов организма (или его искусственных аналогов – программ, роботов), таких как руки, движители, между алгоритмами управления поведением, синтезируемыми различными метафункциональными и специальными функциональными системами для решения различных проблем.

Актуальность работы определяется необходимостью создания агентов универсального искусственного интеллекта, способных автономно онтологизировать, идентифицировать и решать проблемы универсального спектра в системе «агент – среда».

Целью исследования является создание теоретических оснований имитационного моделирования *агентов универсального искусственного интеллекта* на *основе метафоры проектирования мультиагентных нейрокогнитивных архитектур*.

Задача исследования состоит в разработке *основных принципов имитационного моделирования сознания* интеллектуальных агентов под управлением нейрокогнитивных архитектур.

Так как сам термин «универсальный искусственный интеллект» зачастую используется в качестве неформального, несущего, скорее, интуитивно понятное значение компьютерной программы искусственного интеллекта, способной к автономному поиску решений проблем широкого спектра, изначально не заложенных в алгоритмах этой программы, необходимо ввести формальное определение агента универсального искусственного интеллекта (УИИ).

1. Агент универсального искусственного интеллекта

Рассмотрим некоторую вычислительную систему S_i, погруженную во внешнюю по отношению к ней реальную среду W^{-S_i}. *Вычислительный цикл* этой системы под номером h определим как набор *алгоритмов*:

$$a^{ih} = \left\{a_{1\tau_{c_1}}^{ih\tau_{c_2}}, a_{2\tau_{c_3}}^{ih\tau_{c_4}}, \dots, a_{n\tau_{c_{m-1}}}^{ih\tau_{c_m}}\right\},$$

которые выполняются на шагах дискретного времени $\Delta\tau_{c_y}^{c_{y+1}} = \tau_{c_{y+1}} - \tau_{c_y}$. Эти алгоритмы реализуют синтез набора отрезков f^{ih} выходной *эффекторной траектории* f^i системы S_i, которые соответствуют некоторому закону управления: $f^{ih} = u^i\left(r^{ih}\right)$. Этот закон формируется устройством управления U^i на основе набора входных отрезков r^{ih} *сенсорной траектории* r^i, которая определяет движение системы S_i в *сенсорном пространстве* $\mathbf{R}^{\boldsymbol{i}}$, которое, в свою очередь, определяется как декартово произведение наборов значений всех сенсоров системы S_i, регистрирующих некоторые параметры, описывающие состояния как самой этой системы, так и внешней среды W^{-S_i}.

Содержательный смысл выполнения цикла вычислений состоит в том, что устройство управления U^i строит некоторый *закон управления*:

$$u^i\left(r^i\right)\colon \mathbf{R}^{\boldsymbol{i}} \to \mathbf{F}^{\boldsymbol{i}}, \mathbf{F}^{\boldsymbol{i}} = \left\{\mathbf{F}_{\tau_1}^{\boldsymbol{i}} \times \mathbf{F}_{\tau_2}^{\boldsymbol{i}}, \dots, \times \mathbf{F}_{\tau_{c^{\mathbf{max}}}}^{\boldsymbol{i}}\right\},$$

$$\mathbf{F}_{\tau_c}^{\boldsymbol{i}} = \left\{\mathbf{f}_{1\tau_c}^{i\mathbf{int}} \times \mathbf{f}_{2\tau_c}^{i\mathbf{int}} \times \dots \times \mathbf{f}_{q_{\mathbf{int}}^{\mathbf{max}}\tau_c}^{i\mathbf{int}} \times \mathbf{f}_{1\tau_c}^{i\mathbf{ext}} \times \mathbf{f}_{2\tau_c}^{i\mathbf{ext}} \dots \times \mathbf{f}_{q_{\mathbf{ext}}^{\mathbf{max}}\tau_c}^{i\mathbf{ext}}\right\},$$

$$u\left(r_{\tau_{c-y}}^{\tau_c}\right) = f_{\tau_c}^i\left(f_{1\tau_c}^{i\mathbf{int}}, f_{2\tau_c}^{i\mathbf{int}}, \dots, f_{q_{\mathbf{int}}^{\mathbf{max}}\tau_c}^{i\mathbf{int}}, f_{1\tau_c}^{i\mathbf{ext}}, f_{2\tau_c}^{i\mathbf{ext}}, \dots, f_{q_{\mathbf{ext}}^{\mathbf{max}}\tau_c}^{i\mathbf{ext}}\right), q_{\mathbf{int}}^{\mathbf{max}} + q_{\mathbf{ext}}^{\mathbf{max}} = q^{\mathbf{max}},$$

$$\mathbf{f}_{q\tau_c}^{i\mathbf{int}} = \left\{f_{q\tau_c}^{i\mathbf{int}} \mid \forall\, u^i\left(r_{\tau_{c-y}}^{i\tau_c}\right)\right\}, \mathbf{f}_{q\tau_c}^{i\mathbf{ext}} = \left\{f_{q\tau_c}^{i\mathbf{ext}} \mid \forall\, u^i\left(r_{\tau_{c-y}}^{i\tau_c}\right)\right\}.$$

Здесь $\mathbf{F}^{\boldsymbol{i}}$ – т.н. *эффекторное пространство*, $f_{q\tau_c}^{i\mathbf{int}}$ и $f_{q\tau_c}^{i\mathbf{ext}}$ – состояния *интероэффекторов* $f_q^{i\mathbf{int}} \in F^i$ системы S_i, с помощью которых она воздействует на свои составные части, и ее *экстероэффекторов* $f_q^{i\mathbf{ext}} \in F^i$, с помощью которых она влияет на внешнюю среду W^{-S_i}, соответственно.

Набор точек:

$$f_{\tau_c}^i\left(f_{1\tau_c}^{i\mathbf{int}}, f_{2\tau_c}^{i\mathbf{int}}, \dots, f_{q_{\mathbf{int}}^{\mathbf{max}}\tau_c}^{i\mathbf{int}}, f_{1\tau_c}^{i\mathbf{ext}}, f_{2\tau_c}^{i\mathbf{ext}}, \dots, f_{q_{\mathbf{ext}}^{\mathbf{max}}\tau_c}^{i\mathbf{ext}}\right) \in \mathbf{F}_{\tau_c}^{\boldsymbol{i}}$$

задает *эффекторную траекторию* системы S_i:

$$f^i = f_{\tau_1}^{i\tau_{c^{\mathbf{max}}}} = \left(f_{\tau_1}^i, f_{\tau_2}^i, \dots, f_{\tau_{c^{\mathbf{max}}}}^i\right) = u^i\left(r_{\tau_1}^{i\tau_{c^{\mathbf{max}}}}\right).$$

Пространство

$$\Theta^{i\urcorner} = \mathbf{R}^i \times \mathbf{F}^i$$

будем называть *пространством поведения* системы S_i в среде W^{-S_i}.

Систему S_i, для которой определена функция $f^{ih} = u^i\left(r^{ih}\right)$, назовем *агентом* (*агентным вычислителем*) $\aleph_i$ *в реальной сред*е W^{-S_i}.

Зададим *класс агентных вычислителей* **A** (*а-вычислители*) как множество всех агентов $\aleph_i$.

Будем считать, что в силу действия в реальной среде $W^{-\aleph_i}$ закона сохранения энергии цикл вычислений a^{ih} может быть реализован агентом $\aleph_i$ только в случае выполнения этим агентом затрат энергии $e^{ih\mathbf{x}}$, рассеивающейся в среду $W^{-\aleph_i}$.

Примем также, что для нормальной работы агента $\aleph_i$ требуются и затраты энергии: $e^{ih\mathbf{L}} = \Delta\tau_{c_y}^{c_m}\Delta e^{i\mathbf{L}}$, где $\Delta e^{i\mathbf{L}}$ – потери энергии на одном интервале времени.

Агент $\aleph_i$, при движении по сенсорной траектории r^{ih} переходит в состояния, в которых он (взаимодействуя со средой $W^{-\aleph_i}$) или теряет некоторое количество энергии $e^{ih\mathbf{w}}$, или приобретает некоторое вознаграждения $e^{ih\mathbf{r}}$. Пусть этот агент запасает энергию e^i в своем внутреннем резервуаре $E^i \in \aleph_i$. Пусть также емкость этого резервуара конечна и не превышает значения $c\left(E^i\right)$. Динамика значения количества энергии агента $\aleph_i$ во время реализации цикла вычислений a^{ih} описывается *функцией разметки состояний*:

$$e^{ih} = e^{i(h-1)} - e^{ih\mathbf{L}} - e^{ih\mathbf{x}} - e^{ih\mathbf{w}} + e^{ih\mathbf{r}},$$

а количество энергии, хранящейся в его резервуаре E^i, – *функцией энергии*:

$$e^i = e_1^i + \sum\nolimits_{h=1}^{h^{\max}} e^{ih}, f^i = \bigcup\nolimits_{h=1}^{h^{\max}} f^{ih}, f^{ih} = \left(f_{\tau_{c_1}}^{ih\tau_{c_2}}, f_{\tau_{c_3}}^{ih\tau_{c_4}}, \dots, f_{\tau_{c^{\max}-1}}^{ih\tau_{c^{\max}}}\right), e^i \le c\left(E^i\right),$$

где e_1^i – энергия в резервуаре E^i агента $\aleph_i$ в начальный момент времени τ_1.

Будем считать, что *время жизни* $\tau_{c^{\max}}$ агента $\aleph_i$ есть функция от *материальной траектории* w^i *системы* «*агент* $\aleph_i$*– среда* $W^{-\aleph_i}$», которая есть функция от эффекторной и сенсорной траекторий: $\tau_{c^{\max}} = \tau^i\left(w^i\left(u^i\left(r^i\right)\right)\right)$.

Точки материальной траектории w^i агента $\aleph_i$, для которых верно

$$\exists w_{\tau_c}^{i\mathbf{pos}} \in w^i, e^i\left(w_{\tau_c}^{i\mathbf{pos}}\right) > e^i\left(w_{\tau_{c-1}}^i\right), w_{\tau_c}^{i\mathbf{pos}} \in W^{i\mathbf{pos}},$$

назовем *позитивными* точками, условие

$$\exists w_{\tau_c}^{i\mathbf{neg}} \in w^i, e^i\left(w_{\tau_c}^{i\mathbf{neg}}\right) < e^i\left(w_{\tau_{c-1}}^i\right), w_{\tau_c}^{i\mathbf{neg}} \in W^{i\mathbf{neg}}$$

– *негативными* точками, условие

$$\exists w_{\tau_c}^{i\mathbf{iter}} \in w^i e^i\left(w_{\tau_c}^i\right) = 0, w_{\tau_c}^{i\mathbf{neg}} \in W^{i\mathbf{iter}} \in W^{i\mathbf{neg}}$$

– *терминальными* точками. Вместе все точки $w_{\tau_c}^{i\mathbf{pos}}$, $w_{\tau_c}^{i\mathbf{neg}}$ и $w_{\tau_c}^{i\mathbf{iter}}$ назовем *экзистенциальными точками*, а множество

$$W^{i\mathbf{ex}} = W^{i\mathbf{pos}} \cup W^{i\mathbf{neg}} \cup W^{i\mathbf{iter}}$$

– *множеством экзистенциальных точек*.

Устройство управления $U^i \in \aleph_i$, расходуя энергию, содержащуюся в резервуаре $E^i \in \aleph_i$, формирует эффекторную траекторию, устремляющую к максимуму *целевую функцию времени жизни* (*экзистенции*) агента $\aleph_i$:

$$\tau^i\left(w^i\left(f^i\right)\right) \xrightarrow[f^i=u^i(r^i)\in\mathbf{F}^i, e^i\le c(E^i), e^i(w_{\tau_c}^i)>0\Rightarrow\exists\aleph_i, e^i(w_{\tau_c}^i)=0\Rightarrow\nexists\aleph_i]{} \max.$$

Будем считать, что агенты $\aleph_i$, реализующие вычислительный цикл a^{ih} с целью решения этой задачи, относятся к классу *экзистенциальных вычислителей* **Э** (*э-вычислители*), являющемуся подклассом класса агентных вычислителей **А**.

Примем, что *функция состояний* агента $\aleph_i$

$$s^i(r^i) = s^{i\tau_c\mathbf{max}}_{\tau_1} = \left(s^i_{1\tau_{x_1}}, s^i_{2\tau_{x_2}}, \dots, s^i_{j^{\mathbf{max}}\tau_{x^j\mathbf{max}}}\right)\colon \mathbf{R^i} \to \mathbf{S^i}, s^i_{j\tau_{x_j}} = s^i\left(r^{ij\tau_{x_k}}_{\tau_{x_{k-y}}}\right)$$

размечает сенсорное пространство $\mathbf{R^i}$, ставя в соответствие отрезкам сенсорных траекторий $r^{ij\tau_{x_k}}_{\tau_{x_{k-y}}}$ (для которых введен верхний индекс номера состояния j), некоторым состояниям $s^i_{j\tau_{x_j}}$. Выполняя такую разметку, функция состояний задает *пространство состояний* $\mathbf{S^i}$ системы «интеллектуальный агент – реальная среда» на базе реализации *когнитивного процесса*. Последовательности состояний $s^i_{j\tau_{x_j}}$ задают *траектории состояний* $s^{i\tau_{x_2}}_{\tau_{x_1}}$ агента $\aleph_i$ в пространстве $\mathbf{S^i}$.

Функция разметки состояний агента $\aleph_i$

$$p^{i\tau_c\mathbf{max}}_{\tau_1} = \left(p^i_{1\tau_{x_1}}, p^i_{2\tau_{x_2}}, \dots, p^i_{d^{\mathbf{max}}\tau_{x^d\mathbf{max}}}\right) = p^i(s^i)\colon \mathbf{S^i} \to \mathbf{P^i},$$

$$p^i_{d\tau_{x_d}} = p^i\left(s^{i\tau_{x_k}}_{\tau_{x_{k-y}}}\right), s^i_{j\tau_{x_k}} = s^i\left(r^{ij\tau_{x_l}}_{\tau_{x_{l-m}}}\right), r^{ij}_{\tau_{x_l}} = r^i\left(w^i_{\tau_{x_l}}\right), w^i_{\tau_{x_l}} \in W^{i\mathbf{exs}}$$

определяет *проблемы* $p^i_{d\tau_{x_d}}$, которые принадлежат *пространству проблем* $\mathbf{P^i}$ агента $\aleph_i$. Последовательность проблем $p^i_{d\tau_{x_d}}$ задает некоторые части *траектории проблем* (*потока проблем*) $p^{i\tau_{x_2}}_{\tau_{x_1}}$ агента в пространстве проблем $\mathbf{P^i}$.

Выделим подкласс *когнитивных вычислителей* **К** (*к-вычислители*) класса экзистенциальных вычислителей **Э**, включающий в себя такие вычислители, закон управления которых может быть представлен в виде функции φ^i вида:

$$f^{ih} = u^i(r^{ih}) = \varphi^i\left(p^i\left(s^i(r^{ih})\right)\right).$$

В *эпизодической, динамической, неопределенной, частично наблюдаемой среде* $W^{-\aleph_i}$ агент $\aleph_i$ не может наблюдать все состояния s^{ih}, формирующие проблему $p^{ih}_{d^h\tau_x}$. Будем считать, что *функция ретропроактивного моделирования*

$$m^i(s^{ih}) = \{p^{ih} \cup p^{ih\Xi}_k \mid \forall\, p^{ih\Xi}_k \in P^{ih\Xi}\}$$

возвращает множество отрезков траекторий частично наблюдаемой проблемы p^{ih} и возможных целей $p^{ih\Xi}_k$, связанных с переходом к состояниям агента $\aleph_i$, в которых эта проблема будет решена.

В классе к-вычислителей зададим подкласс *ретропроактивных вычислителей* **Т** (*т-вычислители*), таких, что их закон управления может быть представлен в виде:

$$f^{ih} = u^i(r^{ih}) = \varphi^i\left(m^i\left(p^i\left(s^i(r^{ih})\right)\right)\right).$$

Будем считать, что *функция выбора цели* агента $\aleph_i$

$$p^{ih\Xi*}_k = g^i(m^i) = \max\{e^i(p^{ih\Xi}_k), \forall p^{ih\Xi}_k \in P^{ih\Xi}\}$$

возвращает цель $p_k^{ih\Xi*}$, являющуюся субоптимальной по ожидаемому в будущем значению функции разметки, а *функция синтеза решений* агента $\aleph_i$

$$a^i\left(p^{ih}, p_k^{ih\Xi*}\right) = \left\{ \begin{array}{c} f_l^{ihk\Xi} \mid \\ p^{ih} = p^i\left(s^i\left(r^{ih}\right)\right) \wedge f_l^{ihk\Xi} = \left(u^i\left(r^{ih}\right)\right) \Rightarrow \\ \Rightarrow p_k^{ih\Xi*} = p^i\left(s^i\left(r^i\left(w^i\left(f_l^{ihk\Xi}\right)\right)\right)\right) \end{array} \right\}$$

возвращает *набор решений* (*алгоритмов*) – частей эффекторной траектории $f_l^{ihk\Xi}$, в результате реализации которых система «агент – среда» переходит из *проблемной ситуации* p^{ih} в *целевое состояние* $p_k^{ih\Xi*}$.

Примем, что *функция выбора решений* агента $\aleph_i$

$$f_l^{ihk\Xi*} = c^i\left(a^i\left(p^{ih}, p_k^{ih\Xi*}\right)\right) = \max\left\{e^i\left(p_k^{ih\Xi*}\right), \forall f_l^{ihk\Xi}\right\}$$

возвращает решение $f_l^{ihk\Xi*}$ проблемы p^{ih}, переводящее агента $\aleph_i$ в целевое состояние $p_k^{ih\Xi*}$, такое, что оценка $e^i\left(p_k^{ih\Xi*}\right)$ становится максимальной.

Внутри класса т-вычислителей выделим подкласс *автодетерминированных вычислителей* **Д** (*д-вычислители*), чей закон управления может быть представлен в виде:

$$f^{ih} = u^i\left(r^{ih}\right) = c^i\left(a^i\left(g^i\left(m^i\left(p^i\left(s^i\left(r^{ih}\right)\right)\right)\right)\right)\right).$$

Внутри класса **Д** выделим класс *самообучаемых вычислителей* **О** (*о-вычислители*), которые самостоятельно (с помощью устройства управления U^i, рассеивая энергию из резервуара E^i) генерируют закон управления на всем интервале своего времени жизни, используя *функция обучения*

$$u_{\tau_{d_k}}^{i\tau_{d_{k+1}}}\left(r^i\right) = l^i\left(u_{\tau_1}^{i\tau_{d_{k-1}}}\left(r^i\right), r^i, f^i, e^i\right), U^i \in P_i \in \aleph_i, E^i \in P_i \in \aleph_i,$$

$$u^i\left(r^i\right) = \left\{ \begin{array}{c} u_{\tau_{d_k}}^{i\tau_{d_{k+1}}}\left(r^i\right) \mid k = \overline{[1, \ldots, k^{\max}]}, u_{\tau_{d_k}}^{i\tau_{d_k}}\left(r^i\right) = u_{\tau_{d_k}}^{i\tau_{d_k+c}}\left(r^i\right), \\ c = \overline{[1, \ldots, d_{k+1} - d_k]} \end{array} \right\},$$

задающую процесс изменения закона управления на шагах дискретного времени $\left[\tau_{d_k}; \tau_{d_{k+1}}\right]$, на каждом из которых он имеет вид $u_{\tau_{d_k}}^{i\tau_{d_{k+1}}}\left(r^i\right)$.

В силу *стохастичности* и *неопределенности* среды $W^{-\aleph_i}$ в потоке проблем $p_{\tau_1}^{i\tau_{c^{\max}}}$ на шаге времени τ_c только некоторые проблемы $p_{h\tau_x}^{\mathbf{iont}} \in \mathbf{P}_{\tau_c}^{\mathbf{iont}} \in \mathbf{P}^{\boldsymbol{i}}$ уже *онтологизированы*, т.е.:

$$\forall p_{h\tau_x}^{\mathbf{iont}} \in \mathbf{P}_{\tau_c}^{\mathbf{iont}}: \exists a_{k\tau_x}^{ih} \in \mathbf{A}_{\tau_c}^i, a_{k\tau_x}^{ih} = u_{k\tau_x}^{ih}\left(p_{h\tau_x}^{\mathbf{iont}}\right):$$

$$\mathbf{p}_h^{\mathbf{iont}} \to \mathbf{a}_k^{ih}, f_{l\tau_x}^{ihk\Xi*} = c^i\left(a^i\left(p_{h\tau_x}^{\mathbf{iont}}, p_k^{ih\Xi*}\right)\right),$$

где $u_{k\tau_x}^{ih}\left(p_{h\tau_x}^{\mathbf{iont}}\right)$ – *локальный* (для проблемы $p_{h\tau_x}^{\mathbf{iont}}$) *закон управления*, область определения которого $\mathbf{p}_h^{\mathbf{iont}} = \left\{p_{h\tau_x}^{\mathbf{iont}} \mid \tau_x \in [\tau_1, \ldots, \tau_{c^{\max}}]\right\}$ представляет собой множество точек потока проблем, в которых в качестве текущей проблемы агент $\aleph_i$ определяет проблему $p_{h\tau_x}^{\mathbf{iont}}$, а

область значений $\mathbf{a}_{k\tau_x}^{ih} = \left\{a_{k\tau_x}^{ih} \mid \tau_x \in [\tau_1, \dots, \tau_{c\mathbf{max}}]\right\}$ – это множество т.н. *специальных алгоритмов* $a_{k\tau_x}^{ih}$ решения проблемы $p_{h\tau_x}^{\boldsymbol{iont}}$, которые имеются в распоряжении агента $\aleph_i$ на шаге времени τ_x, являющееся частью

$$\mathbf{A}_{\tau_c}^{i} = \left\{\mathbf{a}_{k\tau_x}^{ih} \mid \forall k, \forall p_{h\tau_x}^{\boldsymbol{iont}} \in \mathbf{P}_{\tau_c}^{\boldsymbol{iont}}\right\} \in K^i$$

– множества всех таких алгоритмов, $\mathbf{P}^{\boldsymbol{iont}}$ – множество онтологизированных проблем. Будем считать, что $p_{h\tau_x}^{\boldsymbol{iont}} \neq p_{h\tau_y}^{\boldsymbol{iont}}$ и $a_{k\tau_x}^{ih} \neq a_{k\tau_y}^{ih}$ в силу того, что во всем периоде жизни агента $\aleph_i$ и проблемы, и алгоритмы модифицируются на основе обучения.

Остальные проблемы $p_{h\tau_x}^{\boldsymbol{inon}} \in \mathbf{P}_{\tau_c}^{\boldsymbol{inon}} \in \mathbf{P}^{\boldsymbol{i}}$ в потоке $p_{\tau_1}^{i\tau_{c\mathbf{max}}}$ агенту $\aleph_i$ на шаге времени τ_c не известны (не *онтологизированы*), т.е.:

$$\forall p_{h\tau_x}^{\boldsymbol{inon}} \in \mathbf{P}_{\tau_c}^{\boldsymbol{inon}}: \nexists a_{k\tau_x}^{ih} \in A^i, a_{k\tau_x}^{ih} = u_{k\tau_x}^{ih}\left(p_{h\tau_x}^{\boldsymbol{iont}}\right),$$

где $\mathbf{P}_{\tau_c}^{\boldsymbol{inon}}$ – все проблемы в потоке $p_{\tau_1}^{i\tau_{c\mathbf{max}}}$, не онтологизированные на шаге τ_c. Для эффективного функционирования агента $\aleph_i$ требуется, чтобы *коэффициент новизны среды* $W^{-\aleph_i}$ *для агента* $\aleph_i$

$$k_{\aleph_i\tau_c}^{W^{-\aleph_i}} = \left|\mathbf{P}_{\tau_c}^{\boldsymbol{inon}}\right| / \left|\mathbf{P}_{\tau_c}^{\boldsymbol{iont}}\right| \xrightarrow[\tau_c \to \tau_{c\mathbf{max}}]{} 0$$

приближался к нулю с увеличением времени функционирования агента. Чтобы коэффициент $k_{\aleph_i\tau_c}^{W^{-\aleph_i}}$ уменьшался, средняя скорость $v_{\tau_c}^{\aleph_i}\left(p_{h\tau_x}^{\boldsymbol{inon}}\right)$ появления неонтологизированных проблем $p_{h\tau_x}^{\boldsymbol{inon}}$ в потоке $p_{\tau_1}^{i\tau_{c\mathbf{max}}}$

$$v_{\tau_c}^{\aleph_i}\left(p_{h\tau_x}^{\boldsymbol{inon}}\right) = \left|\mathbf{P}_{\tau_c}^{\boldsymbol{inon}}\right| / (\tau_{c\mathbf{max}} - \tau_1) < v_{\tau_c}^{\aleph_i}\left(a_{k\tau_x}^{ih}\right) = \left|\mathbf{A}_{\tau_c}^{i}\right| / (\tau_{c\mathbf{max}} - \tau_1)$$

должна быть не больше средней скорости процесса онтологизации проблем $p_{h\tau_x}^{\boldsymbol{inon}}$. Такого соотношения агент $\aleph_i$ может достичь с помощью синтеза *специальных алгоритмов идентификации и решения* этих проблем $a_{k\tau_x}^{ih}$ *метаалгоритмами* $a_{k\tau_x}^{ih\mathrm{M}} \in \mathbf{A}_{\tau_c}^{i\mathrm{M}}$, которые реализуют функции

$$a_{k\tau_x}^{ih} = u_{k\tau_x}^{ih\mathrm{M}}\left(p_{h\tau_x}^{\boldsymbol{inon}}\right) \in U^i, \mathbf{A}_{\tau_c}^{i} = \mathbf{A}_{\tau_x}^{i} \cup a_{k\tau_x}^{ih},$$

при вычислении которых и строятся специальные алгоритмы $a_{k\tau_x}^{ih}$. Здесь $\mathbf{A}_{\tau_c}^{i}$ – множество *специальных алгоритмов* $a_{k\tau_x}^{ih}$, а $\mathbf{A}_{\tau_c}^{i\mathrm{M}}$ – множество *метаалгоритмов* $a_{k\tau_x}^{ih\mathrm{M}}$ синтеза таких специальных алгоритмов.

Внеся соответствующие поправки в закон управления $f^{ih} = u^i\left(r^{ih}\right)$ и вышеприведенную оптимизационную задачу:

$$f^i = u^{i\mathrm{M}a}\left(r^i\right) = f^{ih}, \forall p_{h\tau_x}^{i} \in p_{\tau_1}^{i\tau_{c\mathbf{max}}},$$

$$f^{ih} = u^{ih\mathrm{M}a}\left(p_{h\tau_x}^{i}\right) = \begin{cases} u_{k\tau_x}^{ih}\left(p_{h\tau_x}^{i}\right), \exists a_{k\tau_x}^{ih} \in \mathbf{A}^i \\ u_{k\tau_x}^{ih\mathrm{M}}\left(p_{h\tau_x}^{i}\right), \nexists a_{k\tau_x}^{ih} \in \mathbf{A}^i \end{cases}, p_{h\tau_x}^{i} = p^i\left(s^i\left(r_{h\tau_{x-y}}^{i\tau_x}\right)\right),$$

$$\forall x, \forall y, \forall h, \tau^i \left(w^i \left(f^i \right) \right) \xrightarrow[f^i = u^{i\mathrm{Ma}}(r^i) \in \mathbf{F}^i, e^i \le c(E^i), e^i(w^i_{\tau_c}) > 0 \Rightarrow \exists \aleph_i, e^i(w^i_{\tau_c}) = 0 \Rightarrow \nexists \aleph_i]{} \max,$$

определим внутри класса о-вычислителей подкласс *метаалгоритмических вычислителей* **M** (*м-вычислители*), подчиняющихся этому закону управления с учетом необходимости решения такой задачи.

Зададим состояние *коллективного агента* (*популяции агентов*) $\aleph^{i(\gamma)}_{\tau_c}\left(\aleph^{i(\gamma-1)}_{1\tau_c}, \aleph^{i(\gamma-1)}_{2\tau_c}, \dots, \aleph^{i(\gamma-1)}_{j^{\max}\tau_c}\right)$ (*коллективное состояние*) на шаге τ_c как множество состояний агентов $\aleph^{i(\gamma-1)}_{j\tau_c} \in \text{ה}^{i\gamma}_{\tau_c}$ на этом шаге:

$$S_{i\tau_c}\left(s^i_{1\tau_c}, s^i_{2\tau_c}, \dots, s^i_{j^{\max}\tau_c}\right) = S^i\left(R^i_{\tau_c}\right), R^i = \left(r^{i\tau_c}_{1\tau_{c-x_1}}, r^{i\tau_c}_{2\tau_{c-x_2}}, \dots, r^{i\tau_c}_{j^{\max}\tau_{c-x_{j^{\max}}}}\right).$$

Верхние индексы $\gamma, \gamma-1$ определяют т.н. *ранги* агентов, описывающие отношения вложенности агентов друг в друга.

Функция *разметки коллективных проблем*

$$P^{i\tau_c{}^{\max}}_{\tau_1} = \left(P^i_{1\tau_{x_1}}, P^i_{2\tau_{x_2}}, \dots, P^i_{d^{\max}\tau_{x^{d^{\max}}}}\right) = P^i\left(S^i\left(R^i\right)\right),$$

$$P^i_{d\tau_{x_d}} = \left\{p^{idh_{k_1}}_{k_1\tau_d}, p^{idh_{k_2}}_{k_2\tau_d}, \dots, p^{idh_{k^{\max}}}_{k^{\max}\tau_d}\right\}, \aleph^{i\gamma}_{\tau_c} \ni \aleph^{i\gamma}_{d\tau_c}\left(\aleph^{i(\gamma-1)}_{k_1\tau_c}, \aleph^{i(\gamma-1)}_{k_2\tau_c}, \dots, \aleph^{i(\gamma-1)}_{k^{\max}\tau_c}\right)$$

определяет проблемы $P^i_{d\tau_{x_d}}$, интерпретируемые в качестве коллективных состояний некоторых *коллективных агентов* $\aleph^{i\gamma}_{d\tau_c} \in \aleph^{i\gamma}_{\tau_c}$, в которых все агенты ранга $\gamma-1$ сталкиваются со значительными изменениями энергии. Здесь k_l – номер агента в популяции, h_{k_l} – номер локальной проблемы этого агента, связанной с коллективной проблемой $P^i_{d\tau_{x_d}}$. Будем считать, что агенты $\aleph^{i(\gamma-1)}_{k_l\tau_c} \in \aleph^{i\gamma}_{d\tau_c}$ обмениваются сообщениями $m^{iqd}_{ik_lz\tau_c}$, где i – номер популяции, d – номер коллективной проблемы, k_l – номер агента-отправителя сообщения, q – номер агента-получателя сообщения для координации своих действий при решении коллективной проблемы $P^i_{d\tau_{x_d}}$ с помощью перевода агента $\aleph^{i\gamma}_{d\tau_c}$ в целевое состояние:

$$P^{i\Xi^*}_{d\tau_{x_d+y}} = G^{i\gamma}\left\{u^{i(\gamma-1)\mathrm{Ma}}_{k_1}\left(r^{i\tau_c}_{1\tau_{c-x_{k_1}}}\right), u^{i(\gamma-1)\mathrm{Ma}}_{k_2}\left(r^{i\tau_c}_{2\tau_{c-x_{k_2}}}\right), \dots, u^{i(\gamma-1)\mathrm{Ma}}_{k^{\max}}\left(r^{i\tau_c}_{k^{\max}\tau_{c-x_{k^{\max}}}}\right)\right\} =$$

$$= \left\{p^{idh_{k_1}\Xi^*}_{k_1h^o_{k_1}\tau_{x_d+y}}, p^{idh_{k_2}\Xi^*}_{k_2h^o_{k_2}\tau_{x_d+y}}, \dots, p^{idh_{k^{\max}}\Xi^*}_{k^{\max}h^o_{k^{\max}}\tau_{x_d+y}}\right\},$$

где $p^{idh_{k_l}\Xi^*}_{k_lh^o_{k_l}\tau_{x_d+y}}$ – локальное целевое состояние агента $\aleph^{i(\gamma-1)}_{k_l\tau_c}$, а $G^{i\gamma}$ – *функция коллективного выбора цели* $P^{i\Xi^*}_{d\tau_{x_d+y}}$ агента $\aleph^{i\gamma}_{d\tau_c}$.

Определим в классе м-вычислителей подкласс *коллективных вычислителей* **Л** (*л-вычислители*), закон управления вычислительным циклом которых может быть представлен в виде:

$$F^{i\gamma} = U^{i\gamma} = \Phi^{i\gamma}\left(P^i_{d\tau_{x_d}}, P^{i\Xi^*}_{d\tau_{x_d+y}}\right),$$

где $\Phi^{i\gamma}$ – вспомогательная функция.

Пусть интеллектуальный агент $\aleph_{\tau_{b^i}}^{\upsilon i\gamma\tau_{e^i\mathbf{max}}}$ (например, постановщик задачи, оператор) для решения проблемы $p_{h_i\tau_d}^{id}$ строит алгоритм:

$$a_{l_i\tau_d}^{idh_i} = u_{l_i\tau_d}^{idh_i}\left(p_{h_i\tau_d}^{id}\right), f_{l_i\tau_d}^{idh_i\Xi*} = c^i\left(a^i\left(p_{h_i\tau_d}^{id}, p_{h_i^o\tau_{d+x}}^{idh_il_i\Xi*}\right)\right), \exists\aleph_{k_q\tau_d}^{\text{ם}i(\gamma-1)} \in \mathbf{Л},$$

$$P_{d\tau_{x_d}}^{i} = \left\{p_{h_i\tau_d}^{id}, p_{k_qh_{k_q}\tau_d}^{\text{ם}id}\right\}, P_{d\tau_{x_d+y}}^{i\Xi*} = \left\{p_{h_i^o\tau_{d+x}}^{idh_il_i\Xi*}, p_{k_qh_{k_q}^o\tau_{x_d+y}}^{\text{ם}idh_{k_q}l_{k_q}\Xi*}\right\},$$

$$a_{l_{k_q}\tau_d}^{\text{ם}idh_{k_q}} = u_{l_{k_q}\tau_d}^{\text{ם}idh_{k_q}}\left(p_{k_qh_{k_q}\tau_d}^{\text{ם}id}\right), f_{l_{k_q}\tau_{x_d+y}}^{\text{ם}idh_{k_q}\Xi*} = c^{\text{ם}i}\left(a^{\text{ם}i}\left(p_{k_qh_{k_q}\tau_d}^{\text{ם}id}, p_{h_{k_q}^o\tau_{d+y}}^{\text{ם}idh_{k_q}l_{k_q}\Xi*}\right)\right),$$

где $\aleph_{k_q\tau_d}^{\text{ם}i(\gamma-1)}$ – агент из класса л-вычислителей, который в соответствии с алгоритмом $a_{l_i\tau_d}^{idh_i}$ для решения локальной проблемы $p_{h_i\tau_d}^{id}$ оператора $\aleph_{\tau_{b^i}}^{\upsilon i\gamma\tau_{e^i\mathbf{max}}}$ должен решить свою локальную проблему $p_{k_qh_{k_q}\tau_d}^{\text{ם}id}$ при помощи алгоритма $a_{l_{k_q}\tau_d}^{\text{ם}idh_{k_q}}$. Сделав это, он совместно с оператором решит и коллективную проблему $P_{d\tau_{x_d}}^{i}$, перейдя к целевым проблемам $P_{d\tau_{x_d+y}}^{i\Xi*}$ при помощи генерации эффекторных траекторий $f_{l_i\tau_d}^{idh_i\Xi*}$ и $f_{l_{k_q}\tau_{x_d+y}}^{\text{ם}idh_{k_q}\Xi*}$. Здесь ם – индекс, обозначающий искусственное происхождение агента $\aleph_{k_q\tau_d}^{\text{ם}i(\gamma-1)}$. Оператор $\aleph_{\tau_{b^i}}^{\upsilon i\gamma\tau_{e^i\mathbf{max}}}$ для привлечения агента $\aleph_{k_q\tau_d}^{\text{ם}i(\gamma-1)}$ к совместному решению проблемы $p_{h_i\tau_d}^{id}$ при помощи функции языкового кодирования $\mathfrak{w}_l^{\upsilon i\mathbf{cod}}$ строит для этого агента задание, выраженное в сообщении

$$m_{iz\tau_d}^{\upsilon k_qdh_{k_q}} = \mathfrak{w}_l^{\upsilon i\mathbf{cod}}\left(P_{d\tau_{x_d}}^{i}, P_{d\tau_{x_d+y}}^{i\Xi*}, a_{l_i\tau_d}^{idh_i}, a_{l_{k_q}\tau_d}^{\text{ם}idh_{k_q}}\right),$$

раскодировав которое при помощи функции языкового декодирования $\mathfrak{w}_l^{k_q\mathbf{dec}}\left(m_{iz\tau_d}^{\upsilon k_qdh_{k_q}}\right)$, агент $\aleph_{k_q\tau_d}^{\text{ם}i(\gamma-1)}$ получает информацию о том, что он должен выполнить алгоритм $a_{l_{k_q}\tau_d}^{\text{ם}idh_{k_q}}$ с целью перехода от проблемы $p_{k_qh_{k_q}\tau_d}^{\text{ם}id}$ к проблеме $p_{h_{k_q}^o\tau_{d+y}}^{\text{ם}idh_{k_q}l_{k_q}\Xi*}$. Система «оператор – агент – среда» реализует коллективную траекторию:

$$F^{i\gamma} = U^{i\gamma} = \Phi^{i\gamma}\left(P_{d\tau_{x_d}}^{i}, P_{d\tau_{x_d+y}}^{i\Xi*}\right) = \left(f_{l_i\tau_d}^{idh_i\Xi*}, f_{l_{k_q}\tau_{x_d+y}}^{\text{ם}idh_{k_q}\Xi*}\right).$$

В силу того, что вся информация приходит к агенту $\aleph_{k_q\tau_d}^{\text{ם}i(\gamma-1)}$ во входном сенсорном потоке, можно записать:

$$p_{h\tau_d}^{\text{ם}k_qdh_{k_q}} = p^i\left(s^i\left(r_{h\tau_{d-g}}^{i\tau_d}, \mathfrak{w}_l^{k_q\mathbf{dec}}\left(m_{iz\tau_d}^{\upsilon k_qdh_{k_q}}\right)\right)\right),$$

где $p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}$ – новая локальная проблема, решение которой детерминировано необходимостью выполнения задания $m_{iz\tau_d}^{\mathbb{v}k_q dh_{k_q}}$, принятого от оператора $\aleph_{\tau_{b^i}}^{\mathbb{v}i\gimel\tau_{e^i\mathbf{max}}}$. Эту проблему агент $\aleph_{k_q\tau_d}^{\mathbb{n}i(\gimel-1)}$ идентифицирует в своей системе управления на базе анализа сенсорной траектории $r_{h\tau_{d-g}}^{i\tau_d}$ и декодированных функцией $\mathfrak{w}_l^{k_q\mathbf{dec}}\left(m_{iz\tau_d}^{\mathbb{v}k_q dh_{k_q}}\right)$ данных о $P_{d\tau_{x_d}}^{i}, P_{d\tau_{x_d+y}}^{i\Xi*}, a_{l_i\tau_d}^{idh_i}, a_{l_{k_q}\tau_d}^{\mathbb{n}idh_{k_q}}$.

Построив при помощи функции p^i локальную проблему $p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}$, агент $\aleph_{k_q\tau_d}^{\mathbb{n}i(\gimel-1)} \in \mathbf{Л}$ с помощью своего закона управления генерирует траекторию:

$$f_{j\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}hl_{k_q}\Xi*} = \left(f_{j\tau_{x_d}}^{\mathbb{n}idh_{k_q}hl_{k_q}\Xi*}, f_{l_{k_q}\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}\Xi*}\right) = u^{\mathbb{n}ih\mathrm{M}a}\left(p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}\right) =$$

$$= \begin{cases} u_{k_q\tau_x}^{\mathbb{n}ih}\left(p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}\right), \exists a_{k_q\tau_d}^{\mathbb{n}ih} \in \mathbf{A}^{\mathbb{n}k_q} \\ u_{k_q\tau_x}^{\mathbb{n}ih\mathrm{M}}\left(p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}\right), \nexists a_{k_q\tau_d}^{\mathbb{n}ih} \in \mathbf{A}^{\mathbb{n}k_q} \end{cases},$$

где $f_{j\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}hl_{k_q}\Xi*}$ – *общая эффекторная траектория*, $f_{l_{k_q}\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}\Xi*}$ – ее *целевая часть*, а $f_{j\tau_{x_d}}^{\mathbb{n}idh_{k_q}hl_{k_q}\Xi*}$ – *подготовительная часть* общей траектории. Для того чтобы из текущего состояния $\aleph_{k_q\tau_d}^{\mathbb{n}i(\gimel-1)}$ оператор $\aleph_{\tau_{b^i}}^{\mathbb{v}i\gimel\tau_{e^i\mathbf{max}}}$ смог реализовать целевую часть, ему необходимо сначала исполнить подготовительную часть общей траектории.

Агент $\aleph_{k_q\tau_d}^{\mathbb{n}i(\gimel-1)}$ при условии присутствия специального алгоритма $a_{k_q\tau_d}^{\mathbb{n}ih}$ решения проблемы $p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}$ в базе знаний и алгоритмов (геноме) $\mathbf{A}^{\mathbb{n}k_q}$ использует выполняемую этим алгоритмом функцию $u_{k_q\tau_x}^{\mathbb{n}ih}\left(p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}\right)$ для построения общей эффекторной траектории $f_{j\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}hl_{k_q}\Xi*}$. При условии отсутствия такого алгоритма агент $\aleph_{k_q\tau_d}^{\mathbb{n}i(\gimel-1)}$ применяет метаалгоритмическую функцию $u_{k_q\tau_x}^{\mathbb{n}ih\mathrm{M}}\left(p_{h\tau_d}^{\mathbb{n}k_q dh_{k_q}}\right)$ для того, чтобы создать специальный алгоритм $a_{k_q\tau_d}^{\mathbb{n}ih}$.

Таким же образом агент $\aleph_{k_q\tau_d}^{\mathbb{n}i(\gimel-1)}$ действует и в случае, когда в процессе построения и выполнения общей эффекторной траектории $f_{j\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}hl_{k_q}\Xi*}$ в потоке $p_{\tau_d}^{k_q\tau_{x_d+y}}$ появляются новые проблемы $p_{h^b\tau_{x_d+y_b}}^{\mathbb{n}k_q dh_{k_q}}$, которые ведут к отклонению от целевой части траектории $f_{l_{k_q}\tau_{x_d+y}}^{\mathbb{n}idh_{k_q}\Xi*}$.

Соответственно, для построения и реализации эффекторной траектории $F^{i\gamma}$ оператору $\aleph_{\tau_b i}^{\upsilon i\gamma\tau_{e^i\max}}$ требуется отправить агенту $\aleph_{k_q\tau_d}^{\daleth i(\gamma-1)}$ сообщение $m_{iz\tau_d}^{\upsilon k_q dh_{k_q}}$. После этого все действия, необходимые для генерации общей эффекторной траектории $f_{j\tau_{x_d+y}}^{\daleth idh_{k_q}hl_{k_q}\Xi*}$, с учетом действий, необходимых для онтологизации, идентификации и решения проблем $p_{h^b\tau_{x_d+y_b}}^{\daleth k_q dh_{k_q}}$, искусственный интеллектуальный агент должен выполнить самостоятельно.

Агента $\aleph_{k_q\tau_d}^{\daleth i(\gamma-1)} \in$ **Л,** закон управления которого имеет вид:

$$f_{j\tau_{x_d+y}}^{\daleth idh_{k_q}hl_{k_q}\Xi*} = u^{\daleth ih\mathrm{M}a}\left(p_{h\tau_d}^{\daleth k_q dh_{k_q}}\right),$$

будем называть *агентом универсального искусственного интеллекта* (*универсальным искусственным интеллектом* (УИИ). Таких агентов будем относить к формальному классу **Н** (н-вычислители).

2. Задача моделирования сознания агента универсального искусственного интеллекта

Исходя из вышеизложенного, опираясь на принятое определение, задачу имитационного моделирования сознания агента универсального интеллекта в реальной среде будем рассматривать как задачу создания нейрокогнитивной системы управления нейрокогнитивной системой управления движением такого агента в пространстве его поведения.

Можно предположить, что основная функция сознания интеллектуального агента, интерпретируемого таким образом, состоит в управлении выбором вариантов фрагментов его будущей траекторий в пространстве поведения, синтезируемых функциональными системами нейрокогнитивной архитектуры нижнего уровня (подсознания) этого агента, формировании плана чередования последовательности таких фрагментов исходя из критериев оптимизации сложной целевой функции.

Из этой гипотезы, в частности, вытекает, что имитационная модель мультиагентных нейрокогнитивных пьес в «картезианском театре» сознания должна учитывать цели и способы оптимального решения задачи синтеза кусочной (по составу задач) траектории движения агента универсального искусственного интеллекта в пространстве поведения. Решению этой же задачи должны быть посвящены структурно-функциональные модели всех сенсоров и эффекторов нейрокогнитивной архитектуры сознания. Так как объектом их наблюдения и управления является нейрокогнитивная архитектура подсознания интеллектуального агента, будем, соответственно, называть их *психосенсорами* и *психоэффекторами*. В частности, состав и функции психоэффекторов, приведенных на рисунке 1, должны быть детерминированы задачами редактирования графа проблем и решений интеллектуального агента, должны позволять добавлять и удалять его элементы, выполнять их разметку маркерами предпочтений (верить, считать невыполнимыми, невероятными и т.п.), рекомбинировать с частями других графов и т.д.

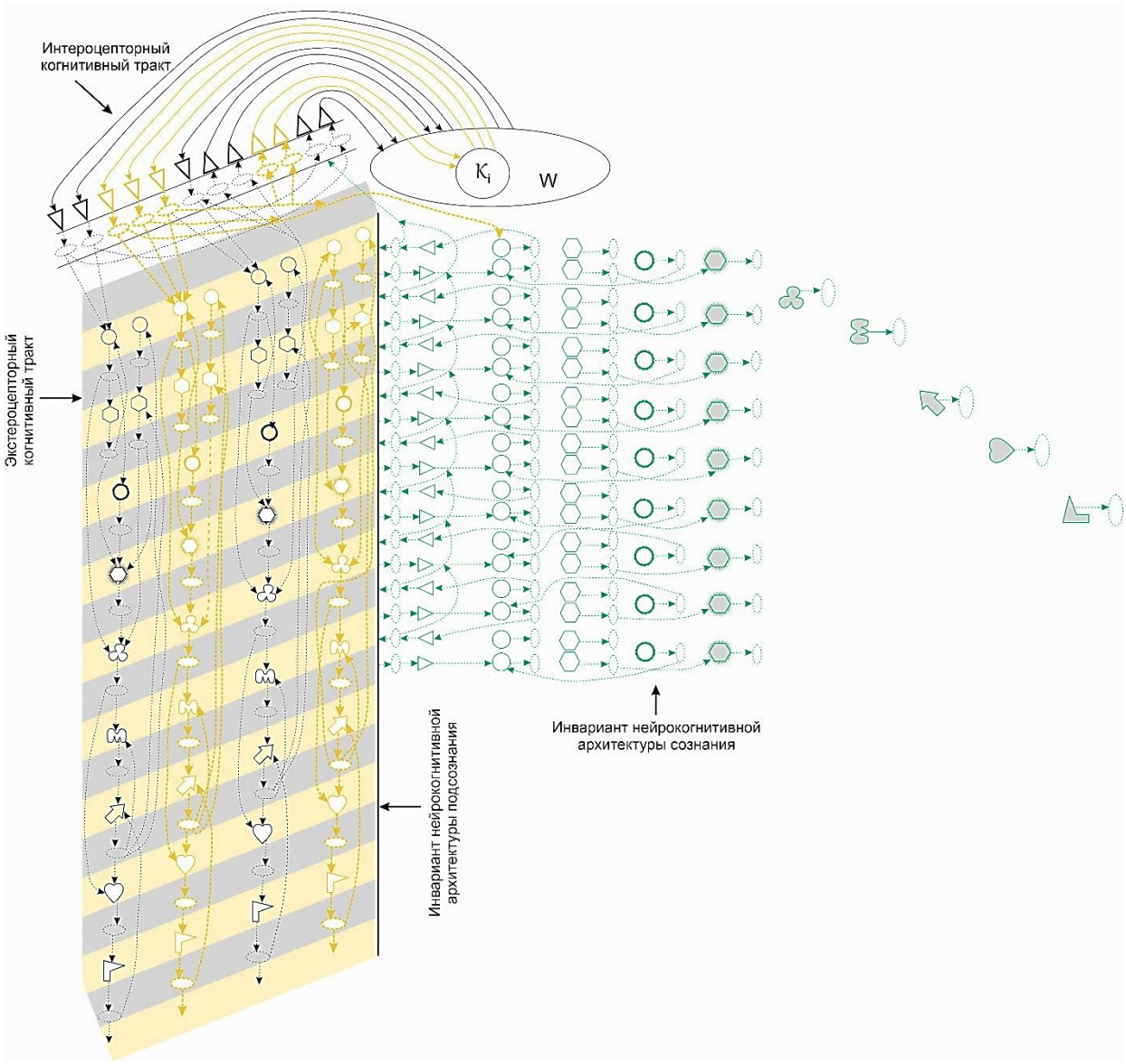

Рис. 1. *Схема взаимодействия нейрокогнитивных архитектур подсознания и сознания агента универсального искусственного интеллекта*

Fig. 1. *Interaction scheme of neurocognitive architectures of subconscious and consciousness of an agent of universal artificial intelligence*

Таким образом, инвариант нейрокогнитивной архитектуры сознания должен реализовывать контур управления субоптимальным выбором альтернатив составных частей проблемы и ее решения, распределенных по инварианту нейрокогнитивной архитектуры нижнего уровня, автономно синтезируемых нейрокогнитонами в ее составе.

3. Гипотеза о структуре сознания агента универсального искусственного интеллекта

С целью разработки метафоры проектирования имитационной модели сознания агента универсального искусственного интеллекта предлагается включить в состав такой модели нейрокогнитивную модель собственного «Я» этого агента, состоящую из четырех взаимодействующих компонентов: нейрокогнитивная модель внешней среды, построенная на ос-

нове экстероцепторного потока данных (экстероцепторная модель реальности); нейрокогнитивная модель своего «тела», построенная на основе интероцепторного потока данных (интероцепторная модель); нейрокогнитивная модель своей мыслительной деятельности, построенная на основе данных психоцепторного потока данных (психоцепторная модель); нейрокогнитивная вербальная модель, построенная на основе функционального представительства в управляющей нейрокогнитивной архитектуре интеллектуального агента феноменологии описания мыслительной деятельности средствами естественного языка, созданными при интерактивном взаимодействии в различных коммуникативных контекстах и ситуациях (рис. 2).

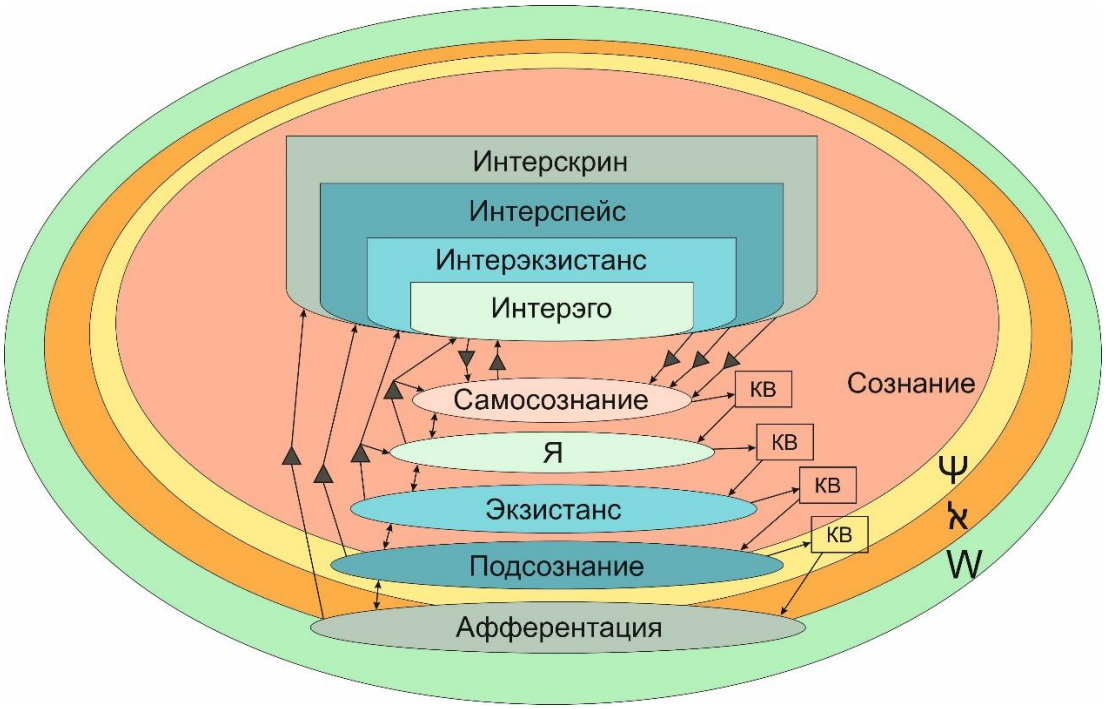

Рис. 2. *Четырехкомпонентная модель собственного «Я» агента универсального искусственного интеллекта и гетерогенные потоки данных*

Fig. 2. *Four-component model of the agent's own "I" of universal artificial intelligence and heterogeneous data flows*

На рисунке 2 приведена функциональная схема взаимодействия нейрокогнитивной модели «Я» интеллектуального агента с потоками интероцепторной, эктероцепторной и психоцепторной информации на основе метафоры проектирования нейрокогнитивного интерскрина и нейрокогнитивного интерспейса.

Нейрокогнитивным интерскрином (*интерскрин*) назовем функциональную систему, состоящую из агентов-нейронов, выполняющих мультиагентные пьесы, обеспечивающие синхронный избирательный доступ агентов-нейронов нейрокогнитивной архитектуры подсознания и агентов-нейронов нейрокогнитивной архитектуры сознания.

Нейрокогнитивным интерспейсом (*интерспейс*) назовем функциональную систему из агентов-нейронов, реализующих пьесы для обеспечения внутреннего представления (интериоризации – по Выготскому) проблемных ситуаций, в деревьях решений которых сознание ведет поиск фрагментов траекторий интеллектуального агента в пространстве состояний в прошлом, настоящем и будущем.

Часть интерспейса, описывающую мультиагентные факты, устанавливаемые нейрокогнитивной архитектурой интеллектуального агента на основании интероцепторных данных, назовем *нейрокогнитивным интерэкзистансом* (*интерэкзистанс*).

Часть интерэкзистанса, описывающую мультиагентные факты, устанавливаемые нейрокогнитивной архитектурой интеллектуального агента на основе психоцепторных данных, назовем *нейрокогнитивным интерэго* (*интерэго*) этого агента.

Согласно нашей гипотезе, онтогенетическое развитие естественных интеллектуальных агентов универсального искусственного интеллекта начинается в условиях сформировавшихся в нейрокогнитивных архитектурах подсознания и сознания минимальных функциональных систем «Я», интерспейса, интерскрина, инерэкзистанса и интерэго, достаточных для того, чтобы в процессе жизни этого агента в реальной коммуникативной среде с помощью вышеописанных нейрогенетических функций и функций обучения выполнялась активная достройка этих подсистем, интерпретируемая как процесс формирования личности интеллектуального агента.

На рисунке 2 видно, что с интерспейсом симметричным образом взаимодействуют афферентный экстероцепторный тракт и *нейрокогнитивный конструктор воображения* – функциональная подсистема, обеспечивающая работу функциональных систем афферентных анализаторов на основе работы психоэффекторов, выполняющих команды инварианта нейрокогнитивной архитектуры сознания.

Как следует из рисунка 2, аналогичным образом взаимодействуют между собой нейрокогнитивный конструктор воображения, интерэкзистанс и афферентный интероцепторный тракт, а также конструкторы воображения, обозначенные на рисунке буквами *КВ*, интерэго и афферентный психоцепторный тракт.

Именно использование интерспейса, интерскрина и интерэго позволяет сознанию варьировать альтернативы кусочных траекторий, интегрируя в единых графах проблемных ситуаций мультиагентные факты, воспринимаемые с помощью афферентных подсистем, и мультиагентные факты, воображаемые с помощью психоэффекторной стимуляции нейрокогнитивных конструкторов воображения.

Принципиальное значение для создания целостного восприятия действительности и себя в этой действительности имеет схема направлений потоков движения сигналов (сообщений) от различных частей нейрокогнитивных архитектур верхнего и нижнего уровней, отнесенных к подсознанию и сознанию. На рисунке 2 видно, что только самосознание имеет доступ к просмотру всех пьес, которые синхронно и на основании одних и тех же сигналов играются на сценах интерэго, интерэкзистанса, интерспейса и интерскрина, что и создает иллюзию целостности.

На рисунке 1 показана часть агентов-нейронов инварианта нейрокогнитивной архитектуры, обеспечивающих нейронное представительство структурно-функциональных компонентов подсознания и сознания, выделенных на рисунке 2. Однако это далеко не полный состав. Например, в качестве нейроморфного субстрата компонентов «афферентация» и «подсознание» должны выступать агенты-нейроны первичной апперцепции, топологические агенты-нейроны, агенты-нейроны-признаки (квалиативные агенты-нейроны), далее последовательно – концептуальные, событийные, эмоциональные, моделирующие агенты-нейроны.

Системная целесообразность доступности нейрокогнитивной архитектуре самосознания всех сцен множественного представления на разных уровнях картезианского театра состоит в обеспечении возможности комбинирования в единой субъективной реальности всех текущих локальных проблем и их деревьев, что необходимо для варьирования и обмена фрагментами

кусочных траекторий в воображении при решении задачи субоптимального синтеза непрерывной траектории интеллектуального агента в пространстве поведения.

Интеграция уровней представления обработанных сигналов позволяет привести афферентные экстероцепторные, интероцепторные и психоцепторные потоки к единому формату восприятия, что обеспечивает возможность комбинирования и синхронной актуализации модели реальности на базе действительных и воображаемых событий.

Принципиальное значение имеет двусторонняя связь конструкторов воображения с интерсистемами (интерэго, интерэкзистанс, интерспейс, интерскрин), позволяющая быстро переходить между сценами для уточнения моделей и перестройки контекстов.

Согласно нашей гипотезе, множественные процессы восприятия, синхронизируемые на сценах разных уровней, на которые попадают сигналы с различной глубиной нейрокогнитивной обработки, лежат в основе осознания осмысленности восприятия, являющегося неотъемлемой функциональной характеристикой сознания.

4. Гипотеза об онтогенезе нейрокогнитивных моделей сознания агента

Развивая идею о нейрокогнитивной архитектуры сознания как о структурно-функциональном комплексе управления выбором траекторий движения интеллектуального агента в пространстве поведения, и учитывая условия и трудности принятия решений в реальной среде, представляется конструктивной гипотеза о филогенезе этого комплекса, детерминированном эволюционным усложнением объекта управления.

В предложенной нами ранее концепции инварианта мультиагентной нейрокогнитивной архитектуры для управления синтезом траектории идентификации и решения одной проблемы достаточно нижнего, подсознательного, уровня. Необходимость синтеза кусочной траектории в пространстве поведения возникает при переходе к многозадачным агентам, способным к субоптимальному распределению ресурсов при синхронном решении нескольких проблем.

Предположив, что в начальный момент жизни интеллектуальный агент оснащен еще и третьим уровнем нейрокогнитивной архитектуры, позволяющим управлять процессами в двух первых ее уровнях, получаем стартовую конфигурацию для экспериментов по онтогенетическому синтезу моделей сознания агента универсального искусственного интеллекта.

Применение нейрогенетических функций, метафункциональных систем и алгоритмов онтонейрогенеза [3, 10] позволяет синхронно обучать все компоненты сознания на всех уровнях нейрокогнитивной архитектуры, представленных на рисунке 2.

Кроме того, при взаимодействии в коммуникативном агенте универсального искусственного интеллекта как агент, относящийся к классу социальных вычислителей, получает всю необходимую информацию для освоения естественного языка [5], с помощью которого он уже может идентифицировать психические состояния и процессы.

Таким образом, индивидуальное развитие сопровождается достройкой нейроморфологического субстрата на всех уровнях управляющей нейрокогнитивной архитектуры интеллектуального агента, обеспечивая синтез и развитие образов мира, себя, себя в мире и мира в себе, – что представляет собой базовый функционал, без которого нельзя представить сознание.

Достройка образа «Я» происходит за счет мультиагентных фактов, построенных над интерцепторными и психоцепторными потоками. Массивы специализированных агентов-нейронов формируют в когнитивной архитектуре своего рода структурно-функцио-

нальный субстрат настоящего, памяти и воображения, манифестация которого в контексте синтеза поведения интеллектуального агента обусловливает феноменологические проявления его психики.

Можно предположить, что к настоящему в подобной нейрокогнитивной архитектуре можно отнести феномены, которые пока находятся в процессе фактологизации (создания внутренних мультиагентных нейрокогнитивных фактов, элементов функциональных репрезентаций событий). В это время на сценах интерскрина и интерспейса идет демонстрация сигналов разной степени обработки. Гипотетически этот процесс должен завершаться с истечением времени жизни афферентных сообщений.

При этом факты и феномены психологической жизни, относящиеся к сценам интерэкзистанса и интерэго, диахронически в процессе индивидуального развития возникают раньше, поэтому процессы достройки и обучения нейроморфологического субстрата экзистанса, «Я» и самосознания в имитационной модели необходимо симулировать в первую очередь.

Заключение

В результате выполнения исследования дано формальное определение агента универсального искусственного интеллекта.

Предложена гипотеза о структурно-функциональной организации сознания универсального искусственного интеллекта на основе метафоры проектирования многокомпонентной мультиагентной нейрокогнитивной архитектуры.

Разработаны некоторые принципы имитационного моделирования сознания агентов универсального искусственного интеллекта на основе контекстно-детерминированного развития управляющей нейрокогнитивной архитектуры в коммуникативной социальной реальной среде.

СПИСОК ЛИТЕРАТУРЫ

1. *Russell S., Norvig P.* Artificial Intelligence: A Modern Approach (AIMA). 2nd ed. Moscow: Williams, 2007. 1424 p.

2. *Goertzel B.* Artificial General Intelligence: Concept, State of the Art, and Future Prospects // Journal of Artificial General Intelligence. 5(1). 1–46. 2014. DOI: 10.2478/jagi-2014-0001

3. *Нагоев З. В.* Интеллектика, или Мышление в живых и искусственных системах. Нальчик: Издательство КБНЦ РАН, 2013. 232 с.

4. *Нагоев З. В.* Мультиагентные экзистенциальные отображения и функции // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 63–71. EDN: QZTFLX

5. *Нагоев З. В., Нагоева О. В.* Обоснование символов и мультиагентные нейрокогнитивные модели семантики естественного языка. Нальчик: Издательство КБНЦ РАН, 2022. 150 с.

6. *Nagoev Z., Nagoeva O., Anchokov M. et al.* The symbol grounding problem in the system of general artificial intelligence based on multi-agent neurocognitive architecture. Cognitive Systems Research. 2023. Vol. 79. Pp. 71–84. DOI: 10.1016/j.cogsys.2023.01.002

7. *Nagoev Z., Pshenokova I., Nagoeva O., Sundukov Z.* Learning algorithm for an intelligent decision making system based on multi-agent neurocognitive architectures. Cognitive Systems Research. 2021. Vol. 66. Pp. 82–88. DOI: 10.1016/j.cogsys.2020.10.015

8. *Нагоев З. В.* Нейрокогнитивные отображения и функции для моделей нейроморфогенеза в системах управления интеллектуальных онтофилогенетических агентов // Известия Кабардино-Балкарского научного центра РАН. 2024. Т. 26. № 6. С. 188–196. DOI: 10.35330/1991-6639-2024-26-6-188-196

9. *Нагоев З. В., Нагоева О. В., Макоева Д. Г., Гуртуева И. А.* Мультиагентный нейрокогнитивный алгоритм управления референцией речевых событий коммуникации агента общего искусственного интеллекта в ситуации синхронных множественных диалогов // Известия Кабардино-Балкарского научного центра РАН. 2023. № 6(116). С. 193–209. DOI: 10.35330/1991-6639-2023-6-116-193-209

10. *Нагоев З. В.* Онтонейроморфогенетическое моделирование // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 56–63. EDN: QZTFLN

REFERENCES

1. Russell S., Norvig P. Artificial Intelligence: A Modern Approach (AIMA). 2nd ed. Moscow: Williams, 2007. 1424 p.

2. *Goertzel B.* Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence.* 5(1). 1–46. 2014. DOI: 10.2478/jagi-2014-0001

3. Nagoev Z.V. *Intellektika, ili Myshleniye v zhivykh i iskusstvennykh sistemakh* [Intellectics, or Thinking in Living and Artificial Systems]. Nalchik: Izdatel'stvo KBNTS RAN, 2013. 232 p. (In Russian)

4. Nagoev Z.V. Multi-agent existential mappings and functions. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2013. No. 4(54). Pp. 63–71. EDN: QZTFLX. (In Russian)

5. Nagoev Z.V., Nagoeva O.V. *Obosnovaniye simvolov i mul'tiagentnyye neyrokognitivnyye modeli semantiki yestestvennogo yazyka* [Justification of symbols and multi-agent neurocognitive models of natural language semantics]. Nalchik: Izdatel'stvo KBNTS RAN, 2022. 150 p. (In Russian)

6. *Nagoev Z., Nagoeva O., Anchokov M. et al.* The symbol grounding problem in the system of general artificial intelligence based on multi-agent neurocognitive architecture. *Cognitive Systems Research*. 2023. Vol. 79. Pp. 71–84. DOI: 10.1016/j.cogsys.2023.01.002

7. *Nagoev Z., Pshenokova I., Nagoeva O., Sundukov Z.* Learning algorithm for an intelligent decision making system based on multi-agent neurocognitive architectures. *Cognitive Systems Research*. 2021. Vol. 66. Pp. 82–88. DOI: 10.1016/j.cogsys.2020.10.015

8. Nagoev Z. V. Neurocognitive mappings and functions for neuromorphogenesis models in control systems of intelligent ontophylogenetic agents. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2024. Vol. 26. No. 6. Pp. 188–196. DOI: 10.35330/1991-6639-2024-26-6-188-196. (In Russian)

9. Nagoev Z.V., Nagoeva O.V., Makoeva D.G., Gurtueva I.A. Multi-agent neurocognitive algorithm for controlling the reference of speech events of communication of an agent of general artificial intelligence in a situation of synchronous multiple dialogues. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2023. No. 6(116). Pp. 193–209. DOI: 10.35330/1991-6639-2023-6-116-193-209. (In Russian)

10. Nagoev Z.V. Ontoneuromorphogenetic modeling. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2013. No. 4(54). Pp. 56–63. EDN: QZTFLN. (In Russian)

Финансирование. Исследование проведено без спонсорской поддержки.

Funding. The study was performed without external funding.

Информация об авторе

Нагоев Залимхан Вячеславович, канд. техн. наук, генеральный директор Кабардино-Балкарского научного центра РАН; вед. науч. сотр. отдела «Мультиагентные системы», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360000, Россия, Нальчик, ул. И. Арманд, 37-а;

zaliman@mail.ru, ORCID: https://orcid.org/0000-0001-9549-1823, SPIN-код: 6279-5857

Information about the author

Zalimkhan V. Nagoev, Candidate of Engineering Sciences, General Director of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences; Leading Researcher, Department of Multi-Agent Systems Institute of Computer Science and Problems of Regional Management – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

360000, Russia, Nalchik, 37-a I. Armand street;

zaliman@mail.ru, ORCID: https://orcid.org/0000-0001-9549-1823, SPIN-code: 6279-5857