DOI: 10.30727/0235-1188-2021-64-1-71-87
Оригинальная исследовательская статья
Original research paper

Нейрофилософия, философия нейронаук и философия искусственного интеллекта: проблема различения

Е.А. Безлепкин
Институт философии и права Сибирского отделения РАН, Новосибирск, Россия

А.С. Зайкова
Институт философии и права Сибирского отделения РАН, Новосибирск, Россия

Аннотация

Под нейрофилософией понимают разные направления философии, в частности философию нейронаук, философию искусственного интеллекта или элиминативный материализм. Чрезмерная нагруженность термина связана с еще не завершившимся процессом осмысления предметной области этой дисциплины. К примеру, в одном из первых определений нейрофилософии П.С. Черчленд говорилось о редукции психологии к нейронаукам. С современных позиций представление о нейрофилософии как о попытке оправдать элиминативный материализм устарело и не соответствует действительности. В статье проведится анализ терминов «философия нейронаук», «нейрофилософия» и «философия искусственного интеллекта», а также предлагается вариант их разделения. Общность и различия показаны на примере теории сознания Дж. Эдельмана и концепции коннекционизма для слабого искусственного интеллекта. Делается вывод, что от интегрального использования понятия «нейрофилософия» целесообразно отказаться. Под термином «нейрофилософия» следует понимать направление в философии начала XXI века, применяющее нейронаучные концепции для решения традиционных философских проблем, а философия нейронаук может быть рассмотрена в первую очередь как раздел философии науки, который формулирует и решает проблемы и частных нейронаук, и нейронаучного направления в целом. «Философия искусственного интеллекта» – направление в философии, отвечающее на вопрос о том, что такое небиологический интеллект и возможен ли он. Иными словами, это – философско-методологическая база для изучения небиологического интеллекта. Авторы приходят к выводу, что в становлении нейронаук и их научно-философского базиса мы пока находимся на первом методологическом этапе анализа и диффе-

ренциации гипотез. Философия нейронаук как база существующих нейронаучных теорий возникнет через определенное время, и именно в соответствующий период будет необходим данный термин.

Ключевые слова: искусственный интеллект, нейронаучная философия, когнитивные науки, нейробиология, П.С. Черчленд, Дж. Эдельман, коннекционизм, нейросети.

Безлепкин Евгений Алексеевич – кандидат философских наук, научный сотрудник Института философии и права Сибирского отделения РАН,
evgeny-bezlepkin@mail.ru
https://orcid.org/0000-0001-9020-5445

Зайкова Алина Сергеевна – младший научный сотрудник Института философии и права Сибирского отделения РАН.
zaykova.a.s@gmail.com
https://orcid.org/0000-0003-3300-0130

Для цитирования: *Безлепкин Е.А., Зайкова А.С.* Нейрофилософия, философия нейронаук и философия искусственного интеллекта: проблема различения // Философские науки. 2021 Т. 64 № 1. С. 71–87.
DOI: 10.30727/0235-1188-2021-64-1-71-87

Neurophilosophy, Philosophy of Neuroscience, and Philosophy of Artificial Intelligence: The Problem of Distinguishing

E.A. Bezlepkin
Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science, Novosibirsk, Russia

A.S. Zaykova
Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science, Novosibirsk, Russia

Abstract

Neurophilosophy is understood as different areas of philosophy, for example, the philosophy of neuroscience, the philosophy of artificial intelligence, or eliminative materialism. This excessive interpretation of the term is due to the fact that the understanding of the subject area of this discipline is still incomplete. For example, one of the earliest definitions of neurophilosophy given by P.S. Churchland stated reduction of psychology to neurosciences. In modern views, the idea of neurophilosophy as an attempt to

justify eliminative materialism is outdated and does not correspond to reality. The article analyzes the terms "philosophy of neuroscience," "neurophilosophy," and "philosophy of artificial intelligence" and also offers a variant of their differentiation. The authors focus on the common and different features, using the example of G.M. Edelman's theory of consciousness and the concept of connectionism for weak artificial intelligence. It is concluded that integral use of the term "neurophilosophy" should be abandoned. As a result, the term "neurophilosophy" should be understood as a direction in philosophy of the early 21st century, applying neuroscientific concepts to solve traditional philosophical problems, while the philosophy of specific neurosciences can be considered primarily as a field in the philosophy of science that formulates and solves problems of specific neurosciences as well as of the entire neuroscientific direction. The philosophy of artificial intelligence is an area in philosophy that answers the question of what non-biological intelligence is and what makes it possible; in other words, it is a philosophical and methodological basis for the study of non-biological intelligence. In the formation of neurosciences and their scientific and philosophical basis, we are still at the first methodological stage of the analysis and differentiation of hypotheses. After some time, there will emerge a philosophy of neuroscience, as the basis of all existing neuroscientific theories, and then this term will acquire greater significance.

Keywords: artificial intelligence, neuroscientific philosophy, cognitive science, P.S. Churchland, G.M. Edelman, connectionism, neural networks.

Evgeny A. Bezlepkin – Ph.D. in Philosophy, Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science.
evgeny-bezlepkin@mail.ru
https://orcid.org/0000-0001-9020-5445

Alina S. Zaykova – Junior Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science.
zaykova.a.s@gmail.com
https://orcid.org/0000-0003-3300-0130

For citation: Bezlepkin E.A. & Zaykova A.S. (2021) Neurophilosophy, Philosophy of Neuroscience, and Philosophy of Artificial Intelligence: The Problem of Distinguishing. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 71–87.
DOI: 10.30727/0235-1188-2021-64-1-71-87

Введение

Философы сегодня по-прежнему спорят о том, что такое философия. В приведенном утверждении присутствует, конечно, доля

лукавства, поскольку определяющие критерии философии известны: предметная область достаточно четко определена, проблемы (как классические, так и современные) поставлены, стратегии их решения, по крайней мере, для философии частных наук, более или менее известны и алгоритмы их применения прописаны. Не хватает только четкого и емкого определения. Не хватает потому, что философия – необычайно разнородный внутри себя предмет, охватывающий области от истории философии до логики и философий конкретных наук.

Следовательно, любые попытки определения термина «философия» приводили и будут приводить к возникновению множества непреодолимых дискуссий. Подобное произошло и с термином «нейрофилософия». Вопрос об определении того, чем является нейрофилософия, можно соотнести с вопросом о том, что такое философия. Отсутствие определенности в постановке предметной области, проблем и методов их решения делает вопрос о том, что такое нейрофилософия, философской проблемой. В настоящей статье предлагается разграничение трех терминов: первый – нейрофилософия (нейронаучная философия), второй – философия нейронаук (т.е. философия частных наук, к которым применима приставка «нейро-») и третий – философия искусственного интеллекта.

Начнем с определения нейронауки. В самом общем смысле под нейронауками понимают междисциплинарную область знаний, которая изучает нейронные процессы. По сути, нейронауки начали отделяться от нейробиологии подобно тому, как в свое время физика и другие науки о природе начали выделяться из философии. Интересный формальный критерий для определения нейронаук предложил А.Ю. Алексеев: «Науке приписывается префикс “нейро-”, если в ее инструментарии преобладают коннекционистские методы исследования» [Алексеев, Кузнецов, Савельев, Янковская 2015, 55]. В данном случае термин «философия нейронаук» должен включать в себя всю философско-методологическую базу указанных наук. Итак, под это определение подходит и часть философии искусственного интеллекта. Действительно, направление, изучающее возможность создания искусственного интеллекта, включает в себя коннекционистские и символьные методы исследования. Если искусственный интеллект рассматривать как нейронауку, то философия искусственного интеллекта не может не являться частью философии нейронаук.

Ревонсуо в книге «Психология сознания» [Ревонсуо 2013] корректно и точно определил, как следует отвечать на вопрос о том, что такое сознание в методологическом смысле: «Описать сознание означает дать его определение, ввести понятия, которые ясно и системно описывают основные особенности сознания, привести показательные примеры этого феномена и дифференцировать его от других феноменов, с которыми его можно легко перепутать. Объяснить сознание означает связать его с другими феноменами, описав механизмы и принципы, лежащие в его основе или несущие ответственность за его возникновение, и показав, как сознание взаимодействует с мозгом и как оно руководит нашим поведением» [Ревонсуо 2013, 256]. Предлагаем найти аналог определения для понятия «нейрофилософия».

Эволюция термина «нейрофилософия»

Исторический анализ – один из основных методов исследования в философии. Полагаем, нет причин отказываться от него и в рамках нашего исследования. Главной отправной точкой нейрофилософии как отдельной области исследований стала книга П.С. Черчленд с одноименным названием «Нейрофилософия» [Churchland 1986]. Рассматриваемое в книге понимание нейрофилософии в значительной степени отличалось от его сегодняшней трактовки.

Нейрофилософия, описанная Черчленд, подана как объединение философии сознания и различных когнитивных наук, прежде всего нейронаук. С философской позиции Черчленд защищала идеи философии элиминативного материализма. С одной стороны, эта концепция предполагает, что сознание либо некоторые классы ментальных состояний не существуют. Они являются иллюзией, возникающей из-за наличия языка или долговременной памяти. С другой стороны, согласно данной концепции, не может существовать нейробиологического базиса для многих наших ментальных состояний (например, веры или желания). Подобных идей придерживались, например, бихевиористы.

Помимо идей элиминативизма Черчленд высказывала предположение о коэволюции психологии, нейробиологии и философии сознания до тех пор «пока в будущем, на некотором более высоком уровне, психологические теории не окажутся редуцированными к более фундаментальной нейрофизиологической теории; именно тогда возникнут предпосылки для разработки единой теории

сознания и мозга» [Философия… 2006, 365]. Под влиянием идей Черчленд сформулировано следующее определение нейрофилософии: это – «направление в современной философии науки, пытающееся обосновать правомерность редукции психологии к нейронаукам (нейробиологии, нейрофизиологии и нейропсихологии)» [Философия… 2006, 365]. Согласно трактовке, предложенной в энциклопедическом словаре под редакцией А.А. Ивина, одной из главных задач нейрофилософии является исследование возможностей компьютерного мышления, а также компьютерного моделирования природы мозга и сознания.

За прошедшие 20 лет количество оформившихся нейронаук необычайно расширилось. Например, появились социальная нейронаука, нейроархитектура, нейроэтика, нейроэкономика и т.д. Философско-методологическим базисом каждой из них должны быть конкретно-научные и философские теории. Общий базис должен быть назван философией нейронаук. Одновременно возникает вопрос о том, что означает нейрофилософия (нейронаучная философия).

Различие между нейрофилософией и смежными дисциплинами

В англоязычной литературе термины «философия нейронауки» и «нейрофилософия» сложно назвать синонимами. И. Голд и А.Л. Роскис пишут о том, что ряд исследователей различают философию нейронаук и нейрофилософию, перечисляя возможные направления философии нейронаук, не затрагивающие нейрофилософию: «анализ теоретических концепций; исследование методологий науки; отношение неврологии к другим наукам» [Gold, Roskies 2008, 350]. Они обращают внимание на то, что философия нейронауки мало распространена и что фактически лишь некоторые исследователи освещают эти вопросы. Поэтому указанные авторы предлагают в своей работе понимать «философию нейронаук» более широко, т.е. «как любое философское исследование, в котором нейронауки играют важную роль». Несмотря на четкое разделение терминов «философия нейронауки» и «нейрофилософия», иногда им сознательно пренебрегают.

М. Юнгерт придерживается более строгого различения. Он пишет: «В то время как философия нейронаук пытается применить методы и классические подходы философии науки к нейронаукам, т.н. нейрофилософия использует другой подход, применяя

результаты, полученные с помощью нейронаук, к классическим философским вопросам» [Jungert 2017]. В частности, он настаивает на том, что обсуждение объяснительных стратегий нейронаук относится к философии нейронаук, а разработка эмпирически обоснованных теорий относительно концепции морали или природы сознания – задача нейрофилософии. Подобного мнения придерживаются Дж. Бикл, П. Мандик и А. Ландрет. Они убеждены в том, что философия нейронауки занимается основополагающими проблемами нейронаук, а нейрофилософия пытается применить нейронаучные концепции к традиционным философским вопросам [Bickle, Mandik, Landreth 2019].

В итоге становится понятным, что если философию нейронаук рассматривают в первую очередь как философию науки, которая решает проблемы нейронаук, то под нейрофилософией понимают попытки применений различных концепций (к примеру, нейробиологических) к традиционным философским вопросам. Однако в русскоязычной литературе такое разделение встречается редко, поскольку возможность применения какой-либо науки в философии, как правило, считается традиционной проблемой философии этой науки. И наоборот: решение проблем нейронаук нередко основано на попытках совместить концепции нейронаук и философскую тематику.

Аналогичная ситуация происходит и при попытке разделить нейрофилософию и философию искусственного интеллекта, т.к. последняя часто не ограничивается методологической базой для попыток построения искусственного интеллекта и ставит более общие вопросы о том, «что такое интеллект», «что делает человека человеком», «где граница человеческого познания – там ли она, где границы познания машинного». Возможно поэтому в России обсуждение проблем и методов нейрофилософии в большей степени тесно связано, во-первых, с изучением работы мозга, во-вторых, с искусственным интеллектом.

Попытаемся выделить понятие нейрофилософии из ряда близких философских направлений. Согласно позиции ряда отечественных исследователей, под нейрофилософией следует понимать «конвенциональное обозначение стратегического направления философской науки, концентрированно характеризующего современные натуралистические интерпретации одного из “основных вопросов философии”» [Алексеев, Кузнецов, Савельев, Янковская 2015, 50]. Подчеркивается, что, несмотря на появив-

шуюся новую терминологию (среди новых терминов указаны субъективная реальность / нейрональная активность, когниция/ реализация), суть проблемы остается прежней. Речь снова идет о взаимоотношении психики и мозга, идеи и материи.

Новизна нейрофилософии заключается в первую очередь в натуралистическом тренде, который обусловлен развитием когнитивных наук и исследованиями в области искусственного интеллекта. А.Ю. Алексеев и его коллеги признают значение нейрофилософии как философии нейронаук, т.е. как философско-методологической рефлексии над основами нейронаук. Однако они дополняют представление о нейрофилософии определением, относящимся к иному направлению нейрофилософии: «Нейрофилософия – это систематическая форма изучения мировоззренческих аспектов, опирающаяся как на категориальные знания о нейрофизиологических основах психических явлений, так и на компьютерные методы имитации, моделирования, репродуцирования мозговой, психической и социальной активности» [Алексеев, Кузнецов, Савельев, Янковская 2015, 51].

Резюмируя, можно утверждать, что в отечественной философии выделено множество аспектов понимания «нейрофилософии»: философия науки, философия с акцентом на использование натурализма нейронаук (не только философия сознания, но и философия искусственного интеллекта, и социальная философия), элиминативный материализм, вариант философии сознания с опорой на нейронаучные теории и данные, а также междисциплинарное направление, изучающее системы мозга и системы, подобные мозгу.

В зависимости от того, какое понимание нейрофилософии нами использовано, на первый план выступает та или иная научная проблема. Перечислим ключевые проблемы, которые ученые пытаются решить в рамках нейрофилософии. Во-первых, проблема, связанная с определением того, что такое нейрофилософия и существуют ли у нее специфические характеристики. Во-вторых, психофизиологическая проблема сознания-тела. В-третьих, проблема построения искусственных нейросетей. В-четвертых, проблема методологии, которая включает в себя вопросы применения философских и метанаучных категорий в области нейронаук и, наоборот, применения результатов нейронаучных исследований в области философских и социальных исследований. И, наконец, актуален междисциплинарный

вопрос интеграции различных когнитивных и нейронаучных исследований.

Наличие одной проблемы для разных дисциплин приводит к тому, что провести разделение между такими дисциплинами непросто. Так, граница между нейрофилософией и философией сознания становится неясной. Причина очевидна: и нейронауки, и философия сознания рассматривают как ключевую проблему «сознание – мозг», но подход для ее решения у них различен. По мнению М. Юнгерта, разница между нейрофилософией и философией сознания существует только для философов, занимающихся исключительно философией сознания. Нейрофилософы не проводят строгой границы между эмпирическими, теоретическими и метатеоретическими вопросами, полагая, что нейрофилософия и философия сознания представляют собой единое направление исследований [Jungert 2017].

На наш взгляд, изложенный подход несколько радикализирован. Если нейронауки исследуют мозг, то нейрофилософия пытается искать общие подходы к его исследованию, а философия сознания исследует проблему «сознание – мозг» со стороны сознания. Чтобы продемонстрировать разделение между перечисленными дисциплинами более четко, рассмотрим две теории – теорию Эдельмана как пример нейронаучного подхода и концепцию коннекционизма в качестве синтеза идей философии нейронаук и философии искусственного интеллекта.

Теория Эдельмана как пример нейронаучного подхода

Ярким примером нейронаучного подхода может стать теория сознания Дж. Эдельмана, которую анализирует Д.И. Дубровский в статье «Нейрофилософия и проблема сознания» [Дубровский 2015], делая акцент на том, что Эдельман придает огромное значение философскому осмыслению своей теории и исследуемой проблемы. Охарактеризуем теорию Эдельмана подробнее и проследим связь его теории с философией.

Эдельман разрабатывает теорию отбора (селекции) нейрональных групп. Теория отбора нейрональных групп предложена им в рамках концепции нейродарвинизма. Ее сторонники настаивают на принципиальной изменчивости и приспосабливаемости нейронных сетей и групп. Эдельман утверждает, что при поступлении некоторого сигнала через органы последовательно активизируются разные этапы обработки этих сигналов. Каждый

из них производит своя группа нейронов. В дальнейшем группы, производящие обработку сигналов, объединяются и образуют систему нейрональных групп, что позволяет оптимизировать восприятие, т.е. ускорить или уточнить восприятие часто встречающихся сигналов.

Эдельман определяет первичное сознание как «состояние наличия ментальной осведомленности о вещах в мире» [Эдельман 2012, 422]. Он полагает, что первичное сознание является т.н. помнимым настоящим – *remembered present,* которое, в свою очередь, служит аналогом кажущегося настоящего – *specious present*, включающего в себя настоящее и краткий интервал недавнего прошлого. Высокоуровневое сознание, в отличие от первичного сознания, воплощает не только настоящее и недавнее прошлое, но и прошлое, и будущее. Оно относится прежде всего к прямому осознанию и сопровождается квалиами как формами высокоуровневой организации. Для построения модели первичного сознания, как правило, он применяет данные нейробиологии и физиологии. Однако представлено три пункта, где его теория тесно связана с философией.

Во-первых, теория Эдельмана построена на трех допущениях (физическом, эволюционном и квалиа). Физическое допущение предполагает, что законы физики не нарушаются, а духов и призраков, выходящих за рамки физического мира, не существует. Тем не менее физическое описание мира, согласно Эдельману, не является достаточным и полным. Эволюционное допущение сводится к убеждению о том, что сознание возникло в рамках эволюционного процесса, что оно реально действует и не является эпифеноменом. Квалиа допущение предусматривает, что квалии присущи всем человеческим существам и что не существует научных наблюдателей, которые были бы лишены квалий. Итак, все допущения предложены и сформулированы на базе ряда современных идей философии сознания.

Во-вторых, Эдельман разрабатывает исследовательскую программу и считает ее наиболее успешной. Он предлагает построение модели первичного сознания, надстроение над ней модели высокоуровневого сознания, а затем анализ связи построенных моделей с феноменальным опытом. Экспериментальное исследование, по его убеждению, должно происходить в обратном порядке: от феноменальных данных к нейрофизиологическому объяснению. Глубокий методологический подход ставит нас в

контекст методологии и философии науки, который Эдельман применяет исключительно к собственной теории.

Кроме того, он пытается решить такие традиционные философские вопросы, как существование квалий и проблему интерсубъективности. Все это позволяет утверждать, что его нейронаучная работа, хотя и не является примером теории философии сознания, но сопровождается философской и методологической работой.

Однако стоит обратить внимание и на следствие этой теории. Известно, что теория нейродарвизима Эдельмана стала основой теории селекции, применяющейся сегодня для ряда разработок из области искусственного интеллекта. В целом его теорию можно причислить к идеям, вдохновившим сторонников направления биовычислений (*Bio-inspired computing*) как метода создания искусственного интеллекта, который использует эволюционный подход, в отличие от традиционного, для искусственного интеллекта креационизма [Edelman 2007].

Целесообразно, как нам кажется, применить к этой концепции описанное выше разделение философии нейронауки и нейрофилософии. Философский анализ концепции Эдельмана, скорее, относится к философии науки, поскольку и его концепция как таковая является научной. Тем не менее поиск и анализ допущений и следствий из теории фактически относится к философии сознания, а проектирование исследовательской программы, как теоретической, так и экспериментальной, можно отнести к философии и методологии науки. Попытка же проанализировать вопрос о том, решает ли теория какие-то из традиционных проблем философии сознания, относится к сугубо философии сознания, несмотря на то, что используются богатые эмпирические данные физиологии и нейронаук. И, наконец, применение теории Эдельмана к проблемам искусственного интеллекта относится к философии искусственного интеллекта.

К нейрофилософии, таким образом, часто обращаются как к философии нейронаук, но подобный подход не совсем верен. Конечно, в рассматриваемой теории проблемы, связанные с искусственным интеллектом, можно легко отделить от остальных проблем. Однако в отношении нейрофилософии и философии нейронаук подобный подход неприемлем. Следует с осторожностью применять к понятию «нейрофилософия» термин «философия нейронауки». Разумнее отдавать предпочтение более точным и устоявшимся терминам «философия нейробиологии», «филосо-

фия когнитивных наук», «философия искусственного интеллекта» и др. Нейрофилософия в большей мере относится к философии сознания и отличается от иных разделов философии сознания тем, что она опирается более всего на нейронаучные теории и данные нейронаучных экспериментов.

Вместе с тем на примере теории Эдельмана становится очевидным тот факт, что разделить философию и методологию науки, философию искусственного интеллекта и философию сознания в полной мере не всегда возможно. Полноценный философский анализ нейронаучных теорий включает в себя и исследование методологии, и анализ исследовательской программы, и анализ философских допущений, находящийся в основе исследовательской программы, а также изучение возможных следствий и выводов из предлагаемых теорий. Фактически такое исследование допустимо считать междисциплинарным, затрагивающим и философию сознания, и когнитивные исследования, и искусственный интеллект. Если же сосредоточиться на методологических и метанаучных основах теории, то можно перейти в контекст философии науки и не использовать термин «нейрофилософия».

Коннекционизм как синтез идей философии нейронаук и философии искусственного интеллекта

Коннекционизм – это один из подходов к моделированию строения мозга, точнее, к моделированию биологических нейронных сетей. Суть главного принципа состоит в следующем: мышление как процесс может быть описано с помощью простых вычислительных элементов, которые связаны в сети. Как и большинство аксиом и принципов, принимаемых в любой науке, исследуемый принцип по своей сути относится к философии науки. Для биологической сети элементами являются нейроны, а связи – это синапсы. Для искусственной сети элементами служат, например, физические элементы памяти, а связями – провода, или все это может быть смоделировано в компьютерной программе.

Наиболее успешная реализация характеризуемого подхода – искусственные нейронные сети. Как правило, они состоят из большого количества соединенных элементов, разделенных на слои. Последние соответствуют трем классам: входной, нейроны которого получают информацию для обработки, выходной, нейроны которого выдают результаты обработки, и т.н. скрытый слой. Как замечает К. Бакнер, «если бы нейронная сеть модели-

ровала всю человеческую нервную систему, входные единицы были бы аналогичны сенсорным нейронам, выходные единицы – моторным нейронам, а скрытые – всем остальным нейронам» [Buckner 2019].

Между нейронами существуют связи, которые имеют вес активации. Сигнал от слоя нижележащих нейронов подается по линиям связи в слой вышележащих. Этот сигнал умножается на число веса активации, и, если суммарное значение от всех входящих линий превышает порог активации нейрона, он переходит в возбужденное состояние и посылает сигнал дальше. В ином случае распространение сигнала блокируется. Обучить сеть – значит подобрать значения веса активации таким образом, чтобы минимизировать расхождение между входными и выходными данными.

Наиболее интересна интерпретация термина «информация», которая в описанной системе представлена весами связей между нейронами. Она сосредоточена не в нейроне, а в связях между ними, т.е. имеет не локальный, а дистрибутивный характер. Среди задач, решаемых искусственной нейросетью, – распознавание образов, кластеризация (разбиение множества входных сигналов на классы), прогнозирование, аппроксимация, сжатие данных и т.д. Методологическая база изложенного подхода относится к философии сознания и представлена теорией функционализма, теориями сознания высших порядков. Функционализм интерпретирует сознание как сложную функцию между входными данными для сенсорных нейронов и поведением, интерпретируемым как выходные данные моторных нейронов. Такая теория, по сути, убирает понятия сознания и самосознания как необходимые для процесса моделирования. Мозг интерпретирован как процессор, а психология становится теорией о том, какие операции нужно произвести, чтобы добиться нужного поведения.

Коннекционистская модель в настоящее время применяется и в обратную сторону – к изучению мозга. С. Сеунг отмечает: «Коннекционисты рассматривают зоны мозга не как некую элементарную ячейку, а как сложную сеть, которая состоит из множества нейронов» [Сеунг 2014, 20]. Теоретически в мозге происходят следующие процессы, если рассматривать его с точки зрения коннекционистской модели: изменение удельного веса существующих связей, рекомбинация, переподключение, реге-

нерация [Сеунг 2014, 14]. Только первое свойство из этого списка характерно и для искусственных нейросетей. Остальные свойства обеспечены биологической природой мозга и, по-видимому, могут быть сведены к такому свойству, как пластичность.

Сегодня искусственные нейросети обладают рядом проблем, наличие которых препятствует видеть в них слабый искусственный интеллект. Самая главная из проблем – отсутствие понимания, т.е. у системы нет семантической интерпретации данных, которые подаются ей на вход. В более общем смысле проблема связана с тем, что в нейросети не запрограммировано понятие «здравого смысла» и концепции общего устройства мира. Иными словами, если биологическое сознание работает с символами и следует причинно-следственным связям, то искусственная нейросеть работает с числами и следует закономерностям. Такая проблема не принадлежит частным наукам – она сугубо философская, поэтому может быть отнесена к философии искусственного интеллекта и, на наш взгляд, решение следует ожидать в этой области.

Другая проблема – тонкие надстройки, которые могут привести к неверной классификации изображений. Например, добавление «шума» к изображению лица способно помешать работе программ распознавания лиц. Речь идет уже о конкретно-научной проблеме. Она описана, к примеру, в одной из недавних статей под названием «Слон в комнате» [Rosenfeld, Zemel, Tsotsos 2018]. Предложена гипотетическая ситуация: классификация объектов, находящихся в комнате. Дан набор фотографий. На некоторые из них посредством программ обработки изображений поместили уменьшенное изображение реального слона. При этом программа распознавания образов могла не замечать слона во многих местах или классифицировать его неверно, или неправильно классифицировать объекты, расположенные рядом.

На примере коннекционизма можно увидеть, где проходит разделение между философией науки, философией нейронаук и философией искусственного интеллекта. Базой подхода являются концепции философии науки в целом и философии сознания в частности. И рефлексия над подходом, и рефлексия проблем – это философия искусственного интеллекта. Ввиду приведенного примера становится очевидным, что любое фундаментальное теоретическое исследование по своей сути всегда будет междисциплинарным.

Заключение

Подведем итоги нашего исследования. Представление о нейрофилософии исключительно как о попытке оправдания элиминативного материализма устарело и не соответствует действительности. Представление о нейрофилософии как о философии нейронаук не вполне корректно: поскольку под нейрофилософией понимается целый ряд направлений, термин «нейрофилософия» становится чрезмерно нагруженным, и каждый раз при его использовании необходимо уточнять, что имеется в виду. Предлагаем разграничить три термина.

Нейрофилософия – это направление в философии начала XXI века, использующее нейронаучные концепции для решения традиционных философских проблем. Философия нейронаук, по нашему мнению, – чрезвычайно абстрактный термин, который содержит исследование научно-философских концепций, лежащих в основе той или иной нейронауки, а также изучение и анализ методов, проблем и целей отдельных нейронаук. Таким образом, философия нейронаук может быть рассмотрена в первую очередь как раздел философии науки, который формулирует и решает проблемы нейронаук. Более корректен в методологическом аспекте отказ от этого понятия в пользу таких терминов, как философия нейроэтики, философия нейробиологии. Философия искусственного интеллекта – это философско-методологическая база для изучения небиологического интеллекта. В таком понимании предложенное определение находится в отношениях пересечения с философией нейронаук.

Методологии науки известно, что любое исследование проходит два этапа на своем пути. Первый этап – анализ и дифференциация гипотез, касающихся объяснений изучаемого факта. Второй этап – синтез общих и отличающихся моментов в гипотезах, который позволяет создать абстрактную обобщенную теорию изучаемого явления. В становлении нейронаук и их научно-философского базиса наблюдается все еще пребывание на первом этапе. Философия нейронаук как база всех существующих нейронаучных теорий возникнет через определенное время, и именно в соответствующий период будет необходим данный термин.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Алексеев, Кузнецов, Савельев, Янковская 2015 – *Алексеев А.Ю., Кузнецов В.Г., Савельев А.В., Янковская Е.А.* Становление отечественной нейрофилософии // Философские науки. 2015. № 11. С. 48–66.

Дубровский 2015 – *Дубровский Д.И.* Нейрофилософия и проблема сознания // Философские науки. 2015. № 11. С. 9–22.

Ревонсуо 2013 – *Ревонсуо А.* Психология сознания. – СПб.: Питер, 2013.

Сеунг 2014 – *Сеунг С.* Коннектом. Как мозг делает нас тем, что мы есть. – М.: БИНОМ. Лаборатория знаний, 2014.

Философия… 2006 – Философия. Энциклопедический словарь / под ред. А.А. Ивина. – М.: Гардарики, 2006.

Эдельман 2012 – *Эдельман Дж.* Сознание: помнимое настоящее // Эволюционная эпистемология. Антология. – М.; СПб.: Центр гуманитарных инициатив, 2012. С. 419–438.

Bickle, Mandik, Landreth 2019 – *Bickle J., Mandik P., Landreth A.* The Philosophy of Neuroscience // Stanford Encyclopedia of Philosophy / ed. E.N. Zalta. 2019. – URL: https://plato.stanford.edu/archives/fall2019/entries/neuroscience

Buckner 2019 – *Buckner C.* Connectionism // Stanford Encyclopedia of Philosophy / ed. by E.N. Zalta. 2019. – URL: https://plato.stanford.edu/entries/connectionism

Churchland 1986 – *Churchland P.S.* Neurophilosophy: Toward a Unified Science of the Mind-Brain (Computational Models of Cognition and Perception). – Cambridge, MA: MIT Press, 1986.

Edelman 2007 – *Edelman G.M.* Learning in and from Brain-Based Devices // Science. Vol. 318. No. 5853. P. 1103–1105.

Gold, Roskies 2008 – *Gold I., Roskies A.L.* Philosophy of Neuroscience // The Oxford Handbook of Philosophy of Biology / ed by M. Ruse. – Oxford: Oxford University Press, 2008. P. 349–380.

Jungert 2017 – *Jungert M.* Neurophilosophy or Philosophy of Neuroscience? What Neuroscience and Philosophy Can and Cannot Do for Each Other // The Human Sciences after the Decade of the Brain. Perspectives on the Neuro-Turn in the Social Sciences and the Humanities / ed. by E. Hildt, J. Leefmann. – London: Academic Press, 2017. P. 3–13.

Rosenfeld, Zemel, Tsotsos 2018 – *Rosenfeld A., Zemel R., Tsotsos J.* The Elephant in the Room // ArXiv. – URL: https://arxiv.org/abs/1808.03305

REFERENCES

Alekseev A.Y., Kuznetsov V.G., Saveliev A.V., & Yankovskaya E.A. (2015) The Formation of the National Neurophilosophy. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* No. 11, pp. 48–66. (In Russian).

Bickle J., Mandik P., & Landreth A. (2019) The Philosophy of Neuroscience. In: Zalta E.N. (Ed.) *The Stanford Encyclopedia of Philosophy.* Retrieved from https://plato.stanford.edu/archives/fall2019/entries/neuroscience

Buckner C. (2019) Connectionism. In: Zalta E.N. (Ed.) *The Stanford Encyclopedia of Philosophy.* Retrieved from https://plato.stanford.edu/entries/connectionism

Churchland P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind-Brain (Computational Models of Cognition and Perception).* Cambridge, MA: MIT Press.

Dubrovsky D.I, (2015) Neurophilosophy and the Problem of Consciousness. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* No. 11, pp. 9–22 (in Russian).

Edelman G.M (2001) Consciousness: The Remembered Present. *Annals of the New York Academy of Sciences. Vol. 929, no.* 1, pp. 111–122 (Russian translation in: Kniazeva E.N. (Ed.) *Evolutionary Epistemology. Anthology* (pp. 419–438). Moscow; Saint Petersburg: Tsentr gumanitarnykh initsiativ, 2012).

Edelman G.M. (2007) Learning in and from Brain-Based Devices. *Science.* Vol. 318, no. 5853, pp. 1103–1105.

Gold I. & Roskies A.L. (2008) Philosophy of Neuroscience. In: Ruse M. (Ed.) *The Oxford Handbook of Philosophy of Biology* (pp. 349–380). Oxford: Oxford University Press.

Ivin A.A. (Ed.) (2004) *Philosophy: Encyclopedic Dictionary.* Moscow: Gardariki. (In Russian).

Jungert M. (2017) Neurophilosophy or Philosophy of Neuroscience? What Neuroscience and Philosophy Can and Cannot Do for Each Other. In: Hildt E. & Leefmann J. (Eds.): *The Human Sciences after the Decade of the Brain. Perspectives on the Neuro-Turn in the Social Sciences and the Humanities* (pp. 3–13). London: Academic Press.

Revonsuo A. (2009) *Consciousness: The Science of Subjectivity.* Hove: Psychology Press (Russian translation: Saint Petersburg: Piter, 2013).

Rosenfeld A., Zemel R., & Tsotsos J. (2018) *The Elephant in the Room.* arXiv. Retrieved from https://arxiv.org/abs/1808.03305

Seung S. (2012) *Connectome: How the Brain's Wiring Makes Us Who We Are.* Boston: Houghton Mifflin Harcourt (Russian translation: Moscow: BINOM, 2014).