МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА

На правах рукописи

Шиллер Александра Викторовна

МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ МОДЕЛИРОВАНИЯ ЭМОЦИЙ В АРХИТЕКТУРЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Специальность: 09.00.08 – Философия науки и техники

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата философских наук

Москва – 2019

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

Научный руководитель: **Шестакова Марина Анатольевна**, кандидат философских наук, доцент

Официальные оппоненты:

Никитина Елена Александровна,
доктор философских наук, доцент,
профессор кафедры гуманитарных и общественных наук Института инновационных технологий и государственного управления ФГБОУ ВО «МИРЭА – Российский технологический университет»

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

Петруня Олег Эдуардович,
кандидат философских наук, доцент,
доцент кафедры №517 «Философия» института №5 «Инженерная экономика и гуманитарные науки» ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)»

Защита диссертации состоится «18» декабря 2019 г. в «17» час.00 мин. на заседании диссертационного совета МГУ.09.01 Московского государственного университета имени М.В.Ломоносова по адресу: 119234, Москва, Ломоносовский проспект, д. 27, к.4 (Шуваловский учебно-научный корпус), философский факультет, ауд. А-518 (Зал заседаний Ученого совета философского факультета).

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

С диссертацией можно ознакомиться в отделе диссертаций Научной библиотеки МГУ имени М. В. Ломоносова (Ломоносовский проспект, д. 27) и на сайте ИАС «ИСТИНА»: https://istina.msu.ru/dissertations/244534572/.

Автореферат разослан «___» ноября 2019 г.

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

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. При моделировании и дальнейшем построении систем искусственного интеллекта решающее значение имеет выбор методологических оснований, на которые опираются разработчики и исследователи, и которые выступают базовыми элементами конечных создаваемых продуктов – моделей архитектур искусственного интеллекта или отдельных моделей эмоций. Философия предоставляет теоретический базис в виде универсальных, обладающих общеметодологическим значением категорий и принципов. Именно поэтому философия будет занимать ключевую роль в развитии междисциплинарных исследований искусственного интеллекта, объединенных в рамках дисциплины «Философия искусственного интеллекта», к которой можно отнести и представленную работу.

Интерес к проблеме эмоций возрастает вместе с развитием исследований искусственного интеллекта. Все чаще возникает потребность в человекоподобном поведении, воспроизвести которое очень сложно без попыток моделирования эмоционального аппарата. Особенно важным становится моделирование эмоций в контексте создания агентов, функциональность которых связана с коммуникацией с человеком. Для многих практических задач и проблем (например, распознавание эмоций, реализация эффектов или последствий эмоций) перспективным кажется развитие биоинформатики и технологий машинного обучения. Однако решение отдельных задач (таких как распознание эмоций по фото/тексту и т.д.) пока не привело к качественному прорыву в моделировании эмоциональных систем. Более того, исследователи все чаще отказываются от создания отдельной эмоциональной системы, апеллируя к тому, что эффекты эмоций реализуются в поведении агента автоматически, если они были включены в данные, использованные для обучения агента. Примером практической реализации задачи является разработанный компанией Microsoft чатбот Тэй (Tay) в Твиттере, который быстро научился писать эмоциональные

тексты, но при этом очевидно, что такое поведение не является следствием работы его собственной эмоциональной системы.

За последние пять лет значительно возросло количество вычислительных моделей эмоций и архитектур аффективных агентов. Исследователи в области когнитивных наук, искусственного интеллекта, биоинженерии, программного обеспечения, робототехники давно занимаются созданием «моделей эмоций». Причины, побуждающие ученых и разработчиков исследовать природу эмоций и развивать теоретическое моделирование эмоций, обусловлены запросами практики: необходимостью качественного повышения уровня взаимодействия человека и машины, а также потребностью в создании более правдоподобных и эффективных систем искусственного интеллекта и роботов. [1]

Несмотря на обширную разработку темы моделирования эмоций, существует методологическая неопределённость, вызванная рядом факторов. Во-первых, нет однозначного понимания понятие «модель эмоций».[2] Во-вторых, не существует однозначной общеупотребимой терминологии, что затрудняет сравнение разработанных к настоящему времени теорий моделирования эмоций и определение применимости и эффективности отдельных теорий и моделей в конкретных случаях.

Следствием такой методологической неопределённости является «отсутствие методологических основ или рекомендаций при создании проектов по моделированию конкретных аффективных феноменов»[3], невозможность ответить на фундаментальные для моделирования эмоций вопросы: «Каковы вычислительные задачи, которые должны быть реализованы? Какие теории наиболее применимы для данной модели? Какие данные необходимо предварительно получить из научной литературы и данных эмпирических исследований?».[4]

---

[1] Hudlicka E. What are we modeling when we model emotion? // AAAI Spring Symposium: Emotion, Personality, and Social Behavior. Stanford University, CA: Menlo Park, CA: AAAI Press. P. 52-59. 2008b.

[2] Шиллер А. В. От теорий к моделям эмоций для искусственного интеллекта — основные методологические вопросы // Ценности и смыслы. Т. 4, № 56. С. 126–137. 2018.

[3] Там же

[4] Там же

Проблематика и степень научной разработанности темы. С позиций философии и методологии науки моделирование эмоций в архитектуре искусственного интеллекта изучалось за рубежом, однако в России подобные исследования до настоящего момента не проводились.

Предлагаемые в работе методологические основания моделирования эмоций основаны на психологических и философских концепциях эмоций, в частности, на концепции комплексных или социальных эмоций, а также на современных когнитивных теориях познания – теории воплощенного познания и теории «небрежных когниций». Используемые представления об архитектуре основаны на подходе А. Сломана к созданию архитектуры искусственного агента. В этой работе представлен аналитический подход к разработке методологии и предложен набор общих методологических рекомендаций, существенных для создания аффективных моделей. Путем создания набора необходимых вычислительных задач и альтернатив для их реализации в работе преследуется цель по разработке методологии моделирования, которая выйдет за границы существующей работы в аффективном моделировании и поможет осуществить первые шаги на пути формирования базовых оснований для создания вычислительных моделей эмоций.

В данной работе под методологическими основаниями понимаются теоретические положения, на которых должно основываться моделирование эмоций в архитектуре искусственного интеллекта. Эмоции рассматриваются как состояния, которые отражают ценностные суждения об окружающем мире, себе и других социальных агентах в разрезе целей и убеждений организма, которые мотивируют и направляют приспособительное поведение. Отметим, что термины «цели» и «убеждения» использованы в общем смысле: цели отражают желаемые состояния, а убеждения отражают текущие знания. Кроме того, понятие «цель» в этом определении включает в себя любые представления о желаемых состояниях: сознательные и неосознаваемые, эксплицитные или имплицитные, врожденные или выученные.

Искусственный интеллект понимается как «свойство систем или агентов выполнять творческие, креативные функции, которые традиционно считаются прерогативой человека». Искусственный интеллект и агент используются в работе как равнозначные понятия с учетом определения искусственного интеллекта, которое дают Рассел и Норвиг: «наука об агентах, получающих из своей среды результаты актов восприятия и выполняющих соответствующие действия». [5]

Искусственный интеллект понимается в работе в рамках философской теории о слабом искусственном интеллекте, однако в некоторых разделах сделаны приписки для отдельных рассматриваемых ситуаций, где искусственный интеллект понимается как сильный искусственный интеллект. Архитектура искусственного интеллекта рассматривается как интегрированный набор разнообразных, но взаимосвязанных способностей, реализованных в искусственной интеллектуальной системе в виде набора блоков-преобразователей данных. Для целей работы важно опираться на понятие архитектуры искусственного интеллекта в понимании А. Сломана.

Важность проблемы поиска и выработки общих методологических рекомендаций или оснований моделирования эмоций предопределила актуальность выполнения данной работы.

Таким образом, анализ научно-методической литературы позволил выявить противоречия между назревшей необходимостью построения общей методологии моделирования эмоций в архитектуре искусственного интеллекта и недостаточной научной разработанностью проблемы создания и оценки существующих теорий эмоций, подходящих в качестве оснований для моделирования отдельных процессов обработки эмоциональной информации в рамках разработки вычислительных моделей эмоций.

Объект исследования: эмоции в архитектуре искусственного интеллекта.

[5] Рассел С., Норвиг П. Искусственный интеллект. Современный подход. "AIMA" ("Artificial Intelligence: A Modern Approach"). – М.: Изд-во Вильямс. 1408 с. – 2019.

Предмет исследования: моделирование эмоций в архитектуре искусственного интеллекта.

Цель исследования - выявить и проанализировать методологические основания моделирования эмоций в архитектуре искусственного интеллекта.

Задачи исследования*:*

1. Показать методологическое значение философских, психологических и когнитивных концепций эмоций для исследований по моделированию эмоций в архитектуре искусственного интеллекта.
2. На основе анализа современных научных дискуссий определить роль эмоций в архитектуре искусственного интеллекта.
3. Сформулировать основные методологические проблемы моделирования эмоций в архитектуре искусственного интеллекта.
4. Проанализировать существующие и предложить новые методологические основания вычислительного моделирования эмоций.
5. Предложить теорию для моделирования эффектов эмоций в архитектуре искусственного интеллекта.

Методы исследования:

1. Теоретико-методологический анализ современных подходов к моделированию эмоций в целом, а также к моделированию эмоций в архитектуре искусственного интеллекта в философии, психологии и когнитивных науках для формулирования основных позиций исследования;
2. Метод экспертной оценки существующих вычислительных моделей эмоций и анализ методологии создания моделей с целью выявления недостатков и проблем методологического характера, а также с целью поиска и создания альтернативных решений выявленных проблем.

Положения, выносимые на защиту:

1. Эмоции связаны со всеми когнитивными функциями и оказывают на них влияние, поэтому можно утверждать, что эмоциональная подсистема - необходимость для искусственного агента, способного к мультимодальному отражению действительности.

2. Наиболее перспективным подходом при построении архитектуры искусственного интеллекта является использование отдельных хорошо разработанных теорий, опирающихся на философские и психологические концепции эмоций, в частности, на теорию воплощенного познания и теорию смешанной когнитивно-аффективной системы («небрежных когниций»).
3. Моделирование эмоций в архитектуре искусственного интеллекта осложняется рядом методологических проблем:

   3.1 Отсутствие общей однозначной терминологии затрудняет сравнение теоретических подходов к моделированию эмоций:

   - отсутствует согласованность и ясность в отношении значения понятия «модель эмоций»;
   - наблюдается неопределённость относительно того, какие аффективные состояния моделируется.

   3.2 Модели эмоций заметно отличаются друг от друга в зависимости от моделируемой функции эмоций.
4. Реконструкция эмоциональной системы возможна путем ее моделирования. Основой моделирования эмоций должны служить вычислительные модели эмоций для двух выделяемых процессов: генерации эмоций и выражения эмоций через эффекты эмоций, а также теоретические основания моделирования в виде различных когнитивно-аффективных теорий.
5. В настоящее время нет общепризнанной теории, подходящей на роль основы для моделирования эффектов эмоций. В работе предлагается использовать в качестве теоретической основы для моделирования эффектов эмоций комбинацию из двух теорий: теории воплощенного познания и теории смешанной когнитивно-аффективной системы («небрежных когниций»), названную автором КАМА - когнитивно-аффективной моделью архитектуры.

Научная новизна и теоретическая значимость исследования определяются тем, что в представленной работе ясно определены и описаны методологические основания моделирования эмоций в архитектуре искусственного интеллекта. Выявлены и описаны теоретические основания,

вычислительные задачи и шаги вычислительного моделирования эмоций. Проанализированы современные когнитивные теории и выделены перспективные для моделирования эмоций – теория воплощенного познания и теория «небрежных когниций», которые могут быть использованы в когнитивно-аффективной модели архитектуры (КАМА), разрабатываемой для описания моделирования эффектов эмоций. Описана роль социальных эмоций в эмоциональной архитектуре искусственного интеллекта. Описаны дилеммы и методологические проблемы, характерные для разработки моделей эмоций.

Важно отметить, что у диссертационной работы есть ряд ограничений:

Во-первых, обсуждается только когнитивная модальность эмоций, применимая для генерации эмоций с помощью когнитивных оценок и моделирования влияния эффектов эмоций на когнитивные процессы. Причиной служит более высокий уровень разработанности моделей эмоций, основанных на когнитивных исследованиях, что ни в коей мере не означает, что другие модальности отбрасываются и не считаются столь ж критически важными для понимания феномена эмоций.

Во-вторых, анализ различных математических методов, служащих для обеспечения вычислительных задач, был также неглубоким. Причиной служит тот факт, что исследования моделей эмоций находятся на начальном этапе, а валидация и систематическая оценка существующих моделей эмоций проводилась в ограниченном объеме.

В-третьих, ограниченность объемом работы не позволяет развернуть глубокую дискуссию по поводу существующих моделей, поэтому о них были сделаны довольно короткие заметки.

Несмотря на это, с учетом всё возрастающей включенности искусственного интеллекта в повседневную жизнь людей необходимо уже сейчас разрабатывать эмоциональные системы для искусственных агентов. Для решения этой задачи необходимы такие подходы, которые выступали бы направляющими для выбора дизайна архитектуры искусственного интеллекта, регулировали возможности развития интеллектуальных систем, чтобы

обеспечить надлежащее управление данными и определить степень вовлеченности человека в процесс управления этими системами.

Практическая значимость исследования состоит в формировании базовых методологических рекомендаций и анализе существующих когнитивно-аффективных теорий, которые можно использовать в качестве основы для моделирования эмоций в архитектуре искусственного интеллекта в рамках разработки вычислительных моделей эмоций.

Достоверность и обоснованность исследования обусловлены обширной научной базой, включающей философские, психолого-когнитивистские работы; применением современной методологии научного исследования; использованием комплекса методов, адекватных целям и задачам исследования; апробацией материалов исследования в педагогической практике; разносторонним анализом и обработкой данных.

Личный вклад соискателя заключается в осуществлении обзора по проблеме исследования; в определении основных методологических проблем моделирования эмоций в архитектуре искусственного интеллекта; в разработке авторской теории для моделирования эффектов эмоций в архитектуре искусственного интеллекта; в выдвижении обобщающих выводов, в разработке спецкурса для магистров.

Апробация и внедрение результатов исследования осуществлялись в процессе научной и педагогической практики с помощью разработки учебного курса по теме «Философия эмоционального искусственного интеллекта», выступлений на научных конференциях и в рамках научных публикаций по теме исследования.

Структура работы. Диссертация состоит из введения, трех глав, заключения, библиографического списка.

II. ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** раскрыты актуальность, сформулированы предмет, объект, цель и задачи исследования, показаны научная новизна, положения, выносимые на защиту, а также теоретическая и практическая значимость работы.

Первая глава диссертационной работы **«Основные трактовки эмоций в философии, психологии и когнитивных науках»** посвящена анализу понимания феномена эмоций в философии, современной психологии и когнитивных науках.

Анализ литературы, посвященной изучению эмоций показывает, что в настоящее время изучение эмоций в рамках их моделирования для искусственных интеллектуальных систем переживает подъем и является предметом интереса многих ученых из разных областей науки – от психологии и философии до математических, когнитивных и технических наук. Это обусловлено наблюдаемым в постмодернистской философии поворотом к телесности и чувственности, который ознаменовал отказ от примата рациональности и позволил поднять статус эмоций до равного другим когнитивным процессам или даже превосходящего их. Согласно современным исследованиям включение телесности в «когнитивное уравнение» отражает прагматический подход, ориентированный на практику и вдохновленный биологическими природными стратегиями живых существ, что в исследованиях искусственного интеллекта вылилось в целое направление создания когнитивных архитектур BICA (Biologically Inspired Cognitive Architectures). Основным направлением в создании когнитивных архитектур человека и искусственного интеллекта является отказ от идеи «точных» подходов и переход к созданию подходов, включающих в себя все те процессы и феномены, которые ранее отбрасывались как «помехи» восприятия и мышления, прежде всего, это касается эмоций. Исследования показывают, что именно присутствие эмоций позволяет обеспечивать мультиэвристическую активность и определяет смешанную природу когнитивной сферы.

Однако, наблюдается сложность в создании целостного междисциплинарного подхода к исследованию эмоций. При создании архитектуры искусственного интеллекта в части эмоциональной подсистемы разработчики и исследователи в явном или неявном виде опираются на философские и психологические концепции эмоций, однако на сегодняшний день более перспективным является использование отдельных хорошо разработанных теорий для отдельных участков архитектуры искусственного интеллекта, нежели попытка создания единой теории моделирования.

По результатам проведенного анализа литературы в качестве теорий для построения нового типа архитектур искусственного интеллекта, учитывающих мультиэвристичскую и мультимодальную природу когнитивных процессов и роль эмоций, предлагается использовать теорию воплощенного познания и теорию «небрежных когниций».

Теория воплощённого познания утверждает, что имитация эмоциональных выражений другого индивида является частью телесного проигрывания переживания состояния другого индивида. Когда имитация эмоций происходит гладко, она служит основанием для возникновения сложных социальных эмоций: эмпатии, стыда, гордости, вины и других.

Теория «небрежных когниций» утверждает, что необходимо отказаться от строгих рациональных моделей и создавать естественный антропоморфный интеллект – с учетом всех несовершенств и ошибок (к которым традиционно относили эмоции), поскольку именно несовершенства делают интеллектуальный аппарат «живым» и креативным.

Также в первой главе рассматривается феномен социальных эмоций и роль социальных эмоций в эмоциональной архитектуре искусственного интеллекта. Социальные эмоции относятся к сложным или комплексным эмоциям и, вероятно, именно они являются той выгодной с эволюционной точки зрения особенностью эмоциональной системы человека, которая позволила существовать сознанию как непрерывному процессу отражения действительности на социальном и чувственном уровнях. Социальные эмоции

как комплексные эмоциональные феномены связаны с ситуациями взаимодействия с другими людьми или, в случае встраивания социальных эмоций в архитектуру искусственного интеллекта, также и с другими искусственными системами как социальными агентами, поэтому важно знать их особенности и работать над декомпозицией с дальнейшим включением в архитектуру искусственного интеллекта, что является амбициозной исследовательской задачей для будущих исследований.

Таким образом, анализ научно-методической литературы показал, что существующие теоретические подходы по исследованию эмоций только частично используются при моделировании эмоций в архитектуре искусственного интеллекта. Кроме того, большинство теорий эмоций принадлежат к исследовательской парадигме «старого типа», предполагающей примат рациональности над чувственностью, поэтому их сложно использовать для проектирования архитектуры искусственного интеллекта, которая создаётся в рамках современной парадигмы. В первой главе освещаются три наиболее перспективных теоретических подхода к эмоциям и когнитивным процессам, которые необходимо использовать в качестве теорий-оснований при разработке нового типа архитектур. Однако не было обнаружено исследований, посвящённых проблеме использования этих теорий в вычислительном моделировании эмоций.

Вторая глава «Эмоции в архитектуре искусственного интеллекта» посвящена обзору и анализу современных научных дискуссий о соотношении эмоций и других когнитивных функций и рассмотрению роли эмоций в архитектуре искусственного интеллекта.

При создании методологии построения архитектур искусственных агентов, включающих эмоциональную подсистему, важно четко понимать, как связаны эмоции и другие когнитивные функции. Во второй главе представлен обзор, который освещает роль эмоций для других когнитивных процессов и поддерживает основную мысль работы: «Эмоциональная подсистема -

необходимость для искусственного интеллекта, способного к мультимодальному отражению действительности».

Рассмотрение соотношения эмоций и языка позволяет сделать вывод, что эмоции и язык тесно связаны, но функции этой взаимосвязи очень различны и всё их разнообразие еще предстоит изучить. Кроме того, необходимо учитывать роль языка, используемого агентом/системой/человеком и возможность кодирования эмоций с помощью этого языка, особенно, если предполагается, что искусственному интеллекту необходимо использовать естественный язык при взаимодействии с другими агентами/системами, что предполагает правдоподобность, осмысленность и эмоциональную окрашенность этого взаимодействия.

Можно сказать, что в современной психологической и философской литературе выработаны две основные точки зрения относительно роли эмоций в мыслительной деятельности. Существуют подходы, в которых подчеркивается деструктивный характер эмоциональных процессов для мыслительной деятельности. С другой стороны, существует регуляционный подход, основанный на контролирующей роли когнитивных процессов относительно эмоциональных процессов.

Вне зависимости от выбранного подхода роль эмоциональных процессов, сопровождающих мыслительную деятельность, не раскрывается и не учитывается в достаточной мере. Оба подхода только констатируют наличие феноменов контроля над эмоциями, не пытаясь осветить психические механизмы воздействия эмоций на когнитивную деятельность. Возможного взаимодополнения двух исследовательских традиций также не происходит, поскольку они противоречивы и отрицают тезисы друг друга.

Кроме того, в настоящий момент происходит масштабный пересмотр представлений о процессах восприятия и связи эмоций с этими процессами. Концепции энактивизма и телесно-воплощенной симуляции (теория воплощенного познания) позволяют по-новому взглянуть на роль эмоций в процессе восприятия.

Одной из наиболее интересных и спорных тем является влияние эмоций на принятие решений, связанных с этическими вопросами и моралью в целом, а также само существование таких сложных социальных эмоций как эмпатия. Традиционно, как и взаимодействие эмоции-мышление, связка эмоции-мораль рассматривалась скорее, как негативное взаимодействие, когда эмоции мешали принимать этичные решения и совершать моральные поступки.

Современные данные свидетельствуют о том, что этическая система не сможет существовать без разработанной аффективной или же целостной аффективно-когнитивной системы в том случае, если за образец этической системы будет взята человеческая модель, поскольку эмоции во многом влияют на принятие решений человеком, как уже было указано выше. Поэтому важной задачей является раскрытие и последующая разработка связей между эмоциональной и этической системами в архитектуре искусственных агентов.

Во второй главе подробно анализируется подход А. Сломана к созданию архитектуры агента, в рамках которого рассматриваются блоки архитектуры искусственного агента и общие принципы ее создания.

Таким образом, анализ современных научных дискуссий о соотношении эмоций и других когнитивных функций показал, что эмоции непосредственно связаны и оказывают влияние на все когнитивные функции, поэтому должны занимать одно из центральных мест в архитектуре искусственной интеллектуальной системы. Роль эмоций является настолько значимой, что можно утверждать, что эмоциональная подсистема – необходимость для искусственного интеллекта, способного к мультимодальному отражению действительности.

Эмоции – необходимость и неотъемлемая часть архитектуры искусственного интеллекта, так как они выполняют ряд важных функций, например, интегрируют разные части системы в единое целое за счет обеспечения мультимодальности когнитивных процессов; являются «ситом» в процессе селекции образов и определяют фокус внимания и те объекты/свойства, которые будут восприняты.

Однако для воплощения всех этих требований к архитектуре искусственного интеллекта в части эмоциональной подсистемы существуют методологические ограничения, связанные со сложностью организации эмоциональной системы у человека и, соответственно, с переносом модели эмоций в архитектуру искусственного агента.

В третьей главе «Вычислительные модели эмоций в архитектуре искусственного интеллекта» представлены общие методологические и теоретические проблемы моделирований эмоций, проведен анализ области вычислительного моделирования эмоций и характерных для вычислительного моделирования теоретических оснований, предложена когнитивно-аффективная теория, подходящая для моделирования эффектов эмоций.

Экспертный анализ существующих моделей эмоций и анализ научно-методической литературы показал, что наблюдается целый комплекс методологических проблем:

1. Отсутствуют согласованность и ясность в отношении значения понятия «модель эмоций».
2. Наблюдается неопределённость относительно того, какие аффективные состояния моделируется. Термин «эмоция» в аффективных моделях может относиться к эмоциям как таковым, настроениям, аттитюдам и состояниям, которые не относятся психологами к эмоциям (например, состояние потока).
3. Модели эмоций сильно различаются в зависимости от того, какая из приписываемых функций эмоций моделируется.
4. Отсутствие общей однозначной терминологии затрудняет сравнение теоретических подходов к моделированию эмоций.
5. Отсутствуют методологические рекомендации и общая методология по созданию моделей эмоций и помещению их в архитектуру искусственного интеллекта.
6. Для моделирования эмоций в архитектуре искусственного интеллекта существуют методологические ограничения, связанные со сложностью

организации эмоциональной системы у человека и, соответственно, с переносом модели эмоций в архитектуру искусственного интеллекта.

Все эти методологические проблемы носят общий характер и относятся к моделированию эмоций в целом, поэтому для выполнения целей и задач исследования в третьей главе предпринят анализ состояния области моделирования эмоций у искусственного интеллекта и выявлены теоретические основания моделирования в виде концепций и теорий, а также предложены методологические рекомендации, которые могут лечь в основу моделирования эмоций, для которого сейчас нет общепризнанных теоретических основ.

Реконструкция эмоциональной системы возможна путем ее моделирования, разделенного на условные процессы работы с эмоциональной информацией и с использованием одного из трех наиболее разработанных подходов, которые можно использовать для вычислительного моделирования эмоций. Это категориальный, пространственный и компонентный подходы.

Следует отметить, что эти теоретические подходы не должны рассматриваться как конкурирующие за единственно правильное решение проблемы. Их стоит расценивать скорее как различные взгляды, каждый из которых появился из конкретной исследовательской традиции, концентрируется на разных наборах аффективных феноменов, рассматривает различные уровни решений и разные фундаментальные компоненты, использует разные экспериментальные методы (например, факторный анализ и данные самоотчетов, результаты нейроанатомических исследований). Разные подходы также с разной глубиной описывают отдельные эмоциональные процессы. Например, компонентные теории представляются хорошо проработанными в аспектах когнитивного оценивания.

Пока не наступил момент, когда эмоции будут полностью понятыми и объясненными, лучше всего рассматривать эти три подхода как альтернативные объяснения, каждое со своим набором преимуществ и недостатков, а также с подтверждающими данными, по аналогии с отношением к волновой и корпускулярной теориям света.

В работе предлагается выделять два основных процесса, существенных для вычислительного моделирования эмоций: генерация эмоций и выражение эмоций через эффекты эмоций.

В процесс генерации эмоций включены все модальности, но он описан наиболее полно со стороны когнитивной модальности, и наиболее «рабочие» модели генерации эмоций содержат в себе когнитивную оценку. В дополнение стоит отметить, что поскольку обладающие телесным воплощением (материальные) агенты/ искусственные системы становятся сложнее, будет возрастать необходимость включать в процесс генерации эмоций и не когнитивные модальности, что можно сделать с помощью использования теории воплощенного познания.

Теории о выражении эмоций не так хорошо разработаны, как теории генерации эмоций. Основными подходами к генерации эмоций являются аффективные модели и когнитивная оценка, однако для создания моделей эмоциональных выражений (эффектов эмоций) таких теорий просто не существует.

Комплексная теория эмоций должна, по сути, объяснять сам феномен эмоций. В работе предлагается идея использования двух современных теорий - теории «небрежных когниций» и теории воплощенного познания, что поможет лучше понять, как можно представить когнитивные эффекты эмоций без потери таких свойств эмоций как мультимодальность и комплексность и позволит включать в архитектуру социальные эмоции. Поскольку исследования эмоций активно развивались в последние десятилетия, использование этих двух теорий в дополнение к трем классам уже используемых и разработанных для вычислительного моделирования теорий хорошо сочетается с представлением о том, что теория не должна представлять собой «единую теорию эмоций». Кроме того, учитывая разнообразие процессов-посредников эмоций, создание единой теории кажется маловероятным.

В работе предложена когнитивно-аффективная модель архитектуры (КАМА), подходящая для моделирования выражения эффектов эмоций. Модель

КАМА, основанная на теории воплощенного познания, является наиболее перспективной в рамках моделирования эмоций у искусственного агента. КАМА - когнитивно-аффективная модель архитектуры, которая фокусируется на моделировании широкого набора индивидуальных различий в процессе когнитивной деятельности, включая аффективные состояния и черты. Особенность этой модели – акцент на моделировании эффектов эмоций (радость, страх, злость и грусть) в когнитивных процессах, являющихся медиаторами принятия решения (внимание, оценка ситуации, создание ожиданий, управление целью и выбор действия).

Можно заключить, что основой моделирования эмоций должны служить вычислительные модели эмоций для процессов генерации эмоций и выражения эмоций, а также теоретические основания моделирования в виде различных когнитивно-аффективных теорий. Кроме того, можно сделать вывод, что эмоциональный искусственный интеллект возможен либо при наличии телесного воплощения у искусственного агента, либо с обязательным условием учета и включения в модель всех модальностей – имитация одних лишь когнитивных процессов не даст информационной полноты, требуемой для правдоподобного и антропоморфного эмоционального ответа.

В процессе анализа было выделено еще несколько ограничений теоретического характера, которые важно учитывать при моделировании эмоций:

1. Существующие теории не способны полностью удовлетворить требования, предъявляемые к теориям-основаниям для моделирования эмоций. Возможным выходом является более тщательная разработка теорий, направленных на объяснение механизмов влияний эмоций на когниции – теории небрежных когниций и теории воплощенного познания в разрезе вычислительного моделирования эмоций. Другой способ – создавать действующие модели эмоций для искусственного интеллекта, в процессе работы над которыми возникнут новые методологические

сложности, которые, в свою очередь, подтолкнут развитие теоретического знания и станут триггерами для создания новых теорий эмоций.

2. При выборе теории для моделирования эмоций необходимо убедиться, что данная теория обладает подходящим уровнем проработанности для обеспечения возможности использования различных вычислительных зада.

Далее в третьей главе подробно рассмотрены вычислительные задачи и представлены рекомендации для дальнейшего развития вычислительного моделирования эмоций.

Определение базовых процессов, требуемые для моделирования эмоций – только первый шаг. Для того, чтобы они стали применимы на практике, необходимо деконструировать высокоуровневые процессы и определить отдельные вычислительные задачи, необходимые для воплощения моделей генерации эмоций и эффектов эмоций. В настоящее время для генерации эмоций Брекенсом и коллегами разработана вычислительная модель, в которой представлен и проведен сравнительный анализ оценочных теорий, подходящих в качестве оснований для вычислительного моделирования генерации эмоций. В работе подход Брекенса был дополнен с учетом требований и необходимых оснований для вычислительного моделирования эффектов эмоций.

Было предложено использовать несколько вычислительных задач, необходимых для генерации эмоций и когнитивной оценки, с акцентом на когнитивной модальности, что выражается в символичности вычислительных моделей эмоций и используется в настоящее время при построении архитектур искусственного интеллекта.

Были выделены следующие вычислительные задачи, для которых предложены разнообразные математические способы вычисления интересующих значений:

1. Определить и воссоздать соотношение триггера и эмоции (в зависимости от принимаемой теоретической основы оно может включать дополнительные подзадачи, которые должны определить место триггера эмоции в промежуточной репрезентации (в пространствах или векторах

оценочных переменных), а затем последующее место в финальной эмоции(ях));

2. Посчитать интенсивность результирующей эмоции;
3. Посчитать время распада эмоции;
4. Объединить сложные эмоции, если они генерировались;
5. Объединить только что сгенерированные эмоции с уже существующими эмоциями или настроениями.

В работе также предложены несколько других вычислительных задач для воссоздания аффективной динамики и моделирования эффектов эмоций:

1. Определить и воплотить соотношение эмоций от настроений к эффектам для модальностей, включенных в модель (когнитивная, экспрессивная, поведенческая, нейрофизиологическая). В зависимости от лежащей в основе теории, к основным задачам могут добавляться дополнительные подзадачи, которые определяют осуществление промежуточных шагов и определены с помощью более абстрактных семантических примитивов, предлагаемых теорией (например, пространства, оценочные переменные);
2. Определить магнитуду результирующих эффектов как функцию эмоции или интенсивности настроения;
3. Определить изменения в этих эффектах, если эмоция или интенсивность настроения угасает со временем;
4. Соединить эффекты сложных эмоций, настроений или нескольких эмоций и настроений, если сложные эмоции и настроения были созданы на подходящем этапе процесса;
5. Соединить эффекты только что появившихся эмоций (созданных) с остаточными уже сопровождающими эффектами для обеспечения правдоподобного перехода между состояниями с течением времени;
6. Подсчитать вариабельность приведенных выше интенсивностей аффективных состояний и особенностей «личности» моделируемого искусственного интеллекта/агента;

7. Координировать видимые проявления эффектов эмоций по разным каналам и модальностям внутри одного временного отрезка для обеспечения правдоподобности выражения эмоций.

Набор задач для каждой из моделей зависит от выбранного теоретического основания модели. Надо также отметить, что не для всех моделей обязательно необходимы все задачи, например, в более простых моделях, где генерируется только одна эмоция, нет необходимости в создании сложных эмоций.

Важной частью вычислительного моделирования эмоций является анализ требований и выбор конкретных шагов моделирования. Для определения степени проработанности эмоций, которые планируется моделировать в данном искусственном интеллекте и для выбора наиболее подходящей парадигмы аффективного моделирования, создатель модели прежде всего должен провести анализ требований: необходимо иметь ясное представление о том, как эксплицитное моделирование эмоций в архитектуре искусственного интеллекта будет усиливать его функциональность или эффективность. В работе сделан акцент на использовании отдельных областей моделирования и вычислительных задач, сопровождающих генерацию и эффекты эмоций. В работе также подробно описаны 10 шагов, необходимых для анализа аффективных требований в процессе разработки модели.

Кроме того, для вычислительного моделирования были выделены возникающие дилеммы и ключевые вопросы при разработке моделей эмоций.

Так или иначе, простые решения в случае требований усиления аффективной сложности и реализма не будут адекватно работать, и для случаев моделирования независимо существующих разнообразных эмоциональных процессов, сопровождающих эффекты эмоций по разным каналам, разработчикам модели эмоций потребуется принимать более сложные решения. Кроме того, некоторые аффективные состояния могут быть определены только с помощью отдельных паттернов их аффективной динамики (например, смущение). Создание правдоподобных выражений этих типов сложных эмоций у искусственного интеллекта является активной сферой исследований.

В **Заключении** подведены итоги исследования. Анализ научно-методической литературы показал, что в области моделирования эмоций наблюдается ряд методологических сложностей, а именно отсутствие методологических оснований для разработки моделей эмоций, недостаток однородной терминологии, неопределённость моделируемых аффективных феноменов, явные различия между существующими моделями эмоций в зависимости от моделируемой функции эмоций.

В связи с этим, в работе представлен широкий обзор подходов к анализу эмоций, а также используется вычислительный аналитический подход для рассмотрения и решения выявленной проблемы.

В работе освещаются существующие методологические сложности и предлагаются методологические рекомендации для разработки подходов к моделированию эмоций в более структурном русле. Представленный здесь анализ имеет некоторые ограничения, так как исследования моделей эмоций находятся на начальном этапе, а валидация и систематическая оценка существующих моделей эмоций проводилась в ограниченном объеме.

Представленный в работе анализ стимулирует исследования и разработку вычислительных моделей эмоций с помощью использования и улучшения предложенного аналитического вычислительного подхода путем уточнения и развития как основных вычислительных задач, так и доступных альтернативных теоретических подходов.

Акцент в этой работе сделан на разработке моделей эмоций для архитектуры искусственного интеллекта, что позволяет присвоить ей более прагматический, прикладной статус по сравнению с изучением и созданием моделей эмоций, целью которых является выявление механизмов, лежащих в основе эмоциональных процессов. Такие совершенно теоретические модели не могут игнорировать мультимодальную природу генерации эмоций, как происходит во многих существующих моделях эмоций в архитектуре искусственного интеллекта, и должны эксплицитно представлять сложные

взаимодействия между разными модальностями, как для моделирования генерации, так и для моделирования эффектов эмоций.

В данной работе не приведены примеры полностью рабочих архитектур, чтобы показать, какие разные типы онтологий для ментальных состояний и процессов могут поддерживаться и насколько хорошо они будут объяснять сложнейшие аспекты человеческой ментальности. Создание собственной модели рабочей архитектуры агента, включающей эмоциональную подсистему – это задача будущих исследований.

Моделирование эмоций в архитектуре искусственного интеллекта в настоящее время - та область, где аффективное моделирование может оказать реальное влияние на усиление производительности и эффективности существующих искусственных интеллектуальных систем/агентов. При этом не менее важным философско-методологическим следствием работы в области моделирования эмоций является потенциальная возможность для исследователей продвинуться на пути понимания природы аффективных феноменов. Вычислительное моделирование эмоций требует детальной операционализации существующих высокоуровневых понятий, включая термин «эмоция», и поэтому потенциально может способствовать освещению механизмов, лежащих в основе аффективных процессов. Разработка моделей эмоций таким образом послужит исследовательским материалом, который вдохновит философскую мысль и теоретическую работу по изучению эмоций. Возможно, это наиболее сложная задача из всех, но при этом и самая перспективная для развития вычислительного моделирования эмоций в архитектуре искусственного интеллекта и философии искусственного интеллекта.

III. СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных изданиях, отвечающих требованиям п. 2.3 Положения о присуждении ученых степеней в МГУ имени М.В. Ломоносова (по философским наукам):

1. Шиллер А.В. Роль теорий воплощенного познания в исследованиях и моделировании эмоций. // Философские науки – 2019. – Т.62, № 5. – С.124-138 (двухлетний Импакт-фактор РИНЦ 2018 – 0,444).
2. Шиллер А.В. Роль эмоций в когнитивно-аффективной системе как фактор развития архитектуры искусственных агентов // Вестник Московского университета. Серия 7. Философия. — 2018. — № 6. — С. 78–90 (двухлетний Импакт-фактор РИНЦ 2018 – 0,081).
3. Шиллер А.В. От теорий к моделям эмоций для искусственного интеллекта — основные методологические вопросы // Ценности и смыслы. — 2018. — Т. 4, № 56. — С. 126–137 (двухлетний Импакт-фактор РИНЦ 2018 – 0,302).
4. Шиллер А.В. Социальные эмоции и проблема интерсубъективности // Философия и общество. — 2018. — Т. 1, № 86. — С. 121–123 (двухлетний Импакт-фактор РИНЦ 2018 – 0,418).

Статьи, опубликованные в иных рецензируемых научных журналах, рекомендованных ВАК при Министерстве образования и науки Российской Федерации (по философским наукам):

5. Шиллер А. В. Теоретические основания моделирования социальных эмоций в мультиагентных средах // Ежеквартальный Интернет – журнал Искусственные общества. — 2018. — Т. 13, № 1-2.
6. Зайцев И. Д., Шиллер А. В. Теории понимания другого в проектах искусственного интеллекта // Искусственный интеллект: философия, методология, инновации. Сборник трудов X Всероссийской конференции студентов, аспирантов и молодых учёных. Москва, МИРЭА, 27–28 апреля 2017 г. Под общей редакцией Е.А. Никитиной. — Т. 1. — М.: Московский технологический университет (МИРЭА) ISBN 978-5-94101-324-1, 2017. — С. 31–36.
7. Шиллер А. В. Роль вычислительных моделей эмоций в проектах искусственного интеллекта // XIV Международный Междисциплинарный Конгресс "Нейронаука для медицины и психологии", Судак, Крым, Россия, 30 мая-10 июня 2018 года. — Москва, 2018. — С. 537–538.