

УДК 004.89

DOI: 10.35330/1991-6639-2025-27-6-186-208

EDN: ZGNGOG

Научная статья

Нейропсихологическая архитектура агента универсального искусственного интеллекта

З. В. Нагоев^{1,2}

¹Кабардино-Балкарский научный центр Российской академии наук
360010, Россия, г. Нальчик, ул. Балкарова, 2

²Институт информатики и проблем регионального управления –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, г. Нальчик, ул. И. Арманд, 37-а

Аннотация. Объектом исследования является нейропсихологическая архитектура мозга.

Цель исследования – разработка теоретических оснований создания агента универсального искусственного интеллекта.

Методы исследования. Используется подход к онтогенетическому формированию универсальных систем управления, сущность которых заключается в имитационном моделировании процессов роста и развития естественных автономных интеллектуальных агентов, оснащенных базовой когнитивной архитектурой, обладающей структурно-функциональным сходством с когнитивной архитектурой головного мозга, в реальной социальной среде. В частности, выдвинута ключевая гипотеза о том, что таким сходством обладают так называемые *нейрокогнитивные архитектуры*, представляющие собой метафору проектирования интеллектуальной системы управления поведением автономного агента, погруженного в реальную среду с помощью сенсоров и эффекторов, основанную на представлениях о составе и способах взаимодействия функциональных узлов так называемого *инварианта организации когнитивной архитектуры акта принятия решений*.

Результаты. Разработана структурно-функциональная схема нейропсихологической архитектуры, мотивированная ее системным назначением, обеспечивающая нейрокогнитивные трансформации пространства определения проблем с помощью согласованной работы нейрокогнитивных компонентов, реализующих каскады и-функций для динамических переходов между пространством поведения, локальными пространствами определения отдельных проблем и пространством психики.

Заключение. Разработана концепция нейропсихологической архитектуры автономной системы управления поведением агента универсального интеллекта. Показано, что универсальный характер спектра проблем, которые способны обрабатывать такие агенты, обеспечивается методологической и алгоритмической базой онтологизации, идентификации и решения проблем на основе синтеза нейропсихологической архитектурой агента его поведения в системе «агент – среда».

Ключевые слова: универсальный искусственный интеллект, нейропсихологическая архитектура, нейрокогнитивная архитектура, нейрокогнитивные функции и отображения, мультиагентные системы и функции

Поступила 06.11.2025, одобрена после рецензирования 27.11.2025, принята к публикации 10.12.2025

Для цитирования. Нагоев З. В. Нейропсихологическая архитектура агента универсального искусственного интеллекта // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 6. С. 186–208. DOI: 10.35330/1991-6639-2025-27-6-186-208

Neuropsychological architecture of a general-purpose artificial intelligence agent

Z.V. Nagoev^{1,2}

¹Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
2, Balkarov street, Nalchik, 360010, Russia

²Institute of Computer Science and Problems of Regional Management –
branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
37-a, I. Armand street, Nalchik, 360000, Russia

Abstract. The object of the study is the neuropsychological architecture of the brain.

Aim. The study is to develop theoretical foundations for the creation of a universal artificial intelligence agent.

Research methods. An approach to the ontogenetic formation of universal control systems is used. The essence of these systems lies in the simulation modeling of the growth and development of natural autonomous intelligent agents, equipped with a basic cognitive architecture possessing structural and functional similarities to the cognitive architecture of the brain, in a real social environment. Specifically, a key hypothesis is advanced that such similarities are possessed by so-called neurocognitive architectures, which represent a metaphor for designing an intelligent system for controlling the behavior of an autonomous agent immersed in a real environment using sensors and effectors. This architecture is based on concepts regarding the composition and interaction modes of functional units—the so-called invariants of the organization of the cognitive architecture of decision-making.

Results. A structural and functional diagram of neuropsychological architecture has been developed, motivated by its systemic purpose, enabling neurocognitive transformations of the problem-defining space through the coordinated work of neurocognitive components implementing cascades of n-functions for dynamic transitions between the behavioral space, local problem-defining spaces, and the mental space.

Conclusion. A concept for the neuropsychological architecture of an autonomous behavioral control system for a universally intelligent agent has been developed. It is shown that the universal nature of the range of problems that such agents are capable of handling is ensured by a methodological and algorithmic framework for ontologization, identification, and problem solving based on the synthesis of the agent's behavior within the agent-environment system by the neuropsychological architecture.

Keywords: general artificial intelligence, neuropsychological architecture, neurocognitive architecture, neurocognitive functions and mappings, multi-agent systems and functions

Submitted 06.11.2025,

approved after reviewing 27.11.2025,

accepted for publication 10.12.2025

For citation. Nagoev Z.V. Neuropsychological architecture of a general-purpose artificial intelligence agent. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2025. Vol. 27. No. 6. Pp. 186–208. DOI: 10.35330/1991-6639-2025-27-6-186-208

ВВЕДЕНИЕ

Грандиозный по объему и планетарный по масштабу запрос на повсеместное применение интеллектуальных агентов в составе человеко-машинных коллективов, сформировавшийся к настоящему времени на фоне успехов генеративных систем, основанных на глубоком обучении больших языковых моделей, сегодня сталкивается с некоторыми взаимосвязанными непреодолимыми проблемами, настолько серьезными и фундаментальными, что некоторые исследователи считают их предвестниками новой продолжительной «зимы» искусственного интеллекта [1]. К числу таких проблем относятся отсутствие теоретической базы интеллектуального принятия решений и автономного управления поведением

программных и роботизированных агентов в условиях реального мира, семантических оснований имитационного моделирования такого поведения, трудности реализации «логики здравого смысла» в условиях отсутствия таких оснований, слабая пригодность существующих формализмов для описания таких оснований, семантики и логики, высокая вычислительная трудоемкость алгоритмов на базе таких формализмов, критическая зависимость таких алгоритмов от количества хорошо структурированных наборов данных, высокие требования к производительности, энергоемкости и стоимости вычислений, чрезмерные затраты труда на адаптацию решений для новых задач, создание онтологий и обучение [1, 2].

Все эти проблемы видятся еще более значимыми и рельефными на фоне очевидной констатации способности мозга человека их успешно преодолевать. В этой связи представляется *актуальным* исследование основных принципов структурно-функциональной организации мозга в перспективе возможности их применения для разработки когнитивной архитектуры интеллектуальных систем управления универсальных автономных агентов, погруженных в реальную коммуникативную среду.

Целью исследования является разработка теоретических оснований создания агента универсального искусственного интеллекта.

Предмет исследования – возможность системного обоснования метафоры проектирования нейропсихологической архитектуры мозга.

Основная задача исследования – построение вычислительной абстракции агента универсального искусственного интеллекта на основе метафоры проектирования нейропсихологической архитектуры интеллектуального агента в реальной среде.

Актуальность исследования определяется перспективами применения основных принципов структурно-функциональной организации мозга для разработки когнитивной архитектуры систем управления агентов универсального искусственного интеллекта. В работе исследуется возможность системного обоснования метафоры проектирования нейропсихологической архитектуры мозга. Разработана концепция нейропсихологической архитектуры автономной системы управления поведением автономного агента в реальной среде, объединяющей в своем составе взаимосвязанные нейрокогнитивные компоненты подсознания, сознания, интерскрина, модуля «Я». Построена концепция агента универсального искусственного интеллекта на основе метафоры проектирования нейропсихологической архитектуры. Разработана вычислительная абстракция процессов онтологизации, идентификации и решения проблем универсального спектра нейропсихологической архитектурой на основе концепции синтеза автономным агентом траектории своего движения в пространстве поведения. Показано, что нейропсихологическая архитектура системы управления интеллектуального агента обеспечивает синтез такой траектории на основе интеграции в ее состав фрагментов траекторий движения агента в локальных пространствах определения проблем. Структурно-функциональная схема нейропсихологической архитектуры обоснована ее системным назначением, связанным с обеспечением снижения вычислительной трудоемкости алгоритмов решения универсального спектра проблем путем их переопределения на основе согласованной работы нейрокогнитивных компонентов, реализующих каскады н-функций для динамических переходов между пространством поведения, локальными пространствами определения отдельных проблем и пространством психики интеллектуального агента, выполняемых в процессе такой интеграции.

1. РЕШЕНИЕ ПРОБЛЕМ НА ОСНОВЕ НЕЙРОКОГНИТИВНОГО СИНТЕЗА ПОВЕДЕНИЯ

В работах [3, 4, 5] намечены подходы к онтогенетическому формированию универсальных систем управления, сущность которых заключается в имитационном моделировании процессов роста и развития естественных автономных интеллектуальных агентов, оснащенных базовой когнитивной архитектурой, обладающей структурно-функциональным

сходством с когнитивной архитектурой головного мозга, в реальной социальной среде. В частности, выдвинута ключевая гипотеза о том, что таким сходством обладают т.н. *нейрокогнитивные архитектуры*, представляющие собой метафору проектирования интеллектуальной системы управления поведением автономного агента, погруженного в реальную среду с помощью сенсоров и эффекторов, основанную на представлениях о составе и способах взаимодействия функциональных узлов т.н. *инварианта организации когнитивной архитектуры акта принятия решений* [6, 7]. Сам термин «нейрокогнитивный» в применении к архитектуре интеллектуальной системы управления поведением автономного агента вводится в связи с тем, что в таких функциональных узлах располагаются программные имитационные модели нейронов головного мозга (агенты-нейроны, т.н. *аг-нейроны*), представляющие собой агентов искусственной жизни (Artificial Life Agents), реализующих функционал алгоритмической машины оптимизации многокритериальной целевой функции в своих локальных пространствах состояний. С этой целью они имитируют процессы динамического формирования и деградации «мягких» (контрактных) связей друг с другом в интересах эмерджентного синтеза состояний всей управляющей нейрокогнитивной архитектуры, интерпретируемых в терминах атTRACTоров самоорганизующейся мультиагентной системы, значимых для решения задачи аппроксимации закона управления поведением всего автономного агента в глобальном пространстве состояний системы «агент – среда» [8]. Такое пространство состояний, размеченное значениями целевой функции, получило название «*пространство поведения* интеллектуального автономного агента в реальной среде».

В целом основной функционал инварианта организации акта принятия решений системой управления на основе нейрокогнитивной архитектуры (*инвариант нейрокогнитивной архитектуры*) можно интерпретировать как синтез [8] отрезков траектории движения интеллектуального агента в пространстве поведения в будущем на основе анализа ее отрезков в прошлом и настоящем с целью онтологизации, идентификации и решения проблем универсального спектра, возникающих в системе «агент – среда».

Под *проблемой* агента в реальной среде в данном контексте понимается отрезок траектории движения агента в пространстве поведения, завершающийся скачком значения целевой функции либо его выходом за некоторые пороговые значения. Если при этом значение целевой функции уменьшается, то проблема считается *отрицательной*, если же увеличивается, то – *позитивной*. Если же целевая функция обращается в некоторой точке траектории в ноль, то такая точка считается *терминальной точкой*, а проблема – *терминальной проблемой*.

2. АГЕНТЫ УНИВЕРСАЛЬНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В [8] указывается на то, что проблемы интеллектуального агента определены не в пространстве его поведения, а в своих *локальных редуцированных подпространствах*, оси которых представляют собой множества значений параметров, по которым локальные целевые функции не вырождены в том смысле, что такие значения оказывают влияние на значения этих целевых функций. В соответствии с одной из наших ключевых гипотез свойством динамического формирования таких пространств для каждой онтологизируемой и идентифицируемой проблемы как раз и обладают сети нейронов головного мозга. Разработанная вычислительная абстракция такой сети на базе метафоры проектирования нейрокогнитивной мультиагентной архитектуры нацелена на имитационное моделирование этого свойства.

В соответствии с нашей гипотезой, построив поведение автономного агента на основе такой вычислительной абстракции, мы сможем имитировать процесс рассуждения при решении одной проблемы, однако цели разработки теоретических основ создания

универсальных интеллектуальных программных агентов и роботов, способных к колаборативному выполнению сложных миссий на основе конструктивного взаимодействия (в том числе с помощью естественного языка) в составе человеко-машинного коллектива, предполагают, что системы управления таких агентов и роботов способны синтезировать управление автономным поведением в реальной среде в течение всего времени работы агента в соответствии с функциональным назначением (*время жизни, экзистенции*) агента.

В том числе предполагается, что с целью решения задач функционального назначения, в том числе подзадач, определяемых оператором-партнером по человеко-машинному коллективу с помощью естественного языка, интеллектуальный агент самостоятельно синтезирует свое поведение (включающее также синтез высказываний на естественном языке), реализуя которое он и формирует свою часть коллективной интерпретации и реализации совместной миссии. Автономность имеет принципиальное значение, так как именно ее наличие позволяет интеллектуальному агенту заместить человека в данном функциональном контуре. Она в том числе предполагает, что возмущения, которые в общем случае всегда имеют место в реальной среде, также *самостоятельно* онтологизируются, идентифицируются и решаются интеллектуальным агентом без необходимости непосредственного вмешательства партнеров по человеко-машинному коллективу для модификации закона управления.

Критерии способности к автономному пониманию естественно-языкового описания миссий, их коллективной декомпозиции на подзадачи, формированию согласованного кооперативного плана реализации и автономного решения назначенных коллективом подзадач при условии самостоятельной компенсации возмущений с помощью метаалгоритмов онтологизации проблем и алгоритмов их идентификации и решения, по нашему мнению, задают дифференцирующие признаки агентов *универсального искусственного интеллекта*. Классификация вычислительных агентов, основанная на соотношении их структурно-функциональных свойств с классами задач, которые они способны решать, определяющая в том числе различия между агентами общего и универсального искусственного интеллекта, введена в [8].

В силу вышеизложенного необходимо констатировать, что инвариант нейрокогнитивной архитектуры, реализующий акт принятия решения о синтезе отрезка в траектории движения интеллектуального агента в пространстве его поведения, сам по себе не обладает потенциалом реализации функционала системы управления агентом универсального искусственного интеллекта, он лишь является базовым структурно-функциональным элементом, с помощью которого такая система управления может быть построена. Инвариант нейрокогнитивной архитектуры в этом смысле является метафорой проектирования *динамической функциональной системы* (по П.К Анохину), представляющей собой базовую структурно-функциональную единицу мышления.

Если нейрокогнитивность в указанном выше аспекте относится прежде всего к свойствам самого инварианта, его способности (как динамической функциональной системы) автономно идентифицировать в составе траектории движения агента в пространстве поведения отдельные проблемы и синтезировать их решения, то для описания интегративных свойств системы управления, состоящей из многочисленных разноуровневых, определенным образом взаимосвязанных инвариантов нейрокогнитивной архитектуры, вместе составляющих полнофункциональную систему управления интеллектуального агента, по аналогии с феноменологическими комплексами, реализуемыми мозгом человека, в современном академическом естествознании описываемыми только концептуальным и терминологическим аппаратом нейропсихологии, будем называть *нейропсихологической архитектурой управления интеллектуального агента* (*нейропсихологической архитектурой интеллектуального агента*).

3. НЕЙРОКОГНИТИВНЫЕ ТРАНСФОРМАЦИИ ПРОСТРАНСТВА ОПРЕДЕЛЕНИЯ ПРОБЛЕМ

Как показано в [9], вычислительной абстракцией процессов взаимодействия между группами агнейронов различных типов, располагающихся в различных когнитивных узлах функциональной системы (*когнитонах*), построенной на основе нейрокогнитивной архитектуры, является т.н. *нейрокогнитивное отображение* (*н-отображение*), которое ставит в соответствие множеству выходных сигналов (сообщений) одного когнитона (*точки области определения*), направленных в данный момент дискретного времени группе агнейронов второго когнитона, множество выходных сигналов второго когнитона (*точки области значения*), синтезированных этими агнейронами в ответ (в результате обработки) на присланные сигналы. Такое отображение является т.н. *мультиагентным отображением* и в этом смысле относится к классу *мультиагентных функций*, введенных в [10].

Особенностью н-функций как мультиагентных отображений между агнейронами является то, что корреляции между точками области определения и области значения являются «мягкими» связями, формирование и разрыв которых зависят от фактически наблюдаемой в системе «агент – среда» корреляции между феноменами, функциональную презентацию которых выполняют эти агнейроны в составе данной когнитивной архитектуры [10].

В [8] приведена гипотетическая архитектура афферентного анализатора мозга, которая рассматривается нами в качестве базовой для построения афферентной части нейропсихологической архитектуры интеллектуального программного агента. Показано, в частности, что она состоит из следующих типов агнейронов: *рецепторные, регуляризирующие, топологические (конструкторы квалиа), признаковые (квалиативные), концептуальные, событийные, эмоциональные, моделирующие*. Все эти агнейроны, кроме моделирующих, – конвергентные, т.е. выдают на выход меньшее количество сигналов, чем принимают на вход. Рецепторные агнейроны принимают сигналы от рецепторов агента и преобразуют их в *высказывания*, описывающие *регулярные графические примитивы* (штрихи, углы поворота). Регуляризирующие агнейроны на основе поступивших от рецепторных агнейронов текущих графических примитивов генерируют описания простых *плоских геометрических фигур*. Далее топологические агнейроны формируют описания отношения таких фигур между собой, представляющие собой т.н. *топологические мультиагентные факты (МА-факты)* [3].

МА-факт представляет собой нейрокогнитивную функциональную систему, состоящую из агнейронов, объединенных мягкими (контрактными) связями, позволяющими на основе обмена сообщениями между этими агнейронами выполнять в нейропсихологической архитектуре функциональную презентацию некоторых наблюдавшихся агентом в системе «агент – среда» взаимосвязей между объектами, признаками и явлениями. Мы исходим из гипотезы о том, что любое знание, которое относится к некоторым объектам и явлениям, носит характер МА-факта [11] и в этом смысле является т.н. *мультиагентным знанием* [7].

Сообщения от агнейронов, входящих в состав нейрокогнитивных функциональных систем, реализующих МА-факты, описывающие отношения топологических объектов, далее попадают на входы признаковых агнейронов, синтезирующих на их основе информацию о признаках объектов, наблюдавшихся интеллектуальным агентом во внешней среде и в себе самом и таким образом представленных в афферентном потоке.

На основе этой информации концептуальные агнейроны генерируют функциональные представления таких объектов, отношений и признаков, а событийные агнейроны по этим данным, в свою очередь, формируют МА-факты, описывающие статические и динамические процессы в системе «агент – среда».

Следует отметить, что по мере прохождения через вышеописанные когнитоны, используя свойство н-функций, состоящее в том, что они определены только в пространствах, оси которых не вырождены по значениям этих функций [12], вышеописанная, первая, афферентная

часть инварианта нейрокогнитивной архитектуры последовательно переопределяет наблюдаемую часть исходного афферентного потока, порождаемого рецепторами интеллектуального агента, в подпространствах с последовательно уменьшающимися размерностями. Если первоначально, сразу после срабатывания рецепторов, некоторая часть траектории интеллектуального агента определена непосредственно в пространстве его поведения, размерность которого равна сумме общего количества рецепторов и эффекторов этого агента, то, проходя через афферентную часть инварианта нейрокогнитивной архитектуры и таким образом переопределяясь на каждом шаге в подпространствах, редуцированных за счет выбывания осей, по которым вырождены н-функции, задействованные в обеспечении таких трансформаций, эта часть траектории оказывается определена в очень компактном подпространстве, оси которого соответствуют лишь агнейронам, задействованным в описании МА-фактов, наблюдавшихся в текущей ситуации в системе «агент – среда».

Следует также отметить, что все вышеупомянутые пространства и подпространства типологически относятся именно к *мультиагентным пространствам (МА-пространства)*, представляющим собой упорядоченные наборы множеств сообщений, синтезируемых агентами, каждый из которых соответствует (выполняет презентацию) одной из осей такого пространства.

С точки зрения представления информации о событиях, происходящих в наблюдаемой интеллектуальным агентом части системы «агент – среда», сформированный набор МА-фактов полностью описывает текущее состояние этой системы. В этом смысле вышеописанная афферентная часть инварианта нейрокогнитивной архитектуры реализует *функцию синтеза состояний* всей системы «агент – среда». Однако важно, что состояния эти описаны в терминах МА-фактов и поэтому определены в соответствующих МА-пространствах очень низкой размерности. Тем самым достигается значительная степень редукции вычислительной сложности всей задачи синтеза траектории агента в пространстве его поведения.

На следующем этапе необходимо оценить события, выделенные событийными агнейронами из афферентного потока, по отношению к динамике целевой функции агента. Для этого используется н-отображение из когнитона событийных агнейронов в когнитон эмоциональных агнейронов. Такие агнейроны используют предопределенные знания и знания, накопленные в опыте интеллектуального агента, для оценки текущих и возможных потерь или приобретения *энергии*, рассматриваемой как обобщенная величина, описывающая значения целевой функции интеллектуального агента. *Эмоциональные агнейроны* реализуют *функцию разметки состояний* системы «агент – среда» значениями целевой функции.

После такой разметки выполняется дивергентное н-отображение, формирующее представление о причинах и возможных последствиях текущих наступивших событий. Это н-отображение выполняется т.н. *моделирующими агнейронами*, находящимися в одноименном когнитоне нейрокогнитивной архитектуры. Эволюционный синтез возможности экстраполировать события прошлого и будущего и объединять их в одну *проблемную ситуацию (ситуацию)*, вероятно, был востребован в связи с необходимостью компенсации таких свойств системы «агент – среда», как эпизодичность, частичная наблюдаемость, динамичность, активность и стохастичность.

Само представление о времени, возникающее в нейропсихологической архитектуре интеллектуального агента, базируется на нейроморфологических процессах, протекающих в инварианте когнитивной архитектуры. Можно предположить, что *настоящее* – это период времени, в течение которого в афферентной части инварианта нейрокогнитивной архитектуры происходит построение МА-фактов, описывающих текущие события.

Представляется также вероятным, что, восстанавливая прошлые события, моделирующие агнейроны также редуцируют локальное пространство определения данной проблемы, так как они тем самым исключают из рассмотрения все другие возможные «сценарии прошлого».

Эфферентная часть инварианта нейрокогнитивной архитектуры включает в себя агнейроны: *целеполагания* (*целевые агнейроны*), *выбора действия, потенциалиса, композиции планов, управления режимами работы* инварианта [7].

Целевые агнейроны предназначены для поиска и фиксации завершающего (целевого) события (цели), наступление которого в будущем соответствует достижению решения проблемы. Подобно тому, как каждый сформированный агнейронами нейрокогнитивной функциональной системы МА-факт редуцирует пространство определения проблемы, целевые агнейроны совместно с агнейронами, выполняющими функциональную репрезентацию целевых событий, ограничивают пространство альтернатив решения проблемы. В этом смысле можно говорить о том, что подцели дерева целей, формируемого инвариантом нейрокогнитивной архитектуры для сложных многоходовых решений, соответственно, также последовательно редуцируют это пространство, на каждом шаге определяя в локальном пространстве проблемы подпространства все меньшей и меньшей размерности, которые задают множества возможных траекторий агента при движении по дереву (плану) решения. В результате выбора конкретной подцели поиск переходит в соответствующий условный «гиперквадрант» локального пространства проблемы, значительно редуцируя пространство дальнейшего поиска. Определив подцель, мы также на порядки снижаем размерность пространства альтернатив всего решения, так как укорачиваем горизонт планирования, выводим из рассмотрения лишние МА-оси.

Агнейроны выбора действия совместно с целевыми, эмоциональными и событийными агнейронами реализуют н-функцию, выполняющую поиск действия, позволяющего перейти из текущей ситуации в целевую. Если функциональная репрезентация таких действий не находится либо агнейроны потенциалиса определяют, что найденное действие не представляется возможным осуществить, то агнейроны композиции планов организуют с помощью обмена запросами в формате сообщений поиск промежуточных действий по схемам прямого, обратного либо смешанного планирования.

Агнейроны управления режимами работы инварианта нейрокогнитивной архитектуры принимают решение о переключении между различными типами поведения интеллектуального агента – синтезом плана решения проблемы, выполнением такого плана.

Следует отметить, что при синтезе плана поведения и его реализации происходит процесс, обратный нейрокогнитивной редукции, – это процесс *нейрокогнитивной экспедиции*. В ходе такого процесса решение, найденное в локальном пространстве определения проблемы, реализуется в системе «агент – среда», т.е. в полноразмерном пространстве поведения интеллектуального агента.

Таким образом, функциональные системы, построенные на основе инварианта нейрокогнитивной архитектуры, оснащены необходимой базой для построения мультиагентных алгоритмов идентификации и решения проблем на основе синтеза траекторий движения интеллектуального агента в пространстве его поведения с использованием способности н-отображений к трансформации пространства определения проблемы и ее решения в динамическом режиме.

4. ГИПОТЕЗА О НЕЙРОПСИХОЛОГИЧЕСКОЙ АРХИТЕКТУРЕ МОЗГА

Следует, однако, отметить, что интеллектуальный агент в реальной среде подвержен воздействию значительного количества проблем, многие из которых проявляются синхронно. Такая ситуация возможна прежде всего в силу того, что нейрокогнитивная система, создающаяся для идентификации и решения каждой проблемы, определяет ее в некотором редуцированном локальном пространстве. Зачастую такие пространства даже не

имеют общих осей. С учетом экзистенциальной значимости, которая по определению свойственна проблемам, пренебрежение даже теми из них, которые размечены незначительными потерями энергии, для агента представляется проигрышной стратегией. Поэтому эволюция синтезировала более сложную *нейропсихологическую архитектуру мозга*, ориентированную на управление многочисленными функциональными системами, основанными на инварианте нейрокогнитивной архитектуры, входящими в ее состав. Рассмотрим гипотетическую структурно-функциональную схему такой нейропсихологической архитектуры (рис. 1).

Так как в соответствии с нашей рабочей гипотезой основной задачей интеллектуальной системы управления интеллектуального агента является синтез его траектории в пространстве поведения, субоптимальной по критерию максимизации целевой функции, нейропсихологическая архитектура мозга, являющаяся структурно-функциональной основой такой системы управления, должна обеспечивать не только онтологизацию, идентификацию и решение отдельных проблем, с которыми сталкивается агент, но и их субоптимальный выбор из входного потока проблем.

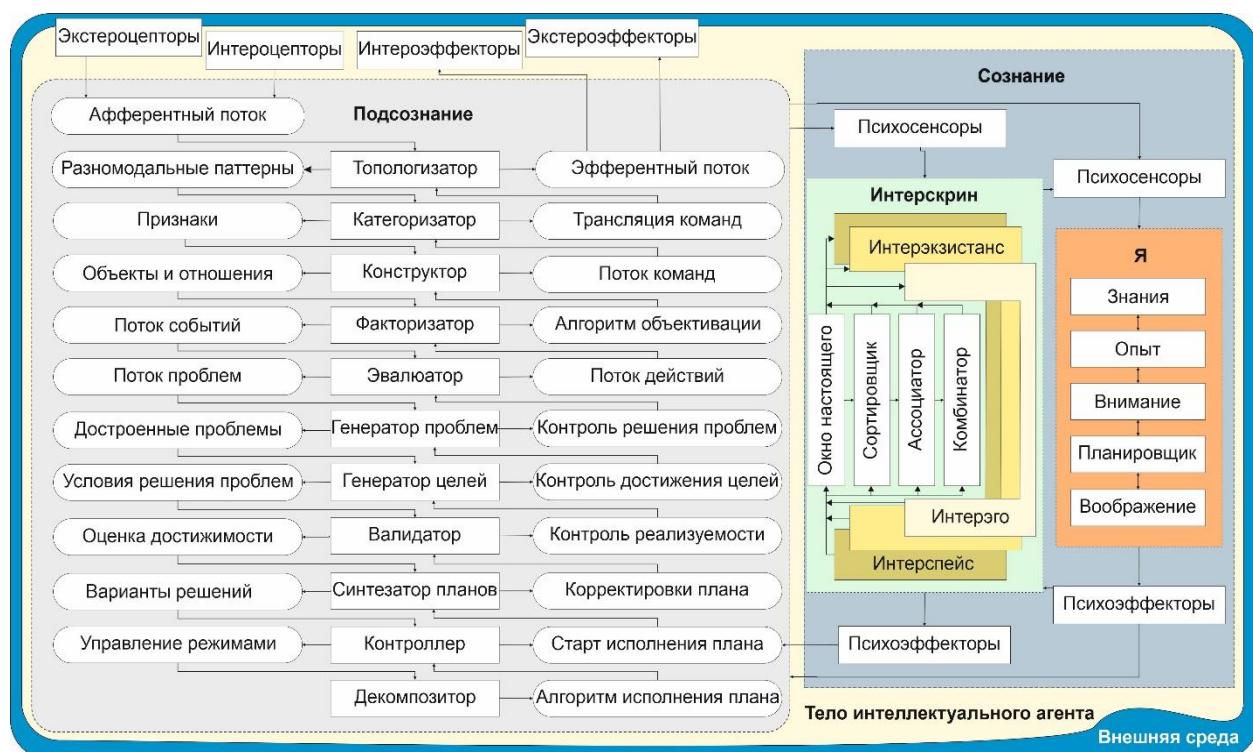


Рис. 1. Гипотетическая нейропсихологическая архитектура мозга

Fig. 1. Hypothetical neuropsychological architecture of the brain

В силу того, что проблемы делятся (развертываются) в течение определенного ненулевого времени и, как правило, состоят из множества синхронных и асинхронных событий, агент вынужден одновременно рассматривать возможность решения нескольких из них. Так как процессы идентификации и решения проблем выполняются нейрокогнитивными функциональными системами, состоящими из множеств взаимодействующих агнейронов, интеллектуальный агент в принципе способен выполнять такие процессы в параллельном режиме. Этому способствует также и то, что каждая синхронно решаемая проблема в силу действия динамической нейрокогнитивной редукции определена в своем собственном ло-

кальном мультиагентном нейрокогнитивном пространстве. В указанном смысле агенты, входящие в состав нейрокогнитивных функциональных систем, можно отнести к *немонопольным ресурсам*, используемым агентом для решения проблем.

Ситуация, однако, осложняется тем, что выполнить найденные алгоритмы решения синхронных проблем агент в общем случае одновременно не может, так как для такой реализации агент должен выполнить движение по некоторому отрезку своей траектории в полноразмерном пространстве поведения, а для этого ему необходимо использовать свои сенсоры и эффекторы, которые, также в общем случае, являются *монопольными ресурсами*, т.е. не могут одновременно использоваться для решения нескольких проблем. Иными словами, при реализации движения по траектории в пространстве поведения агент вынужден перейти от нейрокогнитивного моделирования «внутри» системы управления к непосредственному решению проблемы во внешней для него реальной среде. При этом он должен каким-то образом распределить точки (отрезки) своей траектории в пространстве поведения между частями нейрокогнитивных алгоритмов решения различных актуальных синхронных проблем, определенных в своих локальных мультиагентных нейрокогнитивных пространствах. Очевидно также, что для того, чтобы заполнить часть своей траектории в пространстве поведения фрагментами таких нейрокогнитивных алгоритмов, т.е. для того, чтобы реализовать эти фрагменты в системе «агент – среда», агенту необходимо переопределить соответствующие им решения из локальных нейрокогнитивных пространств, в которых они определены, в полноразмерное пространство поведения. Такое переопределение, как уже указывалось выше, выполняется с помощью нейрокогнитивной экспрессии.

Можно предположить, что интеллектуальный агент под управлением нейропсихологической архитектуры в реальной среде строит свою траекторию в пространстве поведения как непрерывающуюся композицию фрагментов траекторий онтологизации, идентификации, решения отдельных проблем, построенных в локальных нейрокогнитивных пространствах определения этих проблем и приведенных с помощью нейрокогнитивной экспрессии к размерности пространства поведения агента (переопределенных в нем).

Исходя из этой гипотезы можно заключить, что одним из ключевых отличительных свойств нейронных систем (систем, состоящих из взаимодействующих нейронов) как таких является их способность к избирательной трансформации пространства определения проблем, основанной на учете факторов системы «агент – среда», существенных для решения этих проблем. Иными словами, вероятно, корректно было бы применять термин «нейронный» в отношении тех метафор проектирования, вычислительных абстракций, формализмов и систем, которые обладают таким свойством. Например, если мы хотим подчеркнуть, что система способна самостоятельно строить функциональные зависимости в стиле мультиагентных отображений [7, 10], то по отношению к ней применим термин «мультиагентная» система. Если же при этом построенные отображения еще и избирательно содержательны в вышеуказанном смысле, то назовем их «мультиагентными нейрокогнитивными (нейронными)».

Возвращаясь к исследованию возможности построения системного обоснования метафоры проектирования нейропсихологической архитектуры мозга, отметим, что мы придерживаемся эволюционной концепции системогенеза мозга, обосновывающей формирование его новых структурно-функциональных компонентов детерминантами усложнения организмов, их поведения и взаимодействия со средой обитания. Такой подход вполне согласуется с наблюдаемой морфологией и поведенческими способностями видов в филогенетическом ряду и с общепринятыми естественнонаучными представле-

ниями об эволюционном формировании отделов мозга. Соответственно, для обеспечения нового функционала управления многозадачным решением синхронных и асинхронных проблем, идентифицируемых интеллектуальным агентом в системе «агент – среда», мы добавляем в нейропсихологическую архитектуру этого агента новые структурно-функциональные компоненты, представленные на рисунке 1.

Так, рассмотренные ранее структурно-функциональные подсистемы объединены в составе модуля *подсознания*, отвечающего в нейропсихологической архитектуре за процессы синтеза и применения нейрокогнитивных динамических функциональных систем онтологизации, идентификации и решения отдельных проблем. В то же время в нейропсихологическую архитектуру дополнительно введен модуль *сознания*, основной задачей которого является управление синтезом единой траектории движения агента в пространстве поведения на основе переключения монопольных ресурсов агента между нейрокогнитивными функциональными системами подсознания с помощью механизма *внимания*.

В модуле сознания, в свою очередь, выделены компоненты *интерскрина* и *подсистемы «Я»* (модуль «Я») агента. В [8] показано, что все вышеупомянутые компоненты построены на основе нейрокогнитивных функциональных систем, аналогичных по основным принципам организации, структуре и составу вышеуказанному инварианту нейрокогнитивной архитектуры, используемому нами в качестве основной структурно-функциональной единицы подсознания. Разница, однако, состоит в том, что каскад н-функций, последовательно реализуемых инвариантом нейрокогнитивной архитектуры, в случае использования в функциональных системах подсознания выполняет синтез фрагментов траекторий движения агента в пространстве поведения, тогда как в случае сознания инвариант нейрокогнитивной архитектуры управляет синтезом фрагментов траектории агента в нейрокогнитивном *пространстве проблем*, которое мы определяем как множество всех метаалгоритмов и алгоритмов онтологизации, идентификации и решения проблем, представленных в подсознании агента.

Будем считать прямое произведение пространства поведения на пространство проблем *пространством психики* интеллектуального агента.

Если подсознание для анализа процессов в пространстве поведения и управления ими использует рецепторы и эффекторы (в том числе интерорецепторы и экстерорецепторы, интероэффекторы и экстероэффекторы), то сознание – т.н. *психосенсоры* и *психоэффекторы*, с помощью которых оно анализирует процессы в пространстве проблем и, соответственно, управляет ими (рис. 1).

В модуле подсознания на этом рисунке в прямоугольниках показаны структурно-функциональные компоненты, соответствующие вышеописанным когнитонам. Однако в отличие от названий когнитонов на схеме приведены названия таких компонентов, мотивированные их обобщенным функциональным назначением. Кроме того, состав таких компонентов расширен за счет ранее не упоминавшихся субкогнитонов, расположенных внутри когнитона распознавания, назначение которых состоит в организации мультиагентного нейрокогнитивного решения задачи распознавания образов по данным разномодальных потоков сенсорных данных (*топологизатор*, *категоризатор*, *конструктор*). В овалах на рисунке 1 показаны категории, в качестве которых интерпретируются сообщения от соответствующих структурно-функциональных компонентов нейрокогнитивной архитектуры подсознания, при синтезе потоков афферентной (слева) и (эфферентной) информации.

Компонент «Я» как часть сознания, по нашему мнению, выполняет основную функцию выбора конкретной части траектории идентификации или решения наиболее актуальной проблемы для ее использования в процессе синтеза единой траектории движения интеллектуального агента в пространстве его поведения.

Компонент «Интерскрин» выполняет основную функцию трансформации пространства определения проблемы или ее решения между локальными подпространствами их определения, пространством поведения и пространством психики интеллектуального агента.

Каждый из компонентов нижнего уровня, входящих в компоненты «Я» и «Интерскрин» (рис. 1), также представляет собой множество нейрокогнитивных функциональных систем.

Компонент «Знания» содержит мультиагентные знания интеллектуального агента, характеризующие его общие представления об устройстве и функционировании мира (себя самого, других агентов и окружающей среды).

Компонент «Опыт» содержит мультиагентные знания, описывающие ситуации непосредственного взаимодействия агента с внешней средой и другими агентами в ее составе, свой собственный субъективный опыт.

Компонент «Воображение» содержит нейрокогнитивные функциональные системы, способные для каждого шага идентификации и решения проблемы синтезировать пространство альтернатив траекторий дальнейшего движения агента.

Компонент «Внимание» позволяет осуществлять выбор наиболее важных фрагментов синхронно решаемых проблем и достройку полного представления этих проблем в их локальных пространствах определения.

Компонент «Планировщик» позволяет выполнить субоптимальный выбор последовательности и длительности сеансов использования монопольных ресурсов интеллектуального агента при переключении между наиболее актуальными проблемами.

Структурно-функциональные подсистемы компонента «Интерскрин», приведенные на рисунке 1, имеют следующее назначение:

- Интерспейс – нейрокогнитивная трансформация, достройка и интеррепрезентация (между подсознанием и сознанием) компонентов проблем, наблюдавшихся во внешней среде;
- Интерэго – нейрокогнитивная трансформация, достройка и интеррепрезентация компонентов проблем, наблюдавшихся в самом агенте;
- Окно настоящего – формирование и интеррепрезентация МА-фактов, наблюдавшихся в системе «агент-среда»;
- Сортировщик – отнесение МА-фактов, сформированных модулем «Окно настоящего», к одной из актуальных (нерешенных) проблем интеллектуального агента;
- Ассоциатор – актуализация мультиагентных связей текущих нейрокогнитивных функциональных репрезентаций событий и проблем с ассоциированными событиями и проблемами с целью наиболее полного описания состояния системы «агент – среда»;
- Комбинатор – система интеррепрезентации комбинации фрагментов различных проблем и решений, способная трансформировать пространство их определения с учетом всех нейрокогнитивных осей, задействованных в описаниях отдельных проблем.

По нашему мнению, взаимодействие между представленными компонентами нейропсихологической архитектуры интеллектуального агента строится на тех же принципах обмена информацией и энергией, что и взаимодействие между агнейронами в составе инварианта нейрокогнитивной архитектуры.

Так, например, компонент сознания может использовать психоэффекторы для воздействия на подсознание с целью воспоминания некоторой ситуации. В этом случае агнейроны соответствующей нейрокогнитивной функциональной системы сознания предлагают вознаграждение в виде порции энергии управляющему агнейрону или моделирующему агнейрону в составе инварианта нейрокогнитивной архитектуры подсознания с тем, чтобы они, в свою очередь, предложили вознаграждение другим участникам функциональных систем подсознания, чтобы отправить соответствующие сообщения в интерскрин. В этом случае психорецепторы считали бы всю проблемную ситуацию (все дерево проблем

и решений) из локальной области определения этой проблемы и нейрокогнитивные функциональные системы интерскрина достроили бы ее до размерности психического пространства, в котором проблемы представляются в интерспейсе и где они становятся доступными для анализа нейрокогнитивным функциональным системам компонента «Я».

Основной задачей компонента сознания является выбор для решения самой актуальной проблемы. Важно, что события, определяемые подсознанием и демонстрируемые в окне настоящего в интерскрине, постоянно сменяют друг друга и при этом, в общем случае, относятся к совершенно разным проблемам. Поэтому требуется, чтобы внимание и интерспейс, распределяя такие события по проблемам, в случае возникновения необходимости могли бы быстро загрузить для демонстрации и обработки компонентом «Я» всю проблему целиком, включая те ее фрагменты, которые относятся к прошлому и воображаемому будущему. При этом эту проблему также необходимо переопределить в пространстве психики для того, чтобы при синтезе решения не упустить значимые детали.

Достройка размерностей выполняется с учетом того, что нейрокогнитивные функциональные системы компонента «Я» адаптированы под эффективную обработку интеррепрезентации траектории движения агента в пространстве психики, объединяющем информацию как о его траектории в пространстве проблем, так и о его траектории в пространстве поведения. Такой подход представляется конструктивным для решения т.н. *легкой и трудной проблем сознания* [13].

Разработанная концепция нейропсихологической архитектуры системы управления интеллектуального агента позволяет вернуться к метафоре проектирования т.н. картезианского театра, удобной для описания сложного взаимодействия компонентов психики [13].

5. КАРТЕЗИАНСКИЙ ТЕАТР: ПОДСОЗНАНИЕ, «Я» И ИНТЕРСКРИН

В соответствии с предложенной концепцией нейрокогнитивной трансформации пространства определения проблемы онтологизируются в редуцированных локальных подпространствах. Поэтому у интеллектуального агента появляется возможность решать их с помощью различных нейрокогнитивных функциональных систем в последовательно-параллельном многозадачном стиле, организуя управление переключением между локальными подпространствами с помощью компонента сознания. Единая траектория движения интеллектуального агента в полном пространстве поведения, таким образом, формируется из фрагментов траекторий в локальных подпространствах проблем, переключением между которыми управляет сознание.

Так как функция переключения при такой конфигурации с необходимостью отдается сознанию, выбор из вариантов планов решения задачи также морфофизиологически передается на этот уровень, что в нашей модели реализуется в компоненте «Я».

Соответственно, функция интерскрина состоит в полной достройке проблем и их демонстрации для моделирования и анализа последствий и вариантов решения с тем, чтобы все задействованные компоненты сознания были обеспечены всей необходимой информацией, требуемой для осуществления выбора, представленной в удобном для него формате фрагмента траектории агента в пространстве психики.

В этом смысле интерскрин можно, по нашему мнению, рассматривать как компонент психики, реализующий функционал т.н. *картизанского театра* [14], в котором сознанию демонстрируются проблемы, приведенные к формату фрагментов траекторий в пространстве психики. Такая подача позволяет упростить реализацию основной функции сознания, состоящей в управлении разделением задач, определенных в локальных пространствах представления проблем, при движении целостного (интегрированного самой структурой нервной системы) агента в полном пространстве его поведения с учетом монопольного характера части задействованных ресурсов.

Нейрокогнитивные функциональные системы онтологизации, идентификации и решения проблем, расположенные в компоненте подсознания, обрабатывая проблемы, синтезируют точки траектории движения интеллектуального агента в локальных пространствах этих проблем. В общем случае эти процессы происходят параллельно, независимо и синхронно по отношению друг к другу. Интерскрин демонстрирует сознанию эти проблемы под управлением функциональной системы внимания. Сознание анализирует построенные фрагменты траекторий и предлагает модулю «Я» варианты интеграции их в единую траекторию движения агента в пространстве поведения в соответствии с расписанием, построенным Планировщиком с учетом приоритетности решения проблем.

Модуль «Я» принимает окончательное решение по выбору фрагмента траектории локальной проблемы для встраивания в главную траекторию. Сознание и интерскрин обеспечивают интеграцию. Интерскрин при этом обеспечивает трансформацию пространств определения проблемы – между локальным пространством определения проблем, глобальным пространством поведения и пространством психики интеллектуального агента, управляя при этом вырожденными осями. Сознание перебирает варианты будущих траекторий, строит последовательности и комбинации их фрагментов в едином формате траектории в пространстве поведения. Следовательно, можно предположить, что «пьесы» в картезианском театре «ставятся» и «разыгрываются» именно с целью оптимального решения задачи синтеза единой, но «кусочной» по своему характеру (по составу задач) траектории в пространстве поведения агента.

6. «ДЕРЕВО ЖИЗНИ» В ПРОСТРАНСТВЕ ПОВЕДЕНИЯ АГЕНТА

Все психоэфекторы предназначены (их состав и функции детерминированы задачами) для редактирования графов проблем и решений, реализуемых отдельными нейрокогнитивными функциональными системами, – добавить, удалить, верить, считать невыполнимым, невероятным, рекомбинировать с частями других графов и т.д. Важны центральные функции воспоминания и представления: вспомнить означает по некоторому частному параметру (по координате, оси, точке траектории в локальном пространстве проблемы) найти и загрузить в интерспейс всю локальную траекторию задачи (еще и с вариантами возможных альтернатив) для того, чтобы исследовать ее: найти причинно-следственные детерминанты, соотнести с другими вариантами проблем и решений (композиция и рекомбинация).

Так как локальных деревьев синхронных проблем (в локальных подпространствах), как правило, много и интеллектуальный агент, перемещаясь по своей траектории в пространстве поведения с помощью монопольных ресурсов и эффекторов, использует на каждом шаге немонопольные нейрокогнитивные функциональные системы решения таких проблем в параллельном режиме, получается, что каждое реальное состояние (каждая точка в пространстве поведения), идентифицируемое агентом, сопровождается идентификацией и загрузкой в интерскрин некоторого количества воображаемых проблем, актуализирующихся в связи с переходом агента в данное состояние.

Такое движение можно образно представить в виде дерева, ствол которого реален и представлен в пространстве поведения, а ветви динамически возникают и исчезают в воображении по мере продвижения агента вверх по этому дереву. Переходя на следующую вершину дерева (точку траектории в реальном пространстве поведения), с помощью подсистемы окна настоящего агент фиксирует некоторый новый набор актуальных проблем в локальных пространствах определения, что позволяет интерскрину переопределить эти проблемы в пространстве психики и таким образом «вырастить» их из определенных точек ствола с тем, чтобы затем предъявить эту визуализацию компоненту «Я». Такое дерево является удобной метафорой представления процесса движения агента в

пространстве его поведения, развертывания его экзистенции (жизни) в реальной среде, поэтому для удобства будем также называть его *деревом жизни агента*. Отметим, что жизнь агента развертывается в процессе его движения по траектории в пространстве поведения. Следовательно, продолжая метафорический ряд, можно отметить, что дерево жизни агента развивается («растет») в пространстве его поведения.

Таким образом, компонент «Окно настоящего» служит для обновления нейрокогнитивных функциональных систем проблем и решений – локальных веток глобального дерева жизни – путем идентификации текущих событий и их распределения между такими ветками. В процессе обновления на основе ретропроактивного анализа, выполняемого моделирующими агнейронами, реализуется сравнение событий, ожидаемых в соответствии с известным составом (паттерном) проблемы, и фактически идентифицируемыми событиями, что дает дополнительную информацию компоненту сознания для принятия решений о выборе фрагментов траекторий в локальных пространствах определения проблем для их встраивания в ствол дерева жизни агента.

В частности, на основе сопоставления МА-фактов из окна настоящего с паттернами проблем может быть выполнена корректировка самих таких паттернов, в том числе путем обнаружения и устранения противоречий в текущих МА-фактах событий, логических посылок, причинно-следственных зависимостей, прогнозов и выводов.

Как следует из вышеприведенного описания и рисунка 2, фактическая траектория у агента (ствол дерева жизни) одна, но на каждом ее шаге компоненту «Я» доступно для выбора ее продолжения много вариантов. При переходе в очередную вершину дерева жизни специализированные нейрокогнитивные функциональные системы выполняют синтез доступных из этой вершины вариантов сценариев дальнейшего движения (потенциальных направлений роста) – локальных ответвлений ствола дерева жизни. Формирующаяся таким образом динамическая структура дерева определяет возможности и порядок движения между локальными ответвлениями. Переместиться между любыми произвольными вершинами дерева в реальности агент в общем случае не может. Такое перемещение, однако, вполне осуществимо сознанием, не ограниченным физическими свойствами реального мира.

Соответственно, при переключении между фрагментами траекторий идентификации или решения конкретных проблем в локальных пространствах их определения агенту необходимо синтезировать также и траектории перехода между конечными и начальными точками таких фрагментов, так как в реальности перейти между ними за один шаг агент не может.

На рисунке 2 приведено упрощенное изображение дерева жизни агента в пространстве размерности 3. Проблемы, идентифицируемые в вершинах дерева, показаны прозрачными параллелепипедами, символически отражающими тот факт, что такие проблемы определены в своих локальных нейрокогнитивных пространствах. Проблемы, относящиеся к вершинам, сформированным в прошлом и настоящем, показаны в параллелепипедах со сплошным контуром, а к вершинам, синтезированным воображением для описания возможных траекторий движения агента в будущем, – с пунктирным контуром. На рисунке изображены траектории решения трех проблем, из фрагментов решений которых сознание строит основную траекторию движения, которая показана сплошной линией. На рисунке можно, в частности, увидеть, что для переключения между такими фрагментами нейропсихологической архитектуре необходимо также синтезировать траектории переходов между конечными и начальными точками таких фрагментов для различных локальных проблем.

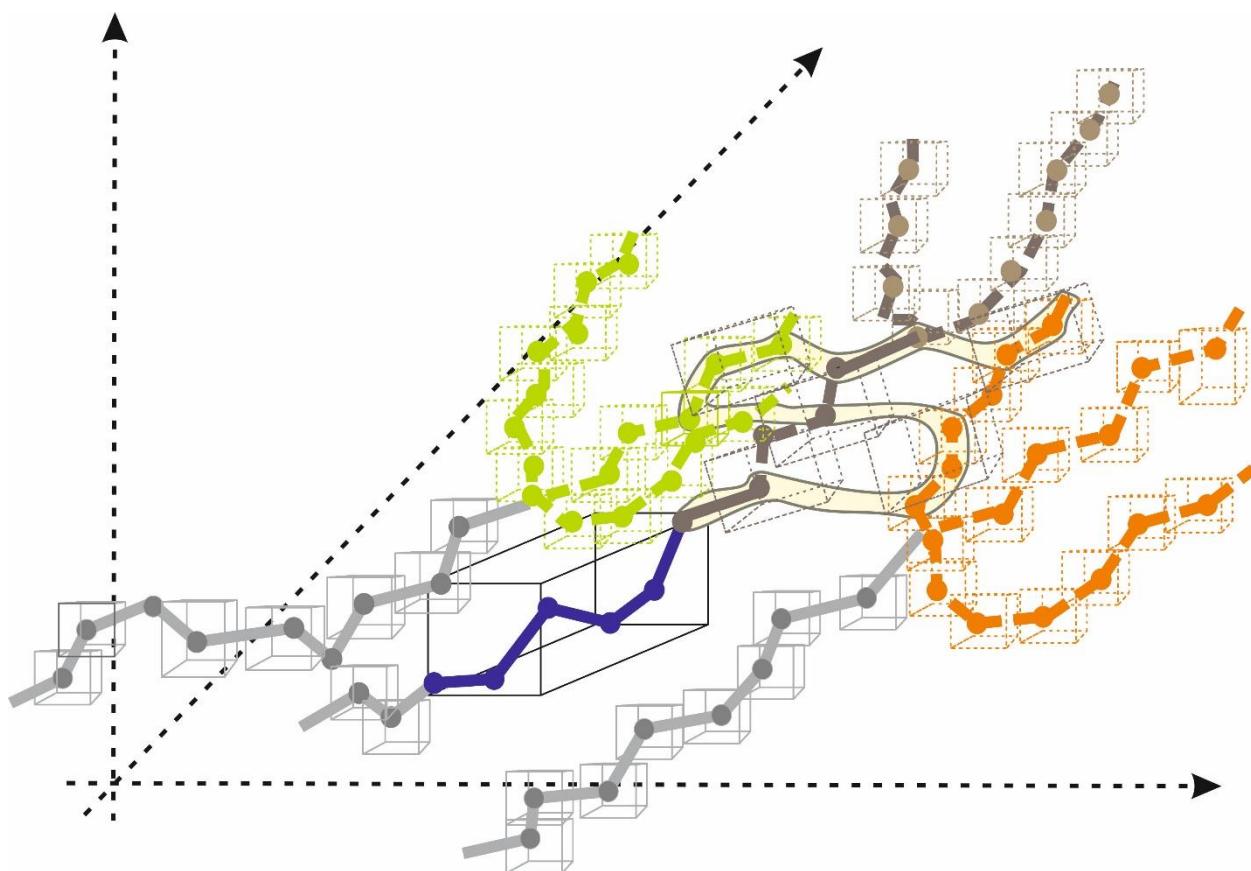


Рис. 2. Схематическое изображение движения агента по дереву жизни

Fig. 2. Schematic representation of the agent's movement along the tree of life

Что же касается нейрокогнитивной архитектуры компонентов сознания, то возможность демонстрации в интерскрине (в единой «реальности») нескольких синхронно обрабатываемых проблем является принципиально важной для обеспечения творческого мышления при их онтологизации, идентификации и решении, так как только в этом случае у нейropsихологической архитектуры системы управления поведением агента появляется возможность комбинировать фрагменты траекторий таких проблем и решений в едином пространстве психики с целью синтеза более эффективных субоптимальных траекторий в пространстве поведения.

7. ВыЧИСЛИТЕЛЬНАЯ АБСТРАКЦИЯ ИЛЛЮЗИИ ЦЕЛОСТНОСТИ «Я»

В работе [8] показано, что нейрокогнитивным компонентам сознания структурно-топологически доступны для восприятия все уровни афферентных анализаторов и все информационные объекты, формируемые агнейронами на этих уровнях, – начиная непосредственно с выходов рецепторов и заканчивая событиями и проблемами. В соответствии с нашей гипотезой это необходимо для того, чтобы сформировать репрезентацию событий и проблем в системе «агент – среда» в пространстве психики, что, в свою очередь, является неотъемлемым условием объединения в таких репрезентациях событий, непосредственно наблюдаемых в окне настоящего и достраиваемых моделирующими агнейронами в процессе загрузки прошлых и будущих воображаемых частей проблемы в интерскрин. Иными словами, в картезианском театре интерскрин должен представить компоненту «Я» текущие и представляемые проблемы в том же формате, в котором нейрокогнитивные компоненты сознания наблюдают

события в составе этих проблем в потоках выходов афферентных агнейронов верхних уровней с тем, чтобы возник эффект единообразного представления при синхронной актуализации модели реальности в пространстве психики на базе действительных и воображаемых событий. Для конструирования такой модели необходимо, чтобы на всех уровнях афферентного тракта существовали агнейроны, которые могли бы достраивать входы этих уровней по запросу, содержащему значения их выходов. Таким образом, при необходимости с использованием таких агнейронов возможно реализовать дивергентную трансформацию пространства определения проблемы – из локального – в пространство психики.

Можно предположить, что динамическое комплексирование сообщений с разных уровней описания проблемы с помощью такого рода нейрокогнитивных трансформаций порождает ощущение целостности и осмыслинности восприятия, составляющее основу представления интеллектуального агента о себе и об окружающем мире.

Итак, модуль первого уровня обработки информации – подсознание – формирует нейрокогнитивные функциональные системы онтологизации, идентификации и решения отдельных проблем в локальных пространствах их определения.

Модуль второго уровня – сознание – определяет субоптимальный план распределения монопольных ресурсов агента между такими нейрокогнитивными функциональными системами.

Внутри модуля второго уровня выделяется компонент интерскрин, формирующий общий и текущий контекст обработки множества актуальных синхронных и асинхронных проблем в пространстве психики.

Кроме того, в модуле второго уровня выделяется также компонент «Я», выполняющий субоптимальный выбор фрагментов траекторий обработки проблем в их локальных пространствах определения с целью их интеграции в единую траекторию агента в пространстве его поведения.

Как показано в [8], неотъемлемым компонентом интерскрина является подсистема т.н. *интерэксистанса*, отвечающая за функциональную интеррепрезентацию процессов, регистрируемых подсознанием в *теле* агента. В свою очередь подсистема *интерэго* регистрирует и выполняет нейрокогнитивную функциональную репрезентацию процессов, протекающих в интерэксистансе и в компоненте «Я». Таким образом, модуль интерскрина всегда (на протяжении всей экзистенции агента) выполняет функциональную интеррепрезентацию подсознательного восприятия окружающего мира, тела агента в этом мире и себя в этом мире и в этом теле.

Операционная модель, в которой именно модуль «Я» принимает окончательное решение о выборе конкретного участка траектории агента в пространстве поведения, в такой трехуровневой модели *интеррепрезентации (самовосприятия)* наделяет этот модуль функцией субъекта экзистенции, от лица которого совершаются действия агента в среде.

Таким образом, в субъективной части такой интеррепрезентации (интерэксистанс и интерэго) формируется интроспективное представление о «Я» как о субъекте агентности, дополнительно оснащенном физическим телом. Структурно-функциональная схема взаимного подключения к потокам генерируемых сигналов, приведенная в [8], когда интерспейс интегрирован в потоки сообщений между экстероцепторами и конструктором, интерэксистанс – между инteroцепторами и конструктором, интерэго – между психоцепторами и конструктором, поддерживает такое представление, составляя топологическую основу целостности самовосприятия агента.

Принципиальное значение имеет то, что такая система интеррепрезентации имеет смешанный генезис – топология и базовая нейрокогнитивная архитектура, позволяющая динамически

формировать функциональные системы в формате инварианта, представлены в ней изначально, в момент старта экзистенции агента, т.е. имеют филогенетический характер. Однако в процессе развертывания экзистенции агента в среде новые нейрокогнитивные функциональные системы генерируются во всех компонентах интерскрина динамически в режиме «по требованию» с использованием т.н. *нейрокогнитической функции* [8] и строят ассоциативные связи друг с другом с использованием алгоритмов *онтонейроморфогенеза* [15], т.е. имеют онтогенетическое происхождение. Таким образом, все компоненты нервной системы агента достраиваются в процессе его жизни.

Такой смешанный, *онтоФилогенетический* [15, 16], подход к формированию системы управления агентом составляет основу ее адаптивности и обеспечивает основной процесс, в который, по определению [8], вовлечен агент универсального искусственного интеллекта – *онтологизация, идентификация и решение проблем универсального спектра на основе автономного синтеза поведения в реальной среде*.

Как показано выше, в этом процессе фактическая траектория движения агента в пространстве его поведения строится динамически на основе фрагментов траекторий в локальных редуцированных пространствах определения конкретных проблем, ассоциированных с вершинами дерева жизни агента, задающем схему допустимых взаимных комбинаций точек глобальной траектории агента в пространстве поведения, связанных с ограничениями на использование монопольных ресурсов агента.

Компонент сознания предназначен для управления ресурсами с целью субоптимального переключения между фрагментами локальных траекторий. Поэтому интерспейс демонстрирует сознанию различные комбинации фрагментов и их вариантов. Взаимосвязанные компоненты интерскрина, встроенные в афферентные потоки, позволяют делать это именно в том формате, в котором такие фрагменты, определенные в локальных пространствах проблем, могут быть дивергентно переопределены (на основе нейрокогнитивной экспрессии) в пространстве психики, в котором они уже могут комбинироваться с событиями, непосредственно детерминированными текущей перцептивной реальностью.

Афферентные потоки изначально структурируют хаос реального мира в перцептивные форматы, на основе которых строятся редукции в виде топологий, признаков, объектов, событий, проблем, а интерскрин как совокупность взаимосвязанных нейрокогнитивных функциональных систем интерспейса, интерэкзистанса и интерэго (интерэго внутри интерскрина как *наблюдатель за зрителем* в картезианском театре) включается в работу только с появлением проблемы.

Нейрокогнитивные трансформации между афферентными данными, определенными в пространстве поведения, проблемами, определенными в своих локальных пространствах, и демонстрациями в интерскрине, определенными в пространстве психики, выполняют топологизатор, категоризатор, конструктор и факторизатор. Делают они это с помощью специализированных нейрокогнитивных функциональных систем, основанных на н-отображениях, сформировавшихся эволюционно и филогенетически наследуемых агентом. В целом метафорически можно обобщить, что зритель в картезианском театре должен увидеть различные постановки пьесы (алгоритмические схемы) о своем движении в прошлом и будущем с тем, чтобы выбрать вариант с самым убедительным счастливым (по возможности) концом и сделать реальностью именно его.

Такая интерпретация основной функции картезианского театра наводит на мысль о том, что субъективность – это способность системы к управлению объективными процессами с помощью интерпрезентативной модели, объединяющей данные подсознания.

тельных функциональных систем о среде, своем теле в среде, с данными функциональных систем верхнего уровня о работе таких подсознательных функциональных системах в едином формате.

Своеобразная оппозиция «Я» агента как субъекта поведения и проблемы в системе «агент – среда» как некоторой части объективно формирующейся траектории агента в пространстве поведения, по нашему мнению, требуется для сохранения иллюзии субъективного восприятия непрерывного движения агента по дереву жизни при переключении между различными синхронными и асинхронными локальными проблемами в процессе их последовательного и параллельного решения. Технологически для этого как раз и используется главная траектория движения агента в пространстве поведения (можно назвать ее «Я»-траекторией), обеспеченная алгоритмами выбора и встраивания в нее фрагментов траекторий локальных проблем, в том числе с применением нейрокогнитивных функциональных систем интерскрина.

Иллюзия «Я» необходима для формирования своеобразной нейрокогнитивной проекции представления агента о себе самом, своих возможностях и структурно-функциональных особенностях, связанных в том числе с констатацией требований последовательного применения монопольных ресурсов агента для псевдопараллельного решения синхронных и асинхронных проблем в системе «агент – среда».

Поддерживая возможность параллельного решения многих проблем и констатируя сугубо последовательный характер использования монопольных ресурсов, система управления агента и допускает загрузку найденных параллельными алгоритмами решений синхронных и асинхронных проблем в последовательную «Я»-траекторию.

В картезианском театре интерскрина интерэкзистанс и интерспейс демонстрируют движение агента по «Я»-траектории – по стволу дерева жизни агента в прошлом и настоящем, а также множество альтернатив движения по этой траектории в будущем. Нейрокогнитивные функциональные системы модуля «Я» принимают решение о выборе конкретного варианта траектории в будущем, что также сопровождается интеррепрезентацией в компоненте интерэго.

Компонент внимания реализует эволюционно приобретенный механизм переключения фокуса ввода окна настоящего между различными локальными проблемами, что также поддерживает универсальный характер спектра решаемых агентом проблем в условиях ограничения на возможность параллельного использования монопольных ресурсов.

8. ПСИХОГЕНЕЗ КОМПОНЕНТА «Я» СИСТЕМЫ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНОГО АГЕНТА

Для достижения универсальности интеллекта агента принципиальное значение имеет процесс онтогенетического формирования (*психогенеза*) нейропсихологической архитектуры его системы управления. Можно предположить, что с феноменами, определенными в пространстве психики, агент сталкивается гораздо раньше и интенсивность взаимодействия с ними значительно выше, чем с проблемами, детерминированными ситуациями во внешней среде. В этом случае условные «слоты» специализированных нейрокогнитивных функциональных систем компонентов «Я» и интерскрина начинают активно заполняться новыми агнейронами в процессе развертывания экзистенции агента.

Рабочее описание компонента «Я» можно построить на базе метафоры проектирования двух взаимно ортогональных деревьев – субъективного отражения дерева жизни агента, связанного с траекторией агента в пространстве поведения, и психоценторно и психоэффекторно базированного в нем *дерева мышления агента*, связанного с траекторией движения системы управления агента по дереву жизни, определенного в пространстве психики,

связанных между собой функциональными системами нейрокогнитивных трансформаций, обеспечивающих синхронные описания проблем в различных пространствах.

Соответственно, нейроморфологически компонент «Я» представлен всеми агнейронами в составе функциональных систем, формирующихся в этих условно ортогональных («психоортогональных») деревьях на протяжении жизни агента. Следовательно, «Я» расширяется в процессе развития экзистенции агента за счет таких новых агнейронов и ситуативно сформированных на их основе нейрокогнитивных функциональных систем.

По нашему мнению, основной причиной эволюционного возникновения структурно-функционального компонента интерскрина является необходимость синтеза решений на основе анализа всей проблемной ситуации с учетом всех входящих в нее событий.

Фрагменты проблемы из прошлого и будущего, загруженные в интерскрин в качестве частей основного ствола дерева жизни агента, должны актуализироваться событиями из окна настоящего, своевременно перестраивая ответвления дерева в виде локальных проблем и тем самым порождая варианты дальнейшего движения по стволу. Данные текущего афферентного потока в интерскрине бесшовно встраиваются на свое место – «миг между прошлым и будущим» проблемы, демонстрируемой компоненту «Я» в пространстве психики. Нейрокогнитивные функциональные системы компонента «Окно настоящего» дстраивают МА-факты, описывающие текущие события. После этого с помощью нейрокогнитивного компонента «Сортировщик» распознанные события распределяются по релевантным проблемам, которые после этого становятся ассоциативно и логически связанными с текущими вершинами в дереве жизни.

Сознание способно произвольно выбирать любые части дерева жизни – как реально имевшие место вершины ствола дерева, определенные в пространстве поведения, так и его воображаемые ответвления, определенные в пространствах локальных проблем, для того, чтобы по необходимости трансформировать (переопределять) их в пространстве психики и «загружать» в интерскрин с целью анализа, модификации, декомпозиции и объединения с другими фрагментами проблем в интересах синтеза алгоритмов онтологизации новых проблем, идентификации и решения уже ранее онтологизированных проблем.

Такой выбор, по нашему мнению, в общем случае обусловлен прежде всего необходимостью решать самые разнообразные проблемы, возникающие в системе «агент – среда», в том числе абсолютно новые для агента проблемы, требующие выполнения метаалгоритмов онтологизации. В этом смысле системные функции компонентов интерскрина и «Я» обоснованы прежде всего задачами решения проблем универсального спектра на основе синтеза поведения интеллектуального агента в реальной среде.

ЗАКЛЮЧЕНИЕ

Разработана концепция нейропсихологической архитектуры автономной системы управления поведением агента универсального интеллекта.

Показано, что универсальный характер спектра проблем, которые способны обрабатывать такие агенты, обеспечивается методологической и алгоритмической базой онтологизации, идентификации и решения проблем на основе синтеза нейропсихологической архитектурой агента его поведения в системе «агент-среда».

Построена вычислительная абстракция процессов онтологизации, идентификации и решения проблем нейропсихологической архитектурой на основе концепции синтеза автономным агентом траектории своего движения в пространстве поведения и метафоры проектирования дерева жизни агента.

Показано, что с целью достижения универсальности спектра решаемых проблем в нейропсихологической архитектуре объединены структурно-функциональные модули, основанные на нейрокогнитивных архитектурах, включающих в свой состав динамические функциональные системы, состоящие из агентов-нейронов.

Показано, что нейропсихологическая архитектура системы управления интеллектуального агента обеспечивает синтез траектории движения агента в пространстве поведения на основе интеграции в ее состав фрагментов траекторий движения агента в локальных пространствах определения проблем.

Разработана структурно-функциональная схема нейропсихологической архитектуры, мотивированная ее системным назначением, обеспечивающая нейрокогнитивные трансформации пространства определения проблем с помощью согласованной работы нейрокогнитивных компонентов, реализующих каскады н-функций для динамических переходов между пространством поведения, локальными пространствами определения отдельных проблем и пространством психики.

Структура и функции нейрокогнитивных компонентов метафоры проектирования нейропсихологической архитектуры – подсознания, сознания, интерскрина, интерспейса, интерэкзистанса, интерэго, модуля «Я» автономного агента – обоснованы в рамках вычислительной абстракции агента универсального искусственного интеллекта.

СПИСОК ЛИТЕРАТУРЫ

1. <http://www.robotspodcast.com/podcast/2012/10/robots-rethink-robotics/#transcript>
2. Russell S., Norvig P. Artificial Intelligence: A Modern Approach (AIMA). 2nd ed. Moscow: Williams, 2007. 1424 p.
3. Нагоев З. В., Нагоева О. В. Обоснование символов и мультиагентные нейрокогнитивные модели семантики естественного языка. Нальчик: Издательство КБНЦ РАН, 2022. 150 с.
4. Нагоев З. В., Анчёков М. И., Бжихатлов К. Ч. и др. Онтоэписоциофилогенетическое развитие систем общего искусственного интеллекта на основе мультиагентных нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2022. № 6(110). С. 61–75. DOI: 10.35330/1991-6639-2022-6-110-61-75
5. Nagoev Z.V., Nagoeva, O., Gurtueva, I. Multi-Agent neurocognitive models of semantics of spatial localization of events. Cognitive Systems Research. 2020. 59. Pp. 91–102. DOI: 10.1016/j.cogsys.2019.09.015.
6. Nagoev Z. V. Multiagent recursive cognitive architecture // Biologically Inspired Cognitive Architectures 2012. Proceedings of the third annual meeting of the BICA Society, in Advances in Intelligent Systems and Computing series, Springer. 2012. Pp. 247–248.
7. Нагоев З. В. Интеллектика, или Мышление в живых и искусственных системах. Нальчик: Издательство КБНЦ РАН, 2013. 232 с.
8. 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
9. Нагоев З. В., Пшенокова И. А., Нагоева О. В. и др. Имитационная модель нейрокогнитивной системы управления автономным программным агентом, выполняющим кооперативное поведение с целью автоматического пополнения онтологий // Известия Кабардино-Балкарского научного центра РАН. 2023. № 6(116). С. 226–234. DOI: 10.35330/1991-6639-2023-6-116-226-234
10. Нагоев З. В. Мультиагентные экзистенциальные отображения и функции // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 63–71. EDN: QZTFLX

11. 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
12. Нагоев З. В. Мультиагентные экзистенциальные отображения и функции // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 63–71. EDN: QZTFLX
13. Chalmers D. J. The Conscious Mind: In Search of a Fundamental Theory. 1 edition. Oxford: Oxford University Press, 1996. 432 p. ISBN 978-0195117899.
14. John R. Searle, “Minds, brains, and programs”. Behavioral and Brain Sciences 3 (3): 417-457.
15. Нагоев З. В. Онтонейроморфогенетическое моделирование // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 56–63. EDN: QZTFLN
16. Анишев А. З., Аталиков Б. А., Канкулов С. А. и др. Онтофилогенетические алгоритмы синтеза фенотипов интеллектуальных программных агентов для применения в задачах многопоколенной оптимизации управляющих нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2022. № 6(110). С. 76–91. DOI: 10.35330/1991-6639-2022-6-110-76-91

REFERENCES

1. <http://www.robotspodcast.com/podcast/2012/10/robots-rethink-robotics/#transcript>
2. Russell S., Norvig P. Artificial Intelligence: Modern approach (AIMA). 2nd ed. Moscow: Williams, 2007. 1424 p.
3. 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)
4. Nagoev Z.V., Anchakov M.I., Bzhikhatlov K.Ch. et al. Ontoepisociophylogenetic development of general artificial intelligence systems based on multi-agent neurocognitive architectures. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2022. No. 6(110). Pp. 61–75. DOI: 10.35330/1991-6639-2022-6-110-61-75. (In Russian)
5. Nagoev Z.V., Nagoeva O., Gurtueva I. Multi-Agent neurocognitive models of semantics of spatial localization of events. *Cognitive Systems Research*. 2020. 59. Pp. 91–102. DOI: 10.1016/j.cogsys.2019.09.015.
6. Nagoev Z.V. Multiagent recursive cognitive architecture. Biologically inspired cognitive architectures 2012. Proceedings of the Third Annual Meeting of the BICA Society, in Advances in Intelligent Systems and Computing Series, Springer. 2012. Pp. 247–248.
7. 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)
8. Nagoev Z.V. Basic principles of neurocognitive modeling of the consciousness of a universal artificial intelligence agent. *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. (In Russian)
9. Nagoev Z.V., Pshenokova I.A., Nagoeva O.V. et al. Simulation model of a neurocognitive control system for an autonomous software agent performing cooperative behavior for the purpose of automatic ontology replenishment. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2023. No. 6 (116). Pp. 226–234. DOI: 10.35330/1991-6639-2023-6-116-226-234. (In Russian)
10. 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)

11. 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
12. 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)
13. Chalmers D.J. The conscious mind: In search of a fundamental theory. 1 edition. Oxford: Oxford University Press, 1996. 432 p. ISBN 978-0195117899.
14. John R. Searle, “Minds, brains, and programs”. *Behavioral and Brain Sciences* 3 (3): 417-457.
15. 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)
16. Apshev A.Z., Atalikov B.A., Kankulov S.A. et al. Ontophylogenetic algorithms for synthesizing phenotypes of intelligent software agents for use in multi-generation optimization problems of control neurocognitive architectures. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2022. No. 6(110). Pp. 76–91. DOI: 10.35330/1991-6639-2022-6-110-76-91. (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 Российской академии наук;
 360010, Россия, г. Нальчик, ул. Балкарова, 2;
 вед. науч. сотр. отдела «Мультиагентные системы», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра Российской академии наук;
 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 Technical Sciences, General Director of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;
 2, Balkarov street, Nalchik, 360010, Russia;
 Leading Researcher, Department “Multi-Agent Systems”, Institute of Computer Science and Problems of Regional Management – branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;
 37-a, I. Armand street, Nalchik, 360000, Russia;
 zaliman@mail.ru, ORCID: <https://orcid.org/0000-0001-9549-1823>, SPIN-code: 6279-5857