МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Хасанов Рустам Юрьевич

Коннекционистский подход в современных когнитивных исследованиях

5.7.6. Философия науки и техники

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата философских наук

Москва – 2023

Диссертация выполнена на кафедре философии и методологии науки философского факультета ФГБОУ ВО «Московский государственный университет имени М.В.Ломоносова».

Научный руководитель:

Алексеев Андрей Юрьевич, доктор философских наук

Официальные оппоненты:

Барышников Павел Николаевич, доктор философских наук, доцент,
ФГБОУ ВО «Пятигорский государственный университет», профессор кафедры исторических и социально-философских дисциплин, востоковедения и теологии, руководитель научно-образовательного центра вычислительной философии «Digit»;

Михайлов Игорь Феликсович, доктор философских наук,
ФГБУН Институт философии Российской академии наук, старший научный сотрудник сектора методологии междисциплинарных исследований человека;

Петрунин Юрий Юрьевич, доктор философских наук, профессор,
ФГБОУ ВО «Московский государственный университет имени М.В.Ломоносова», заведующий кафедрой математических методов и информационных технологий в управлении факультета государственного управления.

Защита диссертации состоится «19» апреля 2023 г. в 17:30 на заседании диссертационного совета МГУ.057.1 Московского государственного университета имени М.В.Ломоносова по адресу: 119234, г. Москва, Ломоносовский проспект, д. 27, корп. 4 (учебно-научный корпус «Шуваловский»), философский факультет, аудитория А-518 (Зал заседаний Ученого совета факультета).

E-mail: diss@philos.msu.ru.

С диссертацией можно ознакомиться в отделе диссертаций Научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на портале: https://dissovet.msu.ru/dissertation/057.1/2262.

Автореферат разослан «___» марта 2023 г.

Учёный секретарь
диссертационного совета,
кандидат философских наук, доцент Е.В. Брызгалина

I. Общая характеристика работы

Актуальность темы исследования

В настоящее время методы когнитивной науки позволяют по-новому представить активность сетей нейронов в живом мозге и определить разнообразные функции, которые выполняются биологическими сетями нейронов. Однако современная системная нейробиология сталкивается с проблемой многообразия гипотетических объяснений функционирования ансамблей нейронов, исследования которых развиваются параллельно с исследованиями в области психологии и разработками ИИ. Затруднения экспериментальной нейробиологии состоят в том, что эксперименты, разрабатываемые для уменьшения сложности той проблемы, которую пытаются решать авторы эксперимента, только увеличивают набор экспериментальных данных и не уменьшают сложность рассматриваемых проблем. Одним из решений данных затруднений является коннекционистская программа когнитивных исследований, которая опирается на некоторые предположения относительно нейронных сетевых взаимодействий. Коннекционизм кладёт в основу функционирования мозга набор правил взаимодействия внутри сети простых элементов, которые объясняют каким образом происходит кодирование информации, репрезентация среды внутри сети и психологическая интерпретация этой информации. Как метод, такая программа предлагает возможность разложения новых данных из области нейронауки согласно принципам организации таких сетей для удовлетворительного понимания работы мозга и организации познавательных процессов. Одной из базовых вычислительных моделей таких исследований являются искусственные нейронные сети, в прессе упоминаемые в контексте развития искусственного интеллекта основанного на данных. Нейронные сети служат переходным мостом межу исследованиями в области физики, нейробиологии и искусственного интеллекта. Исследования свойств нейронных сетей в первой половине прошлого века стали предтечей программы коннекционизма. Исследователи ИИ на основе нейронных сетей

стремятся смоделировать и репродуцировать мыслительные операции с помощью современных компьютерных технологий, когнитивная психология на основе нейронных сетей стремится описать процессы решения задач и объяснить особенности ассоциативного мышления, теория познания исследует познавательные способности и познавательную деятельность человека, возникшие на основе коннекционистской программы когнитивных исследований, онтология изучает мировоззренческие следствия такой программы, сообщающие нечто о природе нашего мышления. В когнитивной науке особое место занимает исследование репрезентативной эквивалентности искусственных и биологических нейронных сетей. Из этих сравнений вытекают общая логика организации сетей, функционирование сетей и репрезентация данных в сетях, что является важной вехой в развитии понимания устройства мозга и мышления человека. В России активно разрабатываются биологически инспирированные когнитивные архитектуры, а также исследуются сетевые структуры психики, тогда как сама методологическая программа коннекционизма не проблематизируется и её ограничения детально не исследуются. В самом общем виде коннекционизм считает представимым на основе вычислительных устройств, подобных мозговым структурам, воспроизвести психические свойства в искусственных системах из связанных между собой простых элементов. Современные нейросетевые подходы показывают как непосредственно из сетевых нейронных взаимодействий возникают когнитивные функции. Сегодня коннекционистская метафора стала стандартным исходным предположением во многих областях когнитивных исследований. Ввиду быстрого развития современных нейросетевых подходов возникла необходимость философского теоретико-методологического анализа, систематизации и понятийного оформления теоретических и практических наработок этой исследовательской программы, методологическую роль которой предстоит прояснить в настоящей работе. Также необходимо критически проанализировать смысл и актуальность выводов коннекционистской программы относительно

познавательных процессов человека. В настоящей работе осуществляется исследование методологической роли коннекционистской парадигмы для современных когнитивных исследований.

Степень разработанности темы исследования

К философским исследованиям, повлиявшими на разработку коннекционистских моделей, можно отнести исследования аналитических философов в области логицизма начала XX-го века. Б. Рассел, а также А. Уайтхед разрабатывали логические выражения (expressions) для объяснения любого вида математических выражений. У. Питтс, математик и биолог, был лично знаком с Б. Расселом, он создал первую формальную нейронную сеть, положив в основу булевы операторы. Исследования в области психологии, предвосхитившие коннекционистские модели в когнитивной науке, — это разработки школы ассоцианизма. Дэвид Хартли, один из основоположников ассоцианизма, применил учение об ассоциации идей Дж. Локка к физическим процессам (вибрациям) мозга в своём труде «Размышления о человеке, его строении, его долге и упованиях». Понятия ума связаны между собой с помощью ассоциаций, говорил Д. Хартли. Его современник Д. Юм также предложил гипотезу о психических ассоциациях, выделив целый класс философских проблем, связанных с ассоциативным мышлением. И. Сеченов также может быть назван продолжателем идей ассоцианизма, уже в XIX веке он рассуждает о возможности тождественного рассмотрения психической ассоциации и физической связи элементов в мозге. И. Сеченов предлагает ассоциацию в качестве элемента, связующего психические явления и их физиологическую основу в мозге. И. Павлов в начале XX-го века доказывает, как возможна ассоциация на физиологическом уровне. Исследования И. Павлова предложили механизм условного рефлекса, основанный на одновременном возбуждении коркового центра индифферентного раздражителя и коркового центра безусловного рефлекса, что приводит к установлению связи двух раздражителей. Динамический стереотип в рамках павловского учения возможно представить в виде сети ассоциаций, возникших

внутри ассоциативных зон коры.

Ранние разработки по формализации активности нейронной сети были предложены У. Мак-Каллоком и У. Питтсом в 1943-м году и продолжены в работах Ф. Розенблатта в 1960-х. Но, несомненно, всеобщее признание коннекционистские модели получили после публикации в 1986-м году Д. Румельхарта и Дж. Макклелланда с соавторами, где обосновывалось предложение о параллельной распределённой обработке информации, как самостоятельной парадигме, объясняющей экспериментальные данные, известные в лингвистике и психологии. Смыслообразующей программой здесь стала биологически правдоподобная архитектура вычислительного устройства, а также алгоритмы обучения и целевые функции, как замена классическому программированию. К моменту написания работы по моделям PDP (Parallel Distributed Processing Model) в когнитивной психологии уже были известны случаи, для которых не могла применяться классическая вычислительная метафора. В качестве примера часто используют «эффект превосходства слов», который свидетельствует о том, что в знакомом контексте объекты опознаются быстрее и точнее, чем в незнакомом. В частности, люди могут обнаружить буквы в слове быстрее, чем отдельные буквы в бессмысленной последовательности букв. Для объяснения этого феномена была предложена нейросетевая модель, в её основе простые элементы, связанные между собой, одни из которых отвечают за опознавание букв, другие за части слов и третьи за целые слова. Так можно построить систему, которая может объяснить контекстное распознавание. Эта система показывала, как можно догадаться, какое слово написано, даже если в нем есть ошибка (и часто эта ошибка не заметна). На основе нейронных сетей было также предложено правдоподобное объяснение классической ошибки сверхрегуляризации, которую осмысляли с помощью формальных правил, применяемых человеком к новым данным. «Модель интерактивной активации» продемонстрировала, как ошибка сверхрегуляризации для неправильных глаголов прошедшего времени в английском языке может

возникать в нейронной сети без формулирования правил в явном виде. Эти убедительные свидетельства работоспособности коннекционистской парадигмы заинтересовали когнитивистов. Начиная с 1990-х годов коннекционистские разработки активно используются в когнитивной науке. Разработка моделей PDP позволила смоделировать появление структурированных и абстрактных представлений в многослойных иерархических сетях. Рекуррентные сети помогли понять различные аспекты изучения языка, такие как категоризация, контекстное различение слова и выделение морфем. Сверточные нейронные сети помогли в понимании обработки зрительной информации мозгом и объяснили, как обрабатывается перцептивная информация. Обучение с подкреплением применительно к нейронным сетям дало возможность объяснять действия, желания и мотивацию у когнитивных агентов. Однако выделение нейросетевых методов в отдельную отрасль машинного обучения привело к биологически неправдоподобным алгоритмам обучения и упрощённому пониманию нейронов и архитектур нейронных сетей. Ф. Крик в 1989 году осветил проблему понимания вычислительных свойств мозга посредством искусственных нейронных сетей. С его точки зрения, прямое сравнение обучения искусственных и естественных нейронных сетей неудачно и не отвечает реальным данным из области нейронауки. Критику коннекционистской метафоры продолжают исследователи распределенного хранения данных в сетях нейронов С. Малбург, М. Макклоски, Р. Роджер. Они отмечают, что новое обучение сетей приводит к катастрофе суперпозиции (superposition catastrophe), в результате которой происходит забывание предыдущего обучения. В реальном мозге это приводило бы к потере долгосрочной памяти. Интересным аспектом коннекционистской проблематики является проблема временной и пространственной инвариантности, обусловленная особенностями механизмов естественного интеллекта, а именно врождённой способностью определения единиц объектов, которой нет у нейронных сетей. Также нейронные сети не

формируют правила в явном виде и не могут прямым образом переносить знания на новый тип данных, эта проблема является основной линией критики коннекционизма сегодня. Обучение нейронной сети не позволяет объединить прямым образом элементы относительно их общего смысла, то есть сформулировать правило в явном виде. То же относится к обобщению смыслов и идей, эти вопросы изучены Г. Маркусом, Дж. Фодором, М. Зайденбергом. Полемика вокруг коннекционистской программы сегодня разворачивается во многом из-за новых способов решения этих старых проблем. Аргументы касательно обратного распространения ошибки обходятся коннекционистами за счёт биологически правдоподобных правил обучения сетей выравнивание обратной связи (equilibrium propagation), равновесное распространение (Feedback Alignment), и прогностическое кодирование (например, работы Я. Ястрофа, Л. Перье), кроме того, исследователи, убежденные в работоспособности алгоритма обратного распространения ошибки, ищут его признаки в мозге. Решению проблемы забывания в настоящее время посвящено довольно много работ. К ним относится, например, работа М. Мермиллорд о системах дополнительного обучения, или работа Дж. Боверса, о локальном хранении данных в нейронных сетях. Проблема инвариантности также находит свои решения при условии использования архитектур особого вида. Сегодня происходит бурное развитие нейросетевого подхода и машинного обучения формальных нейронных сетей. Модели предсказательного кодирования, вероятностные модели языковых способностей человека, модели глубокого обучения для нейронауки и даже модели сознательных состояний строятся на основе коннекционистского подхода. В настоящее время в каждой конкретной области когнитивной науки используются нейросетевые подходы. И сегодня уже существуют некоторые доказательства репрезентативной эквивалентности между нейронными сетями и мозгом, на самом деле сегодня публикуются сотни работ предлагающие общие свойства моделей «глубоких нейронных сетей» и когнитивных функций, реализуемых в мозге. В то же самое время разработчики ИИ,

вдохновлённые новыми открытиями в области нейробиологии, активно переосмысляют эмпирический материал, изобретая и тестируя новые искусственные сети. Нейросетевые методы породили целую волну новых исследований в области ИИ и возобновили дискуссии о роли таких моделей в функционировании естественного интеллекта. Эта бурно развивающаяся область философских исследований в русскоязычной философской литературе обсуждена недостаточно подробно.

Современными отечественными учёными и философами коннекционизм изучался в рамках разработок в области неклассической эпистемологии В. А. Лекторским, в рамках функциональных моделей в естественном языке Т. В. Черниговской, в рамках нейрокомпьютерных разработок А. В. Савельевым, в рамках междисциплинарных исследований нейрофилософии В. Г. Кузнецовым, для осмысления соотношения психических феноменов с их нейробиологическим субстратом Д. В. Ивановым, а также в отношении проблематики философии искусственного интеллекта А. Ю. Алексеевым, прояснившим вклад С.Н. Корсакова в развитие коннекционистской парадигмы и Петруниным Ю. Ю в рамках исследований истории искусственного интеллекта. Также стоит отметить работы Е.А. Янковской в гносеологии, И.Ф. Михайлова в области социальной философии, В.Д. Арутюняна и И.Г. Овчинниковой в области лингвистики. Коннекционистская парадигма встречается в большом количестве диссертационных исследований (например, Н. Ю. Клюева Е.М. Панина, А.Г. Сонин, М.А. Сущин, В.А. Герович, А.Ю. Алексеев, Барышников П. Н., Савельев А. В.), но как самостоятельный объект философских исследований не проблематизируется. В настоящем исследовании также нельзя охватить весь спектр существующей литературы, связанной с нейросетевыми методами, который очень велик из-за проникновения машинного обучения (своеобразного технологического аналога методологического понятия «коннекционизм») во многие сферы жизни, но попытаемся выделить наиболее демонстративные примеры.

Цель исследования. На основе комплекса современных междисциплинарных исследований системной нейробиологии, психофизиологии, нейрофизиологии, философии и методологии искусственного интеллекта выявить и критически исследовать методологическую роль коннекционистского подхода в становлении и развитии проблематики современной когнитивной науки.

Для достижения поставленной цели решаются следующие **задачи**:

1. Раскрыть роль и место коннекционизма в философской методологии современной когнитивной науки.

2. Выявить роль коннекционизма в методологии нейробиологии и сформировать набор принципиальных характеристик нейросетевых моделей.

3. Выявить роль коннекционизма в методологии современной психологии и определить преимущества и недостатки моделей машинного обучения.

4. Изучить перспективы применения информационного подхода к сознанию для решения проблемы описания психических явлений на базе коннекционистской методологии.

Объектом настоящего исследования является коннекционистский подход к изучению проблем познания в когнитивной науке.

Предметом исследования является философско-методологическая роль вычислительных моделей на основе распределенной и параллельной обработки данных для современной когнитивной науки.

Научная новизна исследования

1. Предлагаются новые критерии сходства с биообъектом мозгом искусственных нейронных сетей с помощью уточнения и систематизации основных поверхностных и структурных сходств, а также найденных функциональных примитивов на уровне реализации алгоритмов естественных и искусственных нейронных сетей в виде канонических вычислительных операций «вычисления производной по времени» и «нормализации».

2. Впервые верифицирована гипотеза включённости операций «вычисления производной по времени» и «нормализации» в сигнатуру

канонических операций описания динамики естественных нейронных сетей, так и в сигнатуру описания способа функционирования формальных нейронных сетей, которое имеет значение для прояснения подобия свойств между естественными и искусственными нейронными сетями.

3. Впервые аргументирована необходимость коннекционизма для методологической координации психических феноменов и нейробиологических явлений в рамках информационного подхода к сознанию и вычислительной теории разума.

4. Показана роль технической науки (информатики и теории связи) в становлении нейронной теории в прогнозирующем кодировании.

Теоретическая и практическая значимость исследования

Теоретическая значимость исследования формируется его вкладом в современное философское осмысление роли и значения программы моделирования психических феноменов на основе нейросетевых моделей в вычислительной теории разума. Выявлены тенденции в расширении области машинного обучения на основе революционных архитектур коннекционистского типа, применительно к анализу данных современной нейробиологии. Обоснованы гносеологические предпосылки появления и развития коннекционистской парадигмы в рамках работы по созданию изофункциональных систем естественного интеллекта. Аргументы в пользу канонических вычислительных операций, производимых и мозгом, и нейрокомпьютером, могут быть использованы при разработке теоретических моделей в области системной нейробиологии.

Материалы диссертационного исследования могут использоваться в курсах лекций по теории познания, философии науки, а также в спецкурсах «Философия искусственного интеллекта», «Философия информатики», «Философские проблемы когнитивной науки».

Теоретико-методологическая база

Методы диссертационного исследования – это метод аргументации, включающий идентификацию и проверку аргументов, методы исторической

реконструкции и концептуальной инженерии. При освещении темы истории возникновения коннекционистского подхода и самой когнитивной науки, а также при рассмотрении классических моделей PDP Д. Румельхарта, Макклелланда Дж., Элмана Дж., Маркуса Г., Я. Ликуна и др. опора происходит на метод исторической реконструкции научных и философских теорий. Для выявления оснований коннекционизма, а также гипотез «распределённого» и «параллельного» способа обработки информации главным выступал классический метод аргументации, позволивший выделить основные теоретические проблемы и ограничения этих подходов. Во второй главе был выдвинут ряд аргументов в поддержку информационного подхода для объяснения природы высокоуровневых когнитивных функций в рамках коннекционистского подхода в когнитивной науке, а также сформулирована гипотеза о возможности интеграции данных системной нейробиологии в рамках информационного подхода в когнитивной науке. Исторический подход, позволяющий проследить принципиальные и исходные положения коннекционизма в сравнении с символизмом, затрагивает статью А. Лавлейс (1843 г.) и библиографическую находку последних лет – статью С.Н. Корсакова (1832 г.).

Положения, выносимые на защиту

1. В современной когнитивной науке коннекционизм как методология параллельной распределённой вычислительной реализации когнитивных функций, двойственно дополняет символизм как методологию последовательной лингвистически заданной вычислительной реализации этих функций.

2. Коннекционизм является достаточным логико-эпистемологическим условием решения проблемы представления мозговой активности: он выражает принципы обработки информации в мозге посредством настройки весовых коэффициентов нейронных связей, не выражая онтологических притязаний на морфологическое подобие и физический изофункционализм искусственных и естественных нейронных систем.

3. Коннекционизм является достаточным логико-эпистемологическим условием решения проблемы представления познавательной деятельности человека: он объясняет механизм параллельной работы ассоциативного мышления и способность одномоментного восприятия различных чувственных модальностей сознания.

4. Коннекционизм является необходимым логико-эпистемологическим условием психофизиологического взаимодействия: психические феномены одномоментны и нелокальны с их нейросетевыми коррелятами.

Степень достоверности и апробация результатов исследования

Достоверность и обоснованность результатов исследования обеспечивается принятой методологией, соответствием содержания работы ее теме, наукометрическими показателями статей, в которых были опубликованы материалы диссертации, а также опорой на обширный круг исследовательской литературы в различных областях знания.

Основные положения и выводы исследования были изложены в 5-и научных работах, опубликованных в изданиях, отвечающих требованиям п. 2.3 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова.

Диссертация прошла обсуждение на кафедре философии и методологии науки философского факультета Московского государственного университета имени М.В. Ломоносова и получила положительное заключение. Основные результаты диссертационного исследования и возможности их теоретического применения в различных предметных областях были апробированы в качестве докладов на следующих конференциях:

1. Хасанов Р.Ю. Коннекционистские интерпретации социального мышления // Международная междисциплинарная конференция «Искусственный интеллект в новой коммуникативной реальности», Москва, Россия, 18-19 июня 2020 г.
2. Хасанов Р.Ю. Нейротехнологии в фармакологии // Международная молодёжная междисциплинарная конференция «Философия искусственного

интеллекта», Москва, Россия, 30 мая-20 июня 2019 г.

3. Хасанов Р.Ю. Проблема сознания в нейробиологии // Международная молодёжная междисциплинарная конференция «Философия искусственного интеллекта», Москва, Россия, 30 мая-20 июня 2019 г.

4. Хасанов Р.Ю. Параметры качеств субъективной реальности и их естественнонаучные способы изучения // XXVI Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов», Москва, Россия, 8-12 апреля 2019 г.

5. Хасанов Р.Ю. Словарь этических терминов в коннекционистских компьютерных системах // Международная научно-практическая конференция «Искусственный интеллект: этические проблемы цифрового общества», Белгород, Россия, 11-12 октября 2018 г.

6. Хасанов Р.Ю. Дефиниции интеллекта в коннекционизме // Международная молодёжная конференция «Философия искусственного интеллекта – 2018», Москва, Россия, 12-13 апреля 2018 г.

7. Хасанов Р.Ю. Коннекционистский подход в общем искусственном интеллекте // Международная конференция «Актуальные проблемы гуманитарных и социальных исследований», Новосибирск, Россия, 6 октября 2020 г.

Структура исследования

Диссертация состоит из введения, двух глав, заключения и списка литературы.

II. Основное содержание работы

Структура диссертационного исследования отражает решение поставленных в диссертации задач для достижения обозначенной цели. Работа состоит из введения, двух глав, включающих в общей сложности тринадцать параграфов, заключения и списка литературы, состоящего из 194 источников, 145 из которых составляют источники на английском языке.

Настоящее диссертационное исследование посвящено методологической роли коннекционизма в современных когнитивных исследованиях. В нашей работе мы ставим целью выявить и критически исследовать методологическую роль коннекционистского подхода в становлении и развитии проблематики современной когнитивной науки.

Предметом нашего исследовательского интереса становится философско-методологическая роль вычислительных моделей на основе распределённой и параллельной обработки данных для современной когнитивной науки. Структура диссертации устроена следующим образом. Первая глава исследует роль коннекционистских моделей для физиологии и нейробиологии, вторая глава диссертации исследует роль коннекционистского похода в эпистемологии, психологии и когнитивной науке. В последнем параграфе второй главы диссертации обсуждается роль коннекционизма для решения психофизиологической проблемы.

Во **Введении** обосновывается актуальность темы исследования, рассматривается степень её научной разработанности, определяются цель и задачи исследования, раскрывается его научная новизна, теоретическая и практическая значимость, теоретико-методологические основания, а также формулируются положения, выносимые на защиту.

Первая глава - **Теоретико-познавательные и методологические проблемы коннекционизма** – посвящена рассмотрению и критическому анализу основных положений коннекционистского подхода и трудностей использования такого подхода на современном этапе развития когнитивных наук. В главе определяется и исследуется коннекционистская программа когнитивных исследований, дан обзор её современного состояния, исследуются

сущностные и понятийные аспекты коннекционизма. Анализ показал, что коннекционизм двойственно дополняет символизм как методологию последовательной лингвистически заданной вычислительной реализации этих функций. Для устранения проблемы отсутствия критериев сходства с нейросетями мозга уточняются физические свойства нейрона и искусственного вычислительного органа и производится поиск функциональных коррелятов на уровне реализации алгоритмов. Исследуются роль и значение философии для коннекционистских моделей, отмечается роль коннекционизма в философии искусственного интеллекта.

В первом параграфе - **Вычислительные модели познавательных процессов в когнитивной науке** – обозреваются основные положения вычислительной теории разума, в рамках которой предлагается понятие «когнитивная функция» для выявления нейробиологических основ психики. Реконструируются основные положения когнитивной науки в рамках символьной и субсимвольной парадигмы. Показывается особое положение вычислительных устройств, которые воспроизводят интеллектуальные способности человека и прямым образом сравниваются с функциональными элементами мозга. Указывается место коннекционистской архитектуры в рамках классической вычислительной теории разума, как биологически правдоподобной логической структуры, организующей работу механических вычислительных органов машин параллельно и распределённо. Описываются основные следствия, вытекающие из такой структурной организации, а также рассматривается эволюция развития нейронных сетей и таких аспектов мышления человека, как ассоциативное мышление, параллелизм перцептивных и когнитивных актов, исследуемых когнитивной наукой, на которые обратили внимание ученые в свете развития коннекционизма.

В первом подпункте первого параграфа - **Вычисление и шифрование –** предлагается расширение коннекционистской метафоры для когнитивных теорий не вычислительного типа. Осмысляется работа «машины Корсакова», как протонейрокомпьютера, который использует шифрование данных без кодирования этих данных и позволяет создавать сети связей признаков без

создания символьной записи этих признаков.

Во втором параграфе – **История развития нейросетевых моделей обработки сигналов мозгом** – Рассмотрена история возникновения, основные направления и этапы развития коннекционистской программы когнитивных исследований, определяются специфические особенности коннекционизма, оценивается степень новизны и результативности современных моделей искусственных нейронных сетей, формулируются ограничения теории нейронных сетей для когнитивной науки: зависимость от алгоритмов, отсутствие методов и критериев сравнения с нейронными сетями мозга.

В третьем параграфе - **Проблема описания познавательных процессов с помощью моделей коннекционистского типа** – исследуются возможные переходы от причинно связанных электронных элементов к психическим явлениям. Рассматривается функционализм как базовая модель для описания психических феноменов в когнитивной науке. Обозреваются аргументы «за» и «против» этой модели и рассматриваются истоки происхождения сравнения машин и мозга. Выясняется роль кибернетического мировоззрения для становления современной проблематики когнитивной науки и исходящие из этого мировоззрения возможности воплощения когнитивных функций в машине, в качестве примера приводится работа Н. Винера «Бог и голем», созданную для популяризации кибернетической идеи. Показана возможность сравнения машины и мозга с точки зрения цели выполняемых вычислений в рамках вычислительной теории разума. Изучается принципиальное устройство машин, выполняющих логические операции. Рассматриваются машина Беббиджа и современные компьютеры. Обсуждается машина Корсакова как первая машина коннекционистского типа и выявляются принципиальные отличия её функционирования (выполнения логических операций) от операций на классических компьютерах. Обосновывается возможность поиска критериев сходства нейронов и искусственных органов машин на морфологическом уровне.

Принципиальные отличия искусственных нейронных сетей от машины Корсакова — это автоматическая настройка значений весовых коэффициентов

связей сети и наличие промежуточных слоёв сети.

В четвертом параграфе – **Принципы функционирования и общая организация нервных клеток в сравнении с искусственными вычислительными органами** – исследуются особенности функционирования биологических нейронов. Показываются общие места с которыми согласно большинство нейроученых, а именно представление биологического нейрона в качестве элементарного функционального элемента работы мозга, смысл которого сводится к проведению или не проведению сигнала (модель связывающего нейрона). Представляются особенности функционирования нейрона. Указываются отличительные свойства нервной системы от вычислительных органов машин: 1) наличие глиальных клеток и кровеносной системы; 2) разнородность клеток; 3) различные направления токов по мембране; 4) наличие спонтанной активности; 5) явление синхронизации и рассинхронизации на уровне клеток; 6) адаптация нейрона; 7) химическая сигнализация и внесинаптическая сигнализация; 8) нейрогенез и спрутинг аксонов. Описываются системные свойства нервной системы: автономность и самореферентность, они вытекают из базовых принципов самоорганизации живых систем. Показываются свойства нейрона и нервной ткани, которые воспроизводятся в коннекционистских моделях. Разбирается модель связывающего нейрона, приводятся аргументы в пользу невозможности сведения всей нейронной активности к модели связывающего нейрона. Показывается переход от биологических вычислений к вычислениям компьютерного типа.

Делается вывод, что идеализация основных функций нейрона не лёгкий и, во многом, интуитивный и экспериментальный процесс по отделению стандартных повторяющихся характеристик нейрона, которые могут быть полезны при работе с большими сетями таких элементов, от неповторяющихся индивидуальных характеристик нейрона.

В пятом параграфе – **Сравнение физических свойств элементов компьютера в сравнении с элементами мозга (изоморфизм на уровне реализации)** – исследуется морфологическое и функциональное подобие

искусственных и биологических нейронов. Проводится подробный сравнительный анализ физических свойств вычислительных органов машины и физических свойств элементов нервной ткани. Рассматриваются поверхностные сходства, такие как линейные размеры, энергопотребление и скорость функционирования нейронов и элементов машины.

Рассмотрение мозга, как вычислительного устройства коннекционистского типа, приводит к очевидным функциональным различиям в способности порождения сложного поведения при сопоставимых вычислительных мощностях. Это идея оказывается работоспособной в виду (при условии, что мозг — это компьютер) логики проводимых вычислений и представления данных в системе. В работе показывается, что далеко не все механизмы мозга сегодня могут быть смоделированы на основе искусственных механических систем. Обсуждается отсутствие изофункционализма и подобия морфологии искусственных механических систем коннекционистского типа и биологических нейронов на физическом уровне. Выводы, которые делаются из сравнений элементов машин и нейронов такие:

1. Искусственные логические элементы сильно отличаются по своим физическим свойствам: линейным, энергетическим и временным. Активные элементы машин (транзисторы) меньше нейронов на 8 – 9 порядков и меньше синапсов на 3 порядка; превосходят нейроны в скорости на 4 порядка; энергопотребление одного транзистора в сто раз больше, чем у одного нейрона. Эти показатели указывают на отсутствие изоморфизма искусственных систем и мозга на физическом уровне.

2. Несмотря на то, что элементы компьютера мельче и быстрее, то есть количество операций в секунду, которые они выполняют, больше, но их производительность ниже. Для **Cerebras Wafer Scale Engine** этот показатель составляет $3{,}3*10^{15}$ флоп/с, против $2*10^{17}$ флоп/с для мозга (10^{10} нейронов*10^4 синапсов*200 импульсов в секунду). Значит, производительность мозга опережает современные искусственные системы в сто раз.

3. Очевидна разница тех функций, которые реализуются в мозге и в

машине. Поэтому важной частью исследования становится сама *логика вычислительной системы и структура данных,* представленных в этой системе. Логика и алгоритмы обработки информации в современных транзисторных схемах только приближаются к тем, что реализованы в мозге. Это действительно и в обратном случае, поэтому становится важно, что вычисление какой-либо функции при изучении мозга, как вычислительной системы, невозможно без изучения внутренней структуры данных в мозге и самих процедур вычислительных операций.

В шестом параграфе - **Поиск изофункциональных примитивов в искусственных нейронных сетях на алгоритмическом уровне обработки сигнала в мозге** – производится поиск изофункционального подобия свойств искусственных нейронных сетей и биологических сетей нейронов на алгоритмическом уровне по Д. Марру. Делается предположение что алгоритмы настройки соединений биологических нейронов сходны с аналогичными в искусственных нейронных сетях. Переход от уровня реализации к алгоритмическому уровню является вопросом правильной функциональной декомпозиции и если на физическом уровне обе системы не имеют формального подобия, то возможно они обрабатывают сигналы аналогичным образом. Внутренняя (физическая) структура примитивных операций определяет то, как строятся на ее основе алгоритмы. Поэтому интерес представляет поиск в машине примитивных эквивалентных операций, которые происходят в мозге и на основе которых собираются функции. Обучение с подкреплением – объясняет действия агентов в среде, ценностные характеристики, желания и мотивацию; свёрточные нейронные сети – объясняют перцепцию; рекуррентные нейронные сети – объясняют память о предыдущем шаге, внимание и обработку последовательных данных.

В первом подпункте шестого параграфа **– первые модели искусственных нейронных сетей –** изучаются алгоритмы настойки коэффициентов персептрона Ф. Розенблатта в сравнении с клеточными механизмами обучения – ассоциация, сенсибилизация, привыкание. Проясняются этапы реализации алгоритмов, показывается принципиальная схожесть алгоритма обучения с

подкреплением и алгоритмов на основе дофаминового предсказания ошибки и реального обучения млекопитающих на основе павловского обучения. Но очень часто животные проявляют активные действия по изучению среды обитания. Их обучение и познание опосредуется действиями и движениями, например, оперантное обуславливание. Для такого рода поведения систем подкрепления Ф. Розенблатта недостаточно. В настоящее время продолжение исследований обучения с подкреплением приводят к парадигме, которая рассматривает обучение с подкреплением как достаточное условие для любой познавательной деятельности.

Во втором подпункте шестого параграфа – **Современные модели нейронных сетей в когнитивной науке** – рассматривается подобие свойств искусственных нейронных сетей на уровне алгоритмов, указывается подобие схемы обработки информации с помощью операции свертки, общая структура взаимосвязей в вентральном зрительном пути и свёрточной нейронной сети. Указывается появление моделей свёрточных нейронных сетей как попытка реконструкции обработки зрительной информации Д. Хьюбела и Т. Визеля. Рассматриваются элементарный алгоритм навигации с помощью вычисления *относительного* изменения концентрации запаха у C. elegans. Объясняется подобие этого алгоритма с алгоритмами настройки коэффициентов в искусственной нейронной сети. Обсуждаются популярные алгоритмы Д. Марра для зрительной системы, которые также используют производную по времени для кодирования первичных перцептивных сигналов. Этот функциональный примитив в виде вычисления производной по времени предлагается на роль канонической вычислительной операции перевода абсолютной физической характеристики сигнала в относительную единицу кода понятную для обучающейся системы нейронов.

В седьмом параграфе - **Методологическая роль параллельной и распределенной обработки данных в современных когнитивных исследованиях** – на основе произведенного выше критического анализа коннекционистского подхода делается заключение о глубоких методологических отличиях современных методов вычислительной нейронауки

и коннекционистского подхода. Приводятся аргументы в пользу наличия у коннекционистских моделей опоры на вычислительную метафору, в то время как для вычислительной нейронауки такая метафора не всегда пригождается. Феноменологические модели нейронной деятельности воспроизводящие отдельные признаки нейрона опираются на данные экспериментальной нейробиологии, такие модели воспроизводят намного больше разнообразных морфологических и структурных признаков живых нейронов.

В второй главе - **Обработка информации в коннекционистской системе** – рассматриваются познавательные процессы с точки зрения коннекционистской парадигмы. Приводится историческая справка, связывающая работы философов нового времени, таких как Р. Декарт, Т. Гоббс и Г. Лейбниц с современной проблематикой коннекционизма. Усматривается преемственность традиции Г. Лейбница в рассмотрении работы мозга как машины по обработке информации. Рассматриваются модель Дж. Макклелланда и Д. Румельхарта, рассматриваются другие результаты PDP моделей. Рассматриваются трудности таких моделей, связанные с экспликацией правил в явном виде и нахождением каузальных связей между различными физическими сигналами. Исследуется применимость коннекционистской парадигмы к объяснению важных характеристик человеческого сознания.

В первом параграфе - **Репрезентация знаний в когнитивной системе коннекционистского типа –** обсуждается кодирование информации в искусственной нейронной сети. Обсуждаются следствия представления психологических механизмов формирования связей в разуме на основе коннекционистских моделей для философских теорий об установлении истинности между языком, мышлением и предметами этого мира. Коннекционистские модели проясняют именно психологические аспекты установления связей, но не дают ответ на то, как нам поступать с такими высказываниями, а значит они применимы как для поклонников эмпиризма, так и сторонников рационализма. Приводятся аргументы о наличии параллельности и распределённости как признака системы обработки сигналов (мозг) и параллельности и распределённости как признака феноменов

окружающего мира.

Во втором параграфе - **Алгоритмы обучения искусственных нейронных сетей** – обсуждаются правила обучения в нейронной сети, и их возможная биологическая правдоподобность. Отмечаются основные алгоритмы, применяемые для настраивания весов связей в искусственных нейронных сетях, и обсуждаются возможные интерпретации алгоритма обратного распространения ошибки. В этом параграфе исследуется производная по времени как биологически правдоподобная каноническая вычислительная операция, введённая в предыдущей главе, как элементарный функциональный коррелят психических процессов и показывается полезность разработки в этой области для сличения элементарных психических актов и физиологических явлений нейробиологического уровня.

В третьем параграфе - **Память в моделях PDP** – показывается влияние коннекционистской парадигмы на современные исследования памяти. Объясняется устройство памяти в коннекционистской модели, из которой следует что различия между обработкой нового сигнала, репрезентацией и актуализацией памятных следов, которые четко размечаются в классическом символьном подходе, стираются для коннекционизма.

В четвертом параграфе - **Проблемы несимвольного кодирования информации** – рассматривается популярная тема для критики коннекционистского подхода, связанная с отсутствием у нейросетей способности применения знаний за пределами тренировочного пространства. Нейронные сети могут охватывать автоматические процессы, касающиеся чтения слов или образования морфем, но у теоретиков искусственного интеллекта остаётся довольно много сомнений в отношении применения моделей PDP к аспектам семантического познания и многим формам рассуждений, включая причинно-следственные и силлогистические рассуждения. Рассматривается теория дуального процесса мышления для ограничения моделей когнитивных функций на основе нейронных сетей. В качестве примеров аналитического мышления приводятся речь и письмо, «ментальная арифметика» или любого рода вычисления в уме, мысленные

перемещения во времени и процесс принятия решений. Обсуждаются возможные способы моделирования таких видов аналитического мышления с помощью коннекционистских моделей. Что требует пересмотра структуры ментальных состояний с точки зрения коннекционизма для психологической теории рассматривающий в качестве объяснительной модели коннекционизм.

В пятом параграфе - **Философское осмысление понятия информации** – предлагается понятие информации для коннекционистской парадигмы как элементарного вида различение. Рассматриваются популярные информационные подходы Д.И. Дубровского и Д. Чалмерса, использующие свойство двухаспектности информации для решения проблемы сознания. Продолжаются спекулятивные рассуждения Д. Чалмерса и показывается как информационное пространство Д. Чалмерса возможно использовать для нейронных сетей. Обозревается информационная теория Д. И. Дубровского и проблема сличения информационных состояний с нейродинамическими процессами в мозге.

В шестом параграфе - **Информационное пространство коннекционистской архитектуры** – исследуется понятие информационного пространства Д. Чалмерса, где информационные состояния с помощью простого различения двух и более макроскопических состояний маркируются как информационные. Структура информационных состояний может быть самая различная. Информационные состояния группируются в информационные пространства за счёт увеличения количества состояний (различение сразу трех, четырёх и более, до бесконечности (континуум состояний, например, все числа между 0 и 1). Такие пространства очень похожи на топологические многомерные пространства весовых коэффициентов глубокой нейросети. Каждому информационному состоянию в такой сети можно найти уникальный набор весовых коэффициентов. Д. Чалмерс расширяет понятие информационного пространства до такой степени, что оно совпадает с динамическим пространством весовых коэффициентов глубокой ИНС. И в этом смысле любое феноменальное различие, которое можно пронумеровать методом Д. Чалмерса и которое может зафиксировать глаз, ухо,

язык и др., может быть сопоставлено с точкой на таком пространстве весовых коэффициентов.

Коннекционизм предсказывает соответствие типов ментальных феноменов с типами информационных систем мозга. И в случае обработки перцептивной информации в мозге, и в случае представления различных квалитативных характеристик в феноменальном пространстве наблюдается параллельное и распределённое представление информации, которое очевидно в силу наличия разных типов сенсорных систем и соответствующих им типов модальностей. Например, сенсорная система зрительного тракта соответствует зрительному восприятию, то же самое и со слухом. И в сознании, и в мозге информация услышанном и увиденном представлена одновременно и в разных локусах (параллельно и распределённо). Современное развитие науки показывает, как параллельно и распределённо обрабатываются различные типы информации мозгом и эти типы соотносятся с тем, как в сознании представляются различные характеристики опыта (тоже параллельно и распределённо). Поэтому коннекционизм – необходимое условие для сличения ментальных и физиологических характеристик любой теории, которая предлагает способ сличения мозгового механизма с некоторым состоянием сознания.

В **Заключении** кратко суммируются основные рассуждения и идеи, обобщенно представляются результаты и выводы исследования и предлагаются рекомендации для использования коннекционистского подхода в проекте общего искусственного интеллекта.

III. Список публикаций по теме диссертации

I. Публикации в изданиях, отвечающих требованиям п. 2.3 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова:

А) Публикации в рецензируемых изданиях, индексируемых в международных базах Web of Science, Scopus, RSCI:

1. Brak Ivan V., Filimonova Elena, Zakhariya Oleg, Khasanov Rustam, Stepanyan Ivan. Transcranial current stimulation as a tool of neuromodulation of cognitive functions in parkinson's disease[1] // Frontiers in neuroscience. — 2022. — Vol. 16 (Web of Science, Scopus; JCR – отсутствует, JCI – 0,87, CiteScore – 6,6, пятилетний импакт-фактор РИНЦ – отсутствует).

2. Соколов И.С., Татаринцев М.К., Хасанов Р.Ю., Азиева А.М., Макаренко Е.Ю., Бурцев М.С. Устойчивость спонтанной электрической активности нейронных сетей in vitro[2] // Вестник Российского государственного медицинского университета. — 2016. — № 2. — С. 45–49 (Web of Science, Scopus; JCR – 0,04, CiteScore – 0,5, пятилетний импакт-фактор РИНЦ – 0,424).

Б) Публикации в журналах, включенных в Список рецензируемых научных изданий по философским наукам, утвержденный решением Ученого совета МГУ имени М.В. Ломоносова:

3. Хасанов Р. Ю. Время как общий параметр качеств субъективной реальности // Искусственные общества. – 2019. – Т. 14. – Выпуск 2 (пятилетний импакт-фактор РИНЦ: 0,377). DOI – 10.18254/S207751800005656-0.

---

[1] Авторский вклад соискателя в данной работе определяется следующими задачами, выполненными им в рамках исследования, результаты которого были представлены в итоговом тексте, подготовленном к публикации: Хасанов Р.Ю. написал главу о гипотетических механизмах, объясняющих положительные эффекты от транскраниальной стимуляции мозга. Результаты отражены в седьмом параграфе первой главы диссертации, где показаны методологические различия современной вычислительной нейронауки и коннекционистского подхода при разработке моделей когнитивных функций. В статье рассматриваются модели, объясняющие болезнь Паркинсона, которые соединяют вычислительную и экспериментальную нейронауку. Показательно их отличие от коннекционистских моделей когнитивных функций, тем что они учитывают влияние физиологических эффектов на выполнение когнитивной функции.

[2] Авторский вклад соискателя в данной работе определяется следующими задачами, выполненными им в рамках исследования, результаты которого были представлены в итоговом тексте, подготовленном к публикации: Хасанов Р.Ю. участвовал в обеспечении экспериментальной части исследования по обучению нейрональной культуры; планировал постановку экспериментов; производил посадки культур нейронов, выделенных из гиппокампа новорожденных мышей линии C57BL/6, на 60-канальных мультиэлектродных матрицах 60StimMEA200/30-ITO (Multichannel Systems, Германия); записывал спонтанную пачечную активность культур; производил эксперимент по стимуляции культур; а также принимал участие в обсуждении анализа активности нейрональной культуры.

4. Хасанов Р. Ю. Аналитическое мышление как проблема коннекционистского подхода в когнитивной науке // Искусственные общества. – 2020. – Т. 15. – Выпуск 3 (пятилетний импакт-фактор РИНЦ: 0,377). DOI – 10.18254/S207751800011011-1.

5. Хасанов Р. Ю. Как сегодня понимают интеллект // Искусственные общества. – 2021. – Т. 16. – Выпуск 3 (пятилетний импакт-фактор РИНЦ: 0,377). DOI – 10.18254/S207751800016769-4.

II. Другие публикации соискателя по теме диссертации:

1. Хасанов Р. Ю. Изучение адаптаций нейронов для построения элементарных вычислительных процессов в нервной системе // Машины. Люди. Ценности: когнитивные и социокультурные системы в потоке времени. Материалы II Международной научной конференции, посвященной 100-летию со дня рождения доктора философских наук, профессора С. М. Шалютина (г. Курган, 22–23 апреля 2021 г.). — Изд-во Курганского гос. ун-та Курган, 2021. С. 92-97.

2. Хасанов Р. Ю. Социокультурные особенности нейроэтики // Философия. Журнал Высшей школы экономики. — 2020. — Т. 4, № 1. — С. 149–152.

3. Алексеев А. Ю., Алексеева Е. А., Хасанов Р. Ю. Машина Корсакова-Тьюринга как теоретико-алгоритмическая диспозиция символизма // Актуальные проблемы гуманитарных и социальных исследований: материалы XVIII Международной научной конференции молодых ученых. — Новосибирск: ИПЦ НГУ, 2020. — С. 37–38.

4. Алексеев А. Ю., Хасанов Р. Ю. О пользе коннекционизма для мультиагентных суперкомпьютерных исследований[3] // Искусственные общества. – 2018. – Т. 13. – Выпуск 1-2 (пятилетний импакт-фактор РИНЦ: 0,377). DOI – 10.18254/S0000121-2-1.

[3] На момент публикации статьи издание «Искусственные общества» еще не было включено в Список рецензируемых научных изданий по философским наукам, утвержденный решением Ученого совета МГУ имени М.В. Ломоносова.