

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Ковалев Максим Александрович

**Роль символических методов
в задачах общего искусственного интеллекта**

5.7.6. Философия науки и техники

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата философских наук

Москва – 2024

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

Научный руководитель:

Чусов Анатолий Витальевич, кандидат философских наук, доцент

Официальные оппоненты:

Иноземцев Владимир Александрович, доктор философских наук, доцент, ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», профессор кафедры СГН4 «Философия» факультета «Социальные и гуманитарные науки»;

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

Финн Виктор Константинович, доктор технических наук, профессор, ФГБОУ ВО «Российский государственный гуманитарный университет», философский факультет, профессор Учебно-научного центра «Проблемы и методы интеллектуального анализа данных» отделения интеллектуальных систем в гуманитарной сфере.

Защита диссертации состоится «15» мая 2024 года в 15:00 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/2883>.

Автореферат разослан «__» апреля 2024 года.

Ученый секретарь
диссертационного совета,
кандидат философских наук,
доцент

 Е.В. Брызгалина

I. Общая характеристика работы

Актуальность темы исследования

Диссертационное исследование посвящено исследованию роли символических методов как технологического подхода к задачам общего искусственного интеллекта (ИИ). В настоящее время существует практически необозримое количество попыток инженерной реализации систем искусственного интеллекта (ИИ-систем) в рамках разных технологических подходов. Задачи общего ИИ в целом можно охарактеризовать как моделирование процессов интеллектуальной деятельности человека.

Существенная и специфическая особенность ИИ-систем заключается в том, что их архитектура основана на процессах получения и переработки информации. При этом явно или неявно проводится аналогия с человеком в том аспекте, что последний обычно имеет знания, получает знания и использует знания. Здесь присутствует проблемный узел, включающий проблемы и гносеологии, и эпистемологии, и онтологии, и непосредственно инженерной реализации. В гносеологическом аспекте классической является проблема возникновения знаний у субъекта. В эпистемологическом аспекте классической является проблема возникновения у субъекта системы знаний или же проблема возникновения системного знания. В интересующем нас случае технико-технологической проблематики реализации идеальных конструкций проблема типично выражается как проблема обучения и самообучения ИИ-системы. Актуальность этой проблематики также вполне очевидна (библиографии работ, посвящённых обучению и самообучению ИИ-систем, насчитывают многие тысячи позиций).

Инженерные подходы к реализации ИИ-систем основаны на тех или иных общих теоретических предположениях о строении и структуре интеллекта. При этом исторически основные модели искусственного получения результатов преобразования знаний были основаны на аналогии с математическими вычислениями как с символическими преобразованиями. С развитием техники эта

аналогия всё упрочивалась практически до 60-х годов XX века, до тех пор, пока все большее значение ни стали приобретать исследования в области коннекционистских моделей. В настоящее время они почти полностью превалируют среди исследований в области ИИ. Однако такое смещение исследовательского интереса обусловлено не теоретическими причинами. На него повлияли как неудачи прямых реализаций символьных методов, так и появление более дешёвых и производительных вычислительных мощностей. Следствием этого сегодня стала недооцененность символьных методов в задачах общего ИИ, а также недостаточность исследований в этой области. В частности, результатом такого положения дел стал уход из поля зрения разработчиков ИИ-систем таких проблем как анализ роли тех или иных технологических подходов, перспективы их интеграции; определение границ и поиск пределов автономности ИИ, а также ряда других «предельных» проблем.

Степень разработанности темы исследования

Тема нашего исследования непосредственно сочетает проблематику общего ИИ с проблематикой технологических подходов к реализации ИИ-систем. Но философский ракурс рассмотрения расширяет эту область, в частности, посредством обращения к её действительным условиям и предпосылкам. Среди них в первую очередь следует указать на человеческое общество и человеческую деятельность (в том числе и мыслительную), без чего в известной нам реальности не возникает и вопроса о создании ИИ-систем как таковых. При этом естественно наличны разные философские уровни рассмотрения проблемы реализации ИИ-систем: онтологический, гносеологический, эпистемологический и инженерно-технический.

Проблематика исследования возможности построения систем общего ИИ (и связанная с этим, проблема того, что понимать под термином ИИ), является на сегодняшний день одной из наиболее дискутируемых тем. Среди множества работ по данной проблематике можно выделить ставшие уже классическими

работы Алана Тьюринга, Джона Серля¹, Герберта Саймона², Хьюберта Дрейфуса³, и целого ряда других исследователей.

Однако, на сегодняшний день отсутствует консенсус как относительно самой по себе возможности построения ИИ-систем, так относительно того, что понимать под термином ИИ в целом. На наш взгляд наиболее точное (с точки зрения настоящей работы) понимание термина «ИИ» дается в работах В.К. Финна⁴. Аналогичная ситуация складывается и относительно терминов «общий ИИ» и «сильный ИИ».

Одним из основных путей решения вышеуказанных проблем заключается не в общем определении того, «что есть ИИ?», и не в сформулированной в общем виде практической постановке задачи: «можно ли построить систему общего (сильного) ИИ?». Гораздо более значим и продуктивен поиск ответа на вопрос о возможном статусе ИИ-систем. Такой вопрос можно сформулировать следующим образом:

Как можно построить систему общего ИИ в условиях её функционирования в человеческом обществе?

Такая постановка вопроса необходимо связана со следующими вопросами:

- 1) определение возможной степени автономности ИИ-систем;
- 2) определение места и значимости символьных, коннекционистских и др. методов в области ИИ;
- 3) установление теоретических и практических пределов возможностей самообучения в задачах общего ИИ;
- 4) каким образом должно осуществляться получение ИИ-системами первичных знаний.

¹ Searle, J. Minds, brains, and programs // Behavioral and brain sciences., 1980., Т. 3, № 3 (September), P. 417-424

² Науки об искусственном., М.: Книга по требованию, 2013

³ Dreyfus H., Why Heideggerian AI failed and how fixing it would require making it more Heideggerian // Artificial Intelligence, Volume 171, Issue 18, 2007, pp 1137-1160, <https://doi.org/10.1016/j.artint.2007.10.012>.

⁴ Финн В.К. Искусственный интеллект: Методология, применения, философия., М.:КРАСАНД, 2011

Поставленные таким образом вопросы подчеркивают необходимость философского осмысления системы взаимодействий искусственных объектов с миром и человеком в фокусе различий между опытным освоением предметной области и овладением опытом, учитывая существование разных типов опыта⁵:

- «опыт готовых знаний», где важен объяснительно-иллюстративный процесс,
- «опыт способов деятельности» с его репродуктивной (тренинговой) направленностью,
- «опыт творческий», имеющий проблемную направленность.

В том или виде данные вопросы, рассматривались, такими исследователями как Хьюберт Дрейфус, В.Ф. Венда⁶, Ж.-Л. Лорьер⁷ и целым рядом других исследователей. Однако, данные исследования лишь частично касаются проблематики непосредственно процесса обучения. В частности, речь идет о недостаточном внимании, по мнению В.В. Краевского⁸, к «событийности» процесса обучения и обязательности наличия двух оппозиций: «ученик-учитель», «ученик-учебный материал».

Отдельными проблемами являются вопросы самостоятельного принятия решений ИИ-системами, выбора рационального решения из множества доступных альтернатив и формализации условий принятия решений. Среди наиболее интересных исследований в этой области можно указать работы С.И. Шапиро⁹, Ж.-Л. Лорьера, Д.А. Поспелова¹⁰, В.А. Смирнова¹¹, Е.Д. Смирновой¹², В.К. Финна и ряда других исследователей.

⁵ Сериков В.В., Станет ли педагогика наукой? Размышления о методологии В.В.Краевского // Отечественная и зарубежная педагогика. 2016. №5 (32).

⁶ Венда В.Ф., Системы гибридного интеллекта: Эволюция, психология, информатика., М.:ЛЕНАНД, 2020

⁷ Лорьер Ж.-Л., Системы искусственного интеллекта., М.:Мир, 1991

⁸ 94. Краевский В.В., Методология педагогики: новый этап: учеб. Пособие для студ. высш. учеб. заведений / В.В. Краевский, Е.В. Бережнова, М.: Издательский дом «Академия», 2006

⁹ Шапиро С.И., От алгоритмов – к суждениям (Эксперименты по обучению элементам математического мышления), М.: Сов. Радио. 1973

¹⁰ Поспелов Д.А., Ситуационное управление: Теория и практика. М.:ЛЕНАНД, 2021.

¹¹ Смирнов В.А., Символическая логика и теория познания // Философия и социология науки и техники, Ежегодник 1983, М., 1985., с. 45-60

¹² Смирнова Е.Д., Логическая семантика и философские основания логики., М.:ЛЕНАНД, 2021

Ключевыми проблемами понимания места тех или иных технологических подходов в задачах общего ИИ занимались такие исследователи как Д.А. Поспелов, Педро Домингос¹³, Маркус Габриель¹⁴, В.Г. Кузнецов¹⁵, В.А. Успенский¹⁶. На наш взгляд, основным недостатком этих работ является недостаточное внимание к вопросам объединения разного рода технологических подходов единую ИИ-систему, учитывающую преимущества и недостатки каждого технологических подходов.

Отдельной трудностью является наблюдаемая последние десятилетия устойчивая тенденция к изолированию исследовательских программ в рамках того или иного технологического подхода¹⁷. В связи с чем, по нашему мнению, сегодня необходима корректировка этих исследовательских программ. Смещение фокуса исследований в область создания гибридных моделей является более плодотворной задачей в долгосрочном периоде, чем игра в имитацию или развитие какого-то одного из технологических подходов.

Вместе с тем, существующий сегодня список исследований в области создания гибридных систем, явным образом включающих человека в разработку технологических подходов, весьма ограничен. Примерами таких исследований могут служить работы Педро Домингоса, В.Н. Гридина и В.И. Солодовникова¹⁸, Гэри Маркуса¹⁹.

На наш взгляд, центральной проблемой на пути создания систем общего ИИ является техническая проблема инженерной реализации процессов генезиса новых понятий, знаков и правил. Именно эти вопросы являются ключевыми в

¹³ 74. Домингос П., Верховный алгоритм: как машинное обучение изменит наш мир. М.: Манн, Иванов и Фербер, 2016.

¹⁴ 62. Габриель М., Я не есть мозг: Философия духа для XX века., М.:УРСС: ЛЕНАНД, 2020

¹⁵ 95. Кузнецов В.Г., Парадигмальный анализ теории искусственного интеллекта //Философия искусственного Интеллект. Труды всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта, 17-18 марта 2016 г, М.:ИИнтелл, 2017.

¹⁶ Успенский В.А., труды по нематематике, Т.3, М.: ОГИ : Фонд «Математические этюды», 2012

¹⁷ Поспелов Д.А., Десять «горячих точек» в исследованиях по искусственному интеллекту // Искусственный интеллект и принятие решений, М., 2019., № 4

¹⁸ Гридин В.Н., Солодовников В.И., Совместное использование нейросетевых технологий и деревьев решений для поиска логических закономерностей в данных. / В. Н. Гридин, В. И. Солодовников // Информационные технологии и нанотехнологии. 2017. С. 1756–1762.

¹⁹ Marcus, G. F., The Algebraic Mind: Integrating Connectionism and Cognitive Science. Cambridge, MA: MIT Press. 2001

исследовании роли символьных методов в задачах общего ИИ, что подчёркивали Фредерик Хейес-Рот, Дональд Уотерман и Дуглас Ленат, говоря о принципиальной значимости для этой технической проблемы не формальных методов рассуждения, а знаний как таковых и проблематики способов их получения²⁰.

Самому по себе исследованию различных аспектов генезиса новых знаков, понятий и правил посвящено огромное количество работ. Среди них следует отметить работы К.-О.Апеля²¹, рассматривающего проблему фиктивности и реальности поэтического акта. Людвиг Витгенштейн²², исследовавший связь между рациональностью и символической языковой природой нашего сознания. Ю.М. Лотман²³ указывает на наличие «асимметрии» семиотических структур адресата и адресанта как необходимого условия возникновения новизны, что является одним из триггеров, запускающих процессы генезиса новых понятий, знаков и правил. Принципиальное значение для нашего исследования имеют понятие «нонсенса» Жюль Делеза²⁴, а также исследования В.А. Смирнова²⁵ в сфере построения концептуальных систем, как в случае их построения человеком, так и в случае ИИ-систем.

Однако, по нашему мнению, указанные работы носят главным образом теоретический характер и не уделяют достаточное внимание практическим (инженерным) аспектам реализации процессов генезиса новых знаков, понятий и правил. Недостаточное внимание было обращено и на сравнительное исследование роли и места тех или иных технологических подходов. Следствием этого является недооцененность символьных методов в качестве основных механизмов, обеспечивающих генезис новых понятий, правил и знаков.

Таким образом диссертационное исследование позволяет восполнить пробел, связанный с вышеуказанной проблематикой.

²⁰ Хейес-Рот Ф., Уотерман Д., Ленат Д., Построение экспертных систем, М.: Мир, 1987

²¹ Апель К.-О., Трансформация философии., М.: «Логос», 2001

²² Витгенштейн Л., Логико-философский трактат., М.: «Канон+» РООИ «Реабилитация», 2011

²³ Лотман Ю.М., Внутри мыслящих миров, СПб.: Азбука, Азбука-Атикус, 2014

²⁴ Делез Ж. Логика смысла, М.: Академический проект, 2015

²⁵ Логико-философские труды В.А. Смирнова . Под ред. В.И. Шалака, М.:Эдиториал УРСС, 2001

Сказанное позволяет сформулировать общую **проблему** изучения роли символьных методов в задачах общего ИИ. Основными для нашего исследования при этом являются онтологический, гносеологический, эпистемологический и инженерный аспекты.

Цель исследования

Определение роли символьных методов в задачах общего искусственного интеллекта.

Задачи исследования

1. Описание основных этапов процесса исторического развития исследовательских программ в области ИИ.
2. Определение онтологического статуса существования ИИ-систем, сущности техники и граничных условий инженерной реализации ИИ-систем, а также определение граничных условий существования ИИ-систем с точки зрения возможности автономии ИИ-систем в процессах самообучения и самостоятельного принятия решения в ходе их функционирования.
3. Определение основных ролей символьных методов в задачах общего ИИ.

Объект и предмет исследования

Объект исследования – многообразие исследований, подходов и разработок в области концептуализации и реализации искусственного интеллекта.

Предмет исследования – символьные методы в общем контексте задач концептуализации и технической реализации систем общего искусственного интеллекта.

Методология исследования

В ходе исторического анализа исследований в области общего искусственного интеллекта и символьных методов, а также в ходе построения теоретической модели концептуальной структуры процедур генезиса новых понятий, знаков и правил, применялись элементы системного,

герменевтического, структурного и конструктивного подходов. Использовались некоторые методы теории информатики.

Положения, выносимые на защиту

1. ИИ является техническим объектом, готовым к использованию и «подручным» человеку. Первоначальное обучение ИИ-системы невозможно без человека. Анализ того, как должен быть устроен общий ИИ, должен проводиться в рамках прежде всего гибридных человеко-машинных концепций.
2. Граничные условия, которым должен удовлетворять процесс самообучения, и нарушение которых приводит к практической невозможности продолжения существования ИИ-системы или к потере ценности получаемых ею результатов, непосредственно задаются человеческими концептуальными системами.
3. Эффективная система общего ИИ должна с необходимостью представлять собой систему гибридного интеллекта, включающую в себя человека, коннекционистские и символьные методы.
4. Роль символьных методов определяется тем, что они реализуют эффективные инструменты и механизмы, обеспечивающие:
 - a. процессы (само)обучения ИИ-систем путем генезиса или трансформации существующих понятий, знаков и правил в том числе с использованием обучающих выборок;
 - b. процессы коммуникации в рамках моделей «человек-машина» и «машина-машина».

Научная новизна исследования

Решая проблему машинного самообучения, входящую в ряд наиболее критически важных проблем создания ИИ-систем, впервые установлен ряд необходимых философских и методологических оснований проектирования творческих механизмов в рамках ИИ-систем за счет: 1) переосмысления роли символьных методов; 2) обоснования принципиальной возможности средствами

символьных методов выхода за пределы существующего сегодня замкнутого круга аналогия – операция обобщения.

В настоящем диссертационном исследовании впервые:

1. Определена необходимая роль символьных методов – обеспечение коммуникации как между машиной и человеком, так и между машинами. Символьные методы могут частично выполнять роль «учителя» в задачах обучения коннекционистских (сетевых) моделей.
2. Определены граничные условия, которым должен удовлетворять процесс самообучения, нарушение которых приводит к невозможности эффективного функционирования ИИ-системы или потере ценности полученных результатов.
3. Обоснованы тезисы о невозможности и нецелесообразности построения полностью автономных систем общего ИИ, а также о приоритетности развития гибридных, человеко-машинных ИИ-систем.
4. Разработана методология, обеспечивающая процедуры генезиса новых представлений, знаков и правил. В частности, разработаны базовые правила, позволяющие в значительной степени расширить возможности символьных систем в области производства нового знания.

Теоретическая и практическая значимость результатов исследования

Основной результат представленной работы – теоретическая модель, которая может быть положена в качестве основания для разработки методологии, обеспечивающей процессы обучения как с учителем, так и без в рамках символьных методов на базе обучающих выборок аналогично тому, как это сегодня реализуется в рамках коннекционистских подходов.

Определена роль символьных методов как источника новых понятий, знаков и правил.

Кроме того, значимость результатов исследования в теоретическом плане определяется тем, что на их основании:

- 1) могут быть доопределены пределы творческих возможностей систем машинного интеллекта;

- 2) определена роль человека как необходимого и ограничивающего ценность результатов обучения ИИ-системы условия;
- 3) определены условия необходимости реализации устройства ИИ-систем в парадигме гибридного ИИ.

В практическом плане значимость исследования определяется тем, что разработанную теоретическую модель можно реализовать как методологию, позволяющую расширить творческие возможности систем машинного интеллекта посредством обеспечения процессов обучения в ситуациях недостатка и/или отсутствия достаточного объема эмпирических данных.

Степень достоверности и апробация результатов исследования

Достоверность полученных результатов достигается автором при помощи применения научной методологии и принципа объективности. Диссертационное исследование опирается на широкий круг источников и исследовательскую литературу на русском, английском и немецком языках.

Диссертация была обсуждена и рекомендована к защите на заседании кафедры философии и методологии науки философского факультета Московского государственного университета имени М.В. Ломоносова.

Основные положения и выводы диссертационного исследования были изложены в 4-х научных статьях, опубликованных в изданиях, отвечающих требованиям п. 2.3 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова.

Основные результаты диссертационного исследования представлены в рамках научно-практических конференций и семинаров:

1. Международная конференция «Суперкомпьютерные дни в России» (Москва, 24.09.2019) – доклад «Суперкомпьютерные системы в задачах распознавания смысла текста»;
2. Международная конференция «Суперкомпьютерные дни в России» (Москва, 22.09.2020) – доклад «Роль символического подхода в задачах общего искусственного интеллекта»;

3. Международная междисциплинарная конференция «Искусственный интеллект в новой коммуникативной реальности» (Москва, 06.11.2020-07.11.2020) – доклад «От гибридных интеллектуальных систем к гибриднему интеллекту»;
4. Научный семинар «Методологические проблемы современной науки: эпистемология научных объяснений» (Москва, МГУ имени М.В. Ломоносова, 20.10.2021) – доклад «Является ли ИИ чем-то большим, чем техника?»;
5. Всероссийская научно-практическая конференция с международным участием «Философские контексты современности: искусственный интеллект и интеллектуальная интуиция» (Ижевск, 25.02.2022-26.02.2022) – доклад «Онтологические проблемы в задачах общего искусственного интеллекта»;
6. XVIII конгресс «Нейронаука для медицины и психологии» (Судак, 30.05.2022-10.06.2022) – доклад «Гносеологические проблемы в задачах общего искусственного интеллекта».

Структура диссертационного исследования

Диссертационная работа состоит из введения, трех глав, заключения, одного приложения и списка литературы. Текст диссертации составляет 247 страниц.

II. Основное содержание работы

Во **Введении** обосновывается актуальность выбранной темы диссертации, определяются цели и задачи исследования. Вводятся основные понятия и термины:

Сильный ИИ

В рамках настоящего исследования под **«сильным ИИ»** будем понимать определение Джона Сёрля ИИ-системы как системы,

демонстрирующей «точно такое же поведение, которое человек демонстрирует в отношении некоторого когнитивного феномена»²⁶ и в буквальном смысле слова обладающей когнитивными состояниями²⁷.

Общий ИИ

На сегодняшний день не существует устоявшегося определения термина «общий ИИ». На наш взгляд, существующее сегодня понимание «сильного ИИ» является в значительной степени размытым и требующим уточнения. В рамках настоящего исследования ограничим понимание сильного ИИ, отделяя это понятие от термина **«общий ИИ»**, под которым понимается

система (инженерное решение), моделирующая интеллектуальную деятельность человека, в которой **реализованы процессы** обучения и самообучения (производство нового знания) в целях принятия самостоятельных рациональных решений понимаемых как выбор одного из вариантов действий среди множества возможных (а также смоделированных в качестве прогнозных самой ИИ-системой) альтернатив в условиях неполноты исходной и текущей информации, обусловленной динамично изменяющимися условиями внешней среды.

Гибридный ИИ.

В рамках настоящего исследования термин **«гибридный ИИ»** понимается прежде всего как человеко-машинная система в русле идей В.Ф. Венда.

²⁶ Parsing the Turing Test, Springer Science + Business Media B.V. 2008, p. 140

²⁷ Searle, J. Minds, brains, and programs // Behavioral and brain sciences., 1980., Т. 3, № 3 (September), P. 417-424.

В настоящее время существует два понимания термина «гибридный ИИ». Во-первых, как ансамбль различных технических подходов (в работах Педро Домингоса, Гэри Маркуса и т.д). Во-вторых, как человеко-машинная система, как ее понимал В.Ф. Венда.

Символьные методы

На абстрактном теоретическом уровне под **«символьными методами»** будем понимать процедуры манипулирования символами как многоуровневыми представлениями в понимании этого Аланом Ньюэллом и Гербертом Саймоном²⁸.

В аспекте инженерных реализаций, под **«символьными методами»** будем понимать совокупность инженерных решений, реализующих процедуры создания и преобразования символьных выражений.

Примерами такого рода инженерных решений являются экспертные системы Фредерика Хейес-Рота, Дональда Уотермана и Дугласа Лената, ДСМ-метод В.К. Финна и т. д.

Коннекционистские методы

Под **«коннекционистскими методами»** будем понимать совокупность технологических подходов, включающих в себя алгоритмы, основанные на сетевых моделях как моделях архитектуры мозга.

В настоящее время существует большое количество разнообразных моделей, таких как нейронные сети Джона Хопфилда, Ричарда Хэмминга, «машина Больцмана» Джеффри Хинтона и Терри Сайновски, сверточные нейронные сети Яна Лекуна и т. д.

Роль символьных методов

«Роль» будем понимать как определенный набор действий, которые может выполнять тот или иной элемент-актор внутри той или иной системы.

«Роль» также часто понимают как степень значимости тех или иных элементов-актеров относительно друг друга и того вклада, который каждый из

²⁸ 26. Newell A. and Simon H., Computer science as empirical inquiry: Symbols and search. Communications of the ACM 19:113-126, 1976.

них вносит в общий результат. В этом смысле говорят о «ведущей», «подчиненной» или «второстепенной» роли.

Технологический подход

Под «технологическим подходом» будем понимать множество инженерных решений, реализующих символичные, коннекционистские и т.д. методы.

«Подручность»

Впервые понятие подручности вводится Мартином Хайдеггером через обращение Dasein (Человека) с внутримирным сущим. Средство – как вещь – не просто «бывает» в смысле моментального присутствия определённого нечто в человеческом окружении, а существует в несамостоятельном, связанном с присутствием человека модусе «для того чтобы». «Сработанное изделие», по Хайдеггеру, отсылает, в том числе, к своему пользователю. Именно по «телу» пользователя происходит «выкраивание» изделия²⁹.

Однако, такое понимание подручности не является достаточным. Для философа подручность связана с «сакральным» и «неподлинным», где существование техники рассматривается из перспективы индивидуальной экзистенции человека, из базовой включенности человека как в мир людей, так и в мир вообще. В рамках такого понимания невозможно понять место ИИ-системы как некоторого субъекта, более активно (нежели «просто» машины и инструменты), включенного в различного рода взаимодействия. В связи с недостаточностью такого понимания подручности, нами предлагается собственное понимание термина «подручность» как **способ существования ИИ-системы** в человеческом мире как системы, автономной в широком диапазоне условий и задач, причём действия этой системы зависят от человека и человеческого мира в вопросах обучения, самообучения и принятия решений в ходе своего функционирования.

²⁹ Хайдеггер М. Бытие и время. М.:Академический проект, 2015. С.70.

Знание

Под «знаниями» в рамках настоящего исследования будем понимать произведенные в результате процессов генезиса и поэзиса новые/измененные понятия знаки (символы) и правила вывода, организованные в единую концептуальную систему, понимаемую в дальнейшем как систему знаний того или иного экземпляра системы общего ИИ.

«Познавательная деятельность» понимается как процесс производства знания некоторым субъектом (человеком, ИИ-системой).

Под «опытом» будем понимать знания, полученные в результате познавательной деятельности. Производство знания предполагается на основании поступающей извне или уже имеющейся у системы информации.

В случае системы машинного интеллекта речь идет о формировании некоторого экземпляра концептуальной системы, реализуемого средствами программно-аппаратных комплексов. В рамках настоящего исследования примем следующие положения:

Обучение и самообучение

По мнению В.В. Краевского, следует проводить четкое различие между наукой, как деятельностью «в которой происходит выработка и теоретическая систематизация объективных знаний о действительности»³⁰ и педагогической деятельностью, направленной на передачу опыта, в которой наличие пар «ученик-учитель» и «ученик-учебный материал» является обязательной. Важно отметить, что, по его мнению, процесс обучения имеет в значительной степени нормативный характер.

Таким образом, процесс самообучения носит принципиально иной характер относительно процесса обучения. Соглашаясь с пониманием ИИ В.К. Финном как веберовского типа интеллекта, деятельность ИИ-системы в области самообучения может быть понята прежде всего как рациональная, соответствующая принципам научного познания, деятельность.

³⁰ Краевский В.В., Методология педагогики: новый этап: учеб. Пособие для студ. высш. учеб. заведений / В.В. Краевский, Е.В. Бережнова. М.: Издательский дом «Академия», 2006. С.7.

Под «обучением» будем понимать процесс передачи от учителя к ученику (ИИ-системе) опыта в форме знаний, в котором участвуют не менее двух субъектов.

Под «самообучением» будем понимать процесс производства нового знания, производимый одним субъектом.

Понятия. Знаки. Продукционные правила.

Под «понятием» в рамках настоящего исследования будем понимать теоретические объекты как элементы концептуальных структур, порождаемые путем идеализации и схематизации в результате процесса теоретической «реконструкции фрагментов действительности, взятых с определенных сторон и определенным образом схематизированных»³¹.

Под «знаками» будем понимать символные выражения и имена, присвоенные полученным теоретическим объектам, с учетом правил грамматики того языка, на котором осуществляется именование.

Под «правилами» в рамках настоящего исследования будем понимать пропозиционные высказывания, реализуемые теми или иными программно-аппаратными средствами.

Примерами формальных систем являются λ -исчисления, модель Крипке для интуиционистской логики высказываний, другие системы неклассической логики.

Далее определяется объект и предмет исследования, указывается теоретическая база и методология исследования. Приводятся положения, обосновывающие новизну исследования, формулируются положения, выносимые на защиту, отмечается теоретическое и практическое значение диссертации.

Первая глава «Исторический анализ процессов возникновения и эволюции понимания роли символьных методов в задачах проектирования

³¹ Смирнов В.А., Уровни знания и этапы процесса познания // Проблемы логики и научного знания., М., 1964. с. 23-52.

ИИ-систем» посвящена истории возникновения как ИИ вообще, так и истории развития символьных методов.

В параграфе ««Дофизический» период» рассматривается период истории начиная с Пифагора и до возникновения первых прототипов интеллектуальных физических систем, таких как машины Беббиджа³² и Корсакова³³. Таким образом этот период охватывает порядка 25 столетий начиная с VI в.д.э. и продолжается до XIX века н.э.. Особенностью данного периода являются поиск основополагающих принципов, позволяющих моделировать внутренние процессы мышления, присущие человеку.

В параграфе «Первые реализации на физическом субстрате. «Физический» рассматривается период, продолжавшийся чуть больше одного столетия вплоть до появления в середине 1940-х первых электронных машин, таких как «Bomb» Тьюринга³⁴, ENIAC от IBM³⁵ и т.д. Данный период ознаменован первыми относительно успешными попытками создания «интеллектуальных» машин, иными словами, физических систем, способных воспроизводить определенные интеллектуальные операции самостоятельно, хотя и после предварительной настройки оператором. Создание таких машин если и не изменило хода истории в тот период, но в известной степени повлияло на ход философской мысли так как подтверждало идеи Декарта и Гоббса о том, что мыслительные процессы не отличаются от механических. Характерной особенностью второго периода необходимо считать возникновение нового феномена – первых механических вычислительных устройств, значительно усиливших картезианские тенденции в философии науки. К концу этого периода, были сформированы все основные вопросы необходимые для создания интеллектуальных физических систем³⁶:

³² Петренко А. К., Петренко О. Л. Машина Беббиджа и возникновение программирования // Историко-математические исследования.,1979., т. 24., с. 340.

³³ Алексеев А. Ю. Машина Корсакова (1832 г.) как прототип мультиагентного суперкомпьютерного автомата // Искусственные общества. – 2019. – Т. 14. – Выпуск 1.

³⁴ Wright J. The Turing Bombe Victory and the first naval Enigma decrypts, Cryptologia, 2017.

³⁵ McCarthy J., What is artificial intelligence? 2007, <http://www-formal.stanford.edu/jmc/whatisai/>

³⁶ 132. Рассел С. Норвиг П., Искусственный интеллект: современный подход., М.: ООО «И.Д. Вильямс», 2016. С.43.

1. Каковы формальные правила формирования правильных заключений?
2. Как определить границы вычислимости?
3. Как проверить рассуждения с использованием недостоверной информации?

Даны если и не исчерпывающие, то достаточно полные ответы на эти вопросы, имеющие первостепенное значение для создания искусственного интеллекта и реализации символьного подхода, в частности.

В параграфе «Компьютерный период» рассматривается самый короткий период, с середины 1940-х и продолжающийся до сегодня. Данный период ознаменовывается появлением электронных вычислительных устройств, что привело к научно-технической революции середины XX века и смене технологического уклада, свидетелями которых мы являемся в настоящее время.

В параграфе «Символьные методы и исследовательские программы ИИ» исследуется история развития непосредственно символьных методов.

Понятие символьного подхода базируется на: 1) понятиях символьных методов и компьютерной алгебры, 2) гипотезе Ньюэлла-Саймона.

В отличие от численных методов, заключающихся в непосредственной обработке вычислений, символьные методы оперируют равенствами и формулами как последовательностями символов. Исходя из тезисов гипотезы Ньюэлла-Саймона, физическая символьная система имеет все необходимое для решения задач, аналогично тому, как это осуществляет человеческий мозг. Тот тезис, что всякая вычислимая функция вычислима с машиной Тьюринга, звучит так: «все частичные функции, вычисляемые посредством алгоритмов, являются частично рекурсивными» (в терминах Клини³⁷).

Пик интереса к развитию символьных методов относится к концу XX века в связи с разработками экспертных систем³⁸. В России особый интерес представляет метод автоматического порождения гипотез путем извлечения

³⁷ Клини С. К. Математическая логика. — М.: Мир, 1973., с.430

³⁸ Уотерман Д., Руководство по экспертным системам. М.: Мир, 1989

причинно-следственных зависимостей из базы фактов и использования базы знаний посредством формальной эвристики вида «индукция-аналогия-абдукция» (ДСМ-метод В.К. Финна).

В параграфе «Проблемы современного дискурса» представлены основные тенденции, определяющие направление современного дискурса в области ИИ.

В нем обсуждаются понимание общего ИИ в связи с тестом Тьюринга, а также перспективы реализации общего ИИ с использованием исключительно коннекционистских методов.

На основе теста Тьюринга Джон Сёрль дает свое понимание термина «сильный ИИ». Основным критерием, который, с его точки зрения, мог бы классифицировать ИИ-систему как «сильную», является способность последней пройти тест Тьюринга. Сам философ относится довольно критически к такой постановке вопроса³⁹. В настоящее время существует несколько десятков вариантов теста Тьюринга, например, «signal-тест» (MIST) Криса МакКинстри⁴⁰, обратный тест Тьюринга или Completely Automated Public Turing Test to Tell Computers and Humans Apart⁴¹ и т.д.

Одним из главных недостатков этого теста, на наш взгляд, является то, что он выявляет не собственную разумность поведения ИИ-системы, а успешность поддержания иллюзии общения с живым человеком у членов жюри в течение определенного времени. Таким образом, тест предполагает реализацию не только и не столько рационального поведения, сколько воспроизведение иррациональной стороны поведения. Примером тому может быть анализ победителей конкурса Лебнера, где, например, наличие ошибок специальным образом закладывалось при реализации конкурсных программ. Такая постановка вопроса только уводит в сторону от задачи по созданию систем общего ИИ.

На наш взгляд, одним из основных критиков Теста является сам Тьюринг. В частности, речь идет о 4-м пункте возражений относительно сознания. В

³⁹ Parsing the Turing Test, Springer Science + Business Media B.V. 2008, pp. 139-151.

⁴⁰ <http://hps.elte.hu/~gk/Loebner/kcm9512.htm>.

⁴¹ Blum M., Ahn von L. A., Langford J., —The CAPTCHA Project, Completely Automatic Public Turing Test to Tell Computers and Humans Apart, | Nov. 2000.

качестве такой точки зрения Тьюринг приводит слова профессора Джефферсона о возможности написания машиной сонета, основанной не за счет «случайного совпадения символов»⁴², а в результате собственных побуждений.

Кроме того, необходимо отметить, что на протяжении всей истории своего развития символьные методы сталкивались с значительной критикой⁴³, главным образом связанной не с реализацией подхода, а со сложностью обучения ИИ-систем на его основе. Все попытки технического решения проблемы, без обращения к философскому анализу, не привели к желаемым результатам. Одним из первых, кто заговорил о необходимости именно философского переосмысления проблемы – сквозь призму экзистенциализма – был Хьюберт Дрейфус, предложив концепцию «хайдеггерианского ИИ»⁴⁴.

Даже беглый анализ списков публикаций и докладов основных мировых конференций в области ИИ говорит о недооценке роли философии в области проектирования ИИ-систем. Об этом наглядно свидетельствует анализ докладов как зарубежных, так и российских конференций таких как «Annual AGI Conference»⁴⁵, «opentalks.ai»⁴⁶ и т.д. В качестве характерных примеров можно привести обзорный доклад И.О. Пивоварова⁴⁷ на конференции «opentalks.ai» 2022 года и недавнюю работу С.А. Шумского⁴⁸ «Машинный интеллект. Очерки по теории машинного обучения и искусственного интеллекта».

Упрощенное, трактуемое исключительно в бихевиористском ключе, понимание авторами процессов обучения приводит, в частности, к ложным выводам о достаточности коннекционистских методов для инженерной реализации систем сильного ИИ. Критика бихевиоризма достаточно

⁴² Turing A., Computing machinery and intelligence, Mind 1950; LIX (236): 433-460 doi:10.1093/mind/LIX.236.433

⁴³ Дрейфус Х., Чего не могут вычислительные машины: Критика искусственного разума., М.: ЛИБРОКОМ, 2010.

⁴⁴ Dreyfus H., Why Heideggerian AI failed and how fixing it would require making it more Heideggerian // Artificial Intelligence, Volume 171, Issue 18, 2007, pp 1137-1160, <https://doi.org/10.1016/j.artint.2007.10.012>.

⁴⁵ <https://agi-conf.org/>

⁴⁶ <https://opentalks.ai.ru/>

⁴⁷ <https://www.youtube.com/watch?v=u066n1mFJO4>

⁴⁸ Шумский С.А., машинный интеллект. Очерки по теории машинного обучения и искусственного интеллекта., М.:РИОР, 2019.

представлена в работах Дж. Сёрля⁴⁹, А.Р. Лурия⁵⁰, Б.М. Величковского⁵¹, Хьюберта Дрейфуса⁵² и других.

Вторая глава «Анализ онтологического статуса ИИ» посвящена анализу онтологического статуса ИИ. Такого рода анализ необходим, во-первых, для понимания того, как, каким образом может существовать система общего ИИ. Во-вторых, каковы границы автономности системы ИИ-систем.

В параграфе «Сущность техники, пределы и граничные условия функционирования технических объектов» проводится анализ взглядов на технику Э. Каппа, П.А. Флоренского, П.К. Энгельмайегра, О. Шпенглера, Х. Ортега-и-Гассета, К. Ясперса, М. Хайдеггера, Ф. Дессауэра в онтологическом аспекте.

Наиболее важными выводами, которые можно сделать из развития концепций техники этими авторами, являются:

1. Признание особого рода финализма технических объектов, связанного с тем, что форма задачи изобретения человеком полностью определена.
2. Признание того факта, что технические объекты выходят за рамки природных сил, оказывая влияние на социальное, на человека.
3. Признание неспособности технических объектов к решению творческих задач, к изобретательству по ряду объективных причин, и, прежде всего, из-за недоступности области человеческих потребностей для машин.
4. Признание опасности «подчинения» техникой человека. Но подчинение это носит со стороны техники скорее пассивный характер, обусловленный степенью вовлеченности человека в глобальный производственный процесс, добровольным отказом со стороны самого человека от собственной свободы.

⁴⁹ Searle, J. Minds, brains, and programs // Behavioral and brain sciences.,1980., Т. 3, № 3 (September), Р. 417-424.

⁵⁰ Лурия А.Р., Язык и сознание.,СПб.: Питер, 2021

⁵¹ Величковский Б.М., Когнитивная наука. Основы психологии познания. В 2 томах., Т. 2., М.:Издательство Юрайт, 2022

⁵² Дрейфус Х., Чего не могут вычислительные машины: Критика искусственного разума., М.: ЛИБРОКОМ, 2010.

В параграфе «Место техники в социальном» дается анализ проблемы места человека в созданном им самим искусственном мире, в котором техника начинает играть все более значимую роль. К такого рода исследователям относятся такие философы как Жильбер Симондон, Мишель Серра, Бернар Стиглер и целый ряд других философов.

Основными выводами из исследования места техники в социальном мире являются:

1. Все более входя в социальное, проникая в те области, в которых до этого было место лишь человеку, техника, тем не менее, остается «подручной» человеку.
2. Оставаясь «подручной», техника сегодня является одним из конституирующих современную культуру начал, приводя к росту алгоритмизации мыслительных процессов человека.
3. Происходящие изменения следует скорее отнести к изменениям в способах и организации коммуникативных и мыслительных процессов самого человека, что не отрицает опосредованного влияния техники на культурные, социальные и ментальные процессы в обществе.

В параграфе «Критика основных технологических подходов» дается критический анализ сильных и слабых сторон основных технологических подходов, таких как коннекционистские и символные методы.

В параграфе «Критика возможности выхода ИИ за границы техники» проводится анализ оснований, которые могли бы свидетельствовать о возникновении феномена выхода ИИ-систем за пределы «подручности». В рамках настоящего исследования данный вопрос является одним из наиболее важных, так как ответ на него является, в том числе, ключом к пониманию роли, пределов возможностей и граничных условий процессов самообучения и принятия решений в ходе самостоятельного функционирования ИИ-систем. Понимание этого необходимо для ответа на вопрос «каким образом, могут существовать символные методы?».

В результате проведенного исследования делается вывод о том, что пока что нельзя говорить о возможности выхода ИИ-систем за пределы «подручности». Прежде всего речь идет о сформулированной Дрейфусом проблеме «хайдеггерианского ИИ».

Общие выводы об онтологическом статусе ИИ-систем:

1. ИИ-системы являются техническими объектами, и, как технические объекты, «подручны» человеку;
2. На сегодняшний день отсутствуют условия для выхода ИИ-систем за пределы «подручности».
3. Оставаясь «подручной», техника, и ИИ-системы, в частности, сегодня является одним из конституирующих современную культуру начал, приводя к росту алгоритмизации мыслительных процессов человека.
4. Происходящие сегодня изменения скорее следует отнести к изменениям в способах и организации коммуникативных и мыслительных процессов самого человека, не отрицая опосредованного влияния техники на процесс культурных, социальных и ментальных процессов в обществе.

В параграфе «Обоснование необходимости гибридного подхода как следствия вывода о «подручности» ИИ» дается анализ и объяснение искусственного интеллекта как погруженного в мир человека. На наш взгляд, без такого рода анализа невозможно ни создание прочной онтологической и эпистемологической базы исследований, ни точного формирования самой постановки задачи при проведении каких-либо исследовательских программ в области создания систем общего ИИ.

Основным выводом, полученным в ходе проведенного исследования, является тезис о том, что любая система ИИ, приближающаяся к имитации рационального поведения, должна быть системой гибридного ИИ.

В параграфе «Определение граничных условий существования ИИ-систем» дается анализ граничных условий существования ИИ-систем как «подручных» человеку. Исследуя проблемы самообучения ИИ-систем в условиях автономности принятия ими решения в ходе их функционирования,

возникает вопрос об оценке получаемых решений сточки зрения их ценности и применимости с учетом ранее полученного вывода об относительной несамостоятельности техники.

Отвечая на поставленные вопросы, мы приходим к выводам о том, что граничные условия, которым должен удовлетворять процесс самообучения, нарушение которых приводит к невозможности существования ИИ-системы или к потере ценности получаемых ими результатов, задают концептуальные системы человека (в понимании В.А. Смирнова).

В главе 3 «Основные аспекты применимости символьных методов при проектировании ИИ-систем» проводится анализ роли символьных методов в задачах общего ИИ. Прежде всего, необходимо сформулировать проблему, решение которой и могло бы являться ответом на поставленный вопрос о роли символьных методов. В нашем случае проблему можно сформулировать как: Проблему необходимости и достаточности символьных методов для построения системы общего искусственного интеллекта. В связи с этим, важно теоретически исследовать такие аспекты поставленной проблемы, как:

1. Эпистемологический. Существует ли принципиальная возможность инженерной реализации актов непосредственного генезиса новых понятий, знаков и правил вывода как необходимых условий, конституирующих возможность построения собственных концептуальных структур ИИ-системой;
2. Гносеологический. Существует ли принципиальная возможность обеспечения автономности процессов первичного генезиса новых понятий, знаков и продукционных правил со стороны ИИ-системы;
3. Онтологический. Как (каким образом) могут существовать символьные методы в рамках решения проблем 1 и 2.

В параграфе «Эпистемологический и гносеологический аспекты» рассматриваются проблематика генезиса новых понятий и знаков, а также проблематика генезиса новых продукционных правил. Основными выводами

относительно возможности генезиса новых понятий — элементов концептуальной системы являются:

1. Факт обнаружения несоответствия между текущим состоянием концептуальной структуры ИИ-системы и поступающей извне информации является необходимым условием, активирующим процессы генезиса новых понятий.
2. Первичный акт генезиса может происходить путем преднамеренного нарушения существующей концептуальной системы за счет а) реализации принципа случайности при, б) отсутствии требования обязательности получения результата.
3. В качестве триггера, вызывающего начало акта генезиса и, одновременно с этим, верифицирующего факт того, что акт свершился выступает нонсенс или обнаружение «пустого места» в понимании Делеза.
4. Полученные ИИ-системой выводы и новые знания об окружающей среде должны быть одобрены человеком, и не противоречить знаниям полученным непосредственно от человека.

Относительно генезиса новых символов:

1. ИИ-система имеет право давать имена новым понятиям как элементам концептуальных систем.
2. ИИ-система обязана получать одобрение со стороны человека относительно корректности имен; при наличии разногласий система обязан принять точку зрения человека.
3. Генезис новых имён активируется фактом завершения генезиса новых понятийных элементов концептуальной системы.
4. Акт генезиса новых имен является свободным контингентным актом внутри ограничений, установленных языком, на котором осуществляется именование.
5. Принцип контингентности может быть реализован как псевдослучайный процесс, реализованный на принципах существующих алгоритмов генерации случайных чисел.

Относительно первичного генезиса новых продукционных правил:

1. Процедура первичного генезиса новых правил представляет собой сложный контингентный «a priori» процесс, состоящий из двух шагов: 1) стохастического (псевдослучайного) акта генезиса ансамбля новых правил-кандидатов; 2) процесса верификации полученных правил, с целью отбора правил, непротиворечащих эмпирическим данным.
2. Генерация правил может осуществляться как 1) генерация одиночного правила, 2) встраивания правила (ансамбля правил) в существующую серию 3) генерация новой серии правил.
3. Результатом процесса верификации может быть ансамбль более чем из одного достоверного правила, прошедшего верификацию (следствие парадоксов Делеза).
4. Принцип контингентности может быть реализован как (псевдо)случайный процесс, реализованный на принципах аналогичных существующим алгоритмам генерации случайных чисел, ограниченный свойствами, атрибутами и диапазоном допустимых значений исследуемого класса.
5. Ограничивающими факторами, частично препятствующими возникновению «комбинаторного взрыва», являются механизмы, эксплуатирующие концепцию нонсенса Делеза, поиска избытков и недостатков в исследуемых сериях.
6. Если в вопросах, так или иначе связанных с проблемой генезиса новых символов процесс в значительной степени ограничен рамками феноменологии, эстетики, квалиа (физиологией, или конструкцией сенсоров в случае машинного интеллекта) и т.д., то в задачах генезиса новых правил такого рода ограничений или не существует, или они несущественны.

В параграфе «Онтологический аспект» проводится онтологический анализ регионов бытия символьных методов. Объединяя выводы, полученные относительно эпистемологического и гносеологического аспектов проблемы, на наш взгляд, с необходимостью, можно сформулировать следующие посылки:

1. Если и не полностью, то в значительной степени регион бытия процессов познания пересекается с регионом бытия символьных систем.
2. В свою очередь регион бытия символьных систем пересекается с регионом бытия естественного языка и языков науки.

Далее, получается тот важный для понимания роли символьных методов предварительный вывод, что пересечение этих регионов дает возможность их воспроизводства в системах машинного интеллекта как языковых моделей.

В параграфе «Информационный аспект» проводится анализ процессов коммуникации и обмена информацией с точки зрения проблематики генезиса новых понятий, знаков и правил. Основными полученными выводами являются:

1. Обнаружение факта асимметрии концептуальной структуры адресата, относительно концептуальной системы адресанта должна пониматься как нонсенс, обуславливающий генезис новых понятий и правил.
2. Та информация, что распознается в качестве нонсенса в концептуальной системе адресата, является триггером, запускающим производство (генезис) новых или корректировку существующих концептуальных понятий и правил.

В параграфе «Коммуникативный аспект» проводится анализ роли символьных методов в процессах обмена информацией. Прежде всего речь идет о механизмах кодирования и декодирования информации в рамках предложенной Р.О. Якобсоном и развитой позднее Ю.М. Лотманом⁵³ модели «адресант (коммуникатор) ← некоторый кодирующий механизм ← канал передачи ← декодирующий механизм ← адресат». В результате проведенного анализа были получены следующие выводы:

1. Акт коммуникации невозможен исключительно в рамках коннекционистских подходов, по меньшей мере по той причине, что целью такого акта является изменение собственных представлений о мире,

⁵³ Лотман Ю.М., *Внутри мыслящих миров*, СПб.: Азбука, Азбука-Атикус, 2014, с. 30

сформированных в виде внутренних концептуальных систем или индивидуальной конфигурации мира.

2. Процедура формирования сообщения, связанная с необходимостью кодирования-декодирования информации, представляет собой процесс, специфический для символьных методов.
3. Акт коммуникации, по крайней мере в рамках своей рациональной составляющей, с необходимостью возможен только в рамках семиотических систем, основное содержание которых представляет собой обмен информацией, закодированной в символической форме.

В Заключении подводятся итоги и формулируются основные результаты исследования.

Общие выводы:

1. Основной ролью символьных методов как необходимых компонентов в задачах общего ИИ является обеспечение процессов порождения новых понятий, правил и символов (знаков) как элементов концептуальных систем.
2. Символьные методы являются первичными триггерами, которые запускают тот или иной «мыслительный» процесс, обеспечивают процесс порождения новых смыслов, а также оценку полученных результатов.
3. Гипотезу Ньюэлла-Саймона необходимо уточнить в части понятия «разумного в широком смысле» поведения». А именно: ограничиться пониманием разумного поведения как рационального.
4. Любая система ИИ, приближающаяся к имитации рационального поведения: а) должна быть гибридной (человеко-машинной), б) должна включать символьные методы в качестве основных механизмов порождения новых знаний.
5. Граничные условия, которым должен удовлетворять процесс самообучения, нарушение которых приводит к невозможности

существования ИИ-системы или потере ценности полученных результатов однозначно задаются концептуальными системами человека.

- б. Механизмы символьных методов могут обеспечивать:
 - а. процессы обучения и самообучения символических систем на основании обучающих выборок, аналогично тому, как это происходит при коннекционистских подходах;
 - б. процессы генезиса новых представлений, знаков и правил;
 - в. процессы коммуникации в рамках моделей «человек-машина» и «машина-машина».

III. Список публикаций по теме диссертации

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

I. Публикации в рецензируемых изданиях, индексируемых в международных базах Web of Science, Scopus, RSCI:

1. Ковалев М.А., Является ли искусственный интеллект чем-то большим, чем техника? // Философия и общество. 2022. № 1(102). С. 121-133 (RSCI, импакт-фактор РИНЦ (пятилетний/двухлетний) – 0,450/0,522)).

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

2. Ковалев М.А. Основные методологические подходы для извлечения данных с целью обучения интеллектуальных агентов // Искусственные общества. 2019. Т. 14. Выпуск 2. DOI: 10.18254/S207751800005736 8 (импакт-фактор РИНЦ (пятилетний/двухлетний) – 0,536/0,730).
3. Ковалев М.А. От гибридных интеллектуальных систем к гибридному интеллекту // Искусственные общества. 2020. Т. 15. Выпуск 2. DOI:10.18254/S207751800009722-3 (импакт-фактор РИНЦ (пятилетний/двухлетний) – 0,536/0,730).
4. Ковалев М.А. Концепция когнитивного цикла в задачах общего искусственного интеллекта // Искусственные общества. 2021. Т. 16. Выпуск 2. DOI: 10.18254/S207751800015292-0 (импакт-фактор РИНЦ (пятилетний/двухлетний) – 0,536/0,730).