

Значение нейронаучных исследований сознания для разработки общего искусственного интеллекта (методологические вопросы)*

© 2022 г. Д. И. Дубровский

Институт философии РАН, Москва, 109240, ул. Гончарная, д. 12, стр. 1.

E-mail: ddi29@mail.ru \ Верстка статьи, Вопросы философии, 2022, № 2

Развитие искусственного интеллекта вступило в новый этап, связанный с проблемой создания Общего искусственного интеллекта (в международном наименовании – AGI), приближающегося по своим функциям к способностям естественного интеллекта. Наряду с феноменологическими исследованиями когнитивных структур субъективной реальности важным ресурсом для создания AGI является использование результатов нейронаучных исследований сознания. В статье рассматриваются наиболее важные из них, которые могут служить основой для моделирования специфических когнитивных архитектур AGI. Это требует теоретического решения вопросов о характере связи явлений субъективной реальности с мозговыми процессами. Автором предлагается информационный подход, который позволяет объяснить эту связь и тем самым обосновать использование нейронаучных описаний когнитивных операций для компьютерного моделирования специфических когнитивных архитектур AGI. С этой целью анализируются результаты наиболее значительных нейронаучных исследований сознания. Необходимым условием успешной разработки AGI является организация более тесного сотрудничества между специалистами в области методологии науки и эпистемологии, искусственного интеллекта и нейронауки.

Ключевые слова: искусственный интеллект, естественный интеллект, общий искусственный интеллект (AGI), сознание, субъективная реальность, Я, когнитивная архитектура, информация, нейродинамическая система, расшифровка мозговых кодов, эпистемология.

DOI: 10.21146/0042-8744-2021-?-??-???

Цитирование: *Дубровский Д.И.* Значение нейронаучных исследований сознания для разработки общего искусственного интеллекта (методологические вопросы) // Вопросы философии. 2022. № 2. С. ?–?

* Исследование проведено при финансовой поддержке гранта Министерства науки и высшего образования РФ (проект «Новейшие тенденции развития наук о человеке и обществе в контексте процесса цифровизации и новых социальных проблем и угроз: междисциплинарный подход», соглашение № 075-15-2020-798).

The Importance of Neuroscientific Research of Consciousness for the Development of General Artificial Intelligence (Methodological Issues)*

© 2022 David I. Dubrovsky

Institute of Philosophy, Russian Academy of Sciences, 12/1, Goncharnaya str., Moscow, 109240, Russian Federation.

E-mail: ddi29@mail.ru

Received 07.05.2021

The development of artificial intelligence has entered a new stage associated with the problem of creating General Artificial Intelligence (AGI), approaching in its functions to the abilities of natural intelligence. Along with the phenomenological studies of the cognitive structures of subjective reality, the most important resource for the creation of AGI is the use of the results of neuroscientific studies of consciousness. The article discusses the most important of them, which can serve as a basis for modeling specific cognitive architectures of AGI. This requires a theoretical explaining the nature of the connection between the phenomena of subjective reality and brain processes. The author proposes an informational approach that allows explaining this connection and thereby substantiate the use of a neuroscientific description of cognitive operations for computer modeling of specific cognitive architectures of AGI. For this purpose, the results of the most important neuroscientific studies of consciousness are analyzed for the development of AGI. A prerequisite for this is the organization of closer cooperation of specialists in the field of methodology and epistemology, artificial intelligence and neuroscientific research.

Keywords: artificial intelligence, natural intelligence, general artificial intelligence (AGI), consciousness, subjective reality, I, cognitive architecture, information, deciphering brain codes, epistemology.

DOI: 10.21146/0042-8744-2021-?-???-???

Citation: Dubrovsky, David I. “The Importance of Neuroscientific Research of Consciousness for the Development of General Artificial Intelligence (Methodological Issues)”, *Voprosy Filosofii*, Vol. 2 (2022), pp. ?-?

Развитие искусственного интеллекта (далее – ИИ) выдвинуло новую проблему стратегического масштаба – создание Общего ИИ, обозначаемого в принятой международной номенклатуре – AGI (Artificial General Intelligence). Недавно издан коллективный труд, в котором впервые в нашей научной литературе обозначен круг основных задач, от решения которых зависит создание AGI [Ведяхин (ред) 2021]. Эти задачи выражают мировой тренд развития ИИ в условиях нарастающей конкуренции между крупнейшими научными центрами и специализированными корпорациями, а в более широком масштабе – между Россией и такими лидирующими в области ИИ государствами, как США, Китай, Япония, страны Западной Европы. Догонять их по основным сложившимся приоритетным направлениям развития ИИ – бессмысленно. Нам

* The research was carried out with a financial support of the Ministry of Science and Higher Education of the Russian Federation (Project “New tendencies of the humanities and social sciences development in the context of digitalization and new social problems and threats: interdisciplinary approach”, Agreement № 075-15-2020-798).

нужен прорыв! Прорыв на принципиально новый уровень развития ИИ, о чем говорил Президент России В.В. Путин. От этого зависит наше экономическое развитие, решение проблем безопасности, многих важных социальных вопросов.

Для такого прорыва необходим новый методологический подход, требующий учета и использования результатов исследований сознания, которые выносились за скобки в классической методологии А.Тьюринга; В связи с этим ставятся и решаются вопросы о развитии посттьюринговой методологии ИИ [Efimov 2020], о привлечении к разработке АГІ широких междисциплинарных исследований сознания, особенно на стыке психологии с нейрофизиологией и неврологией.

Первостепенное значение нейронаучных исследований сознания и естественного интеллекта в целом для достижения нового уровня в развитии ИИ не раз подчеркивали ведущие представители как нейронауки [Анохин 2017] и др., так и специалисты по искусственному интеллекту [Кузнецов 2016] и др. Нейронаучные исследования сознания раскрывают «аппаратный», воплощенный в мозговых динамических структурах аспект функционирования сознания. При этом принципиальное значение имеет решение вопросов о способе связи явлений сознания с мозговыми процессами, то есть главных вопросов так называемой «Трудной проблемы сознания»: 1) как объяснить связь явлений сознания (субъективной реальности) с мозговыми процессами, если первым нельзя приписывать физические свойства, а вторые ими необходимо обладают и 2) как теоретически корректно объяснить способность явлений субъективной реальности служить причиной телесных изменений. Возникает также вопрос о свободе воли и ряд других. Однако главное значение имеют первые два вопроса. Они решаются на основе предложенного автором информационного подхода. Его основные положения были сформулированы еще в давних работах [Дубровский 1968; Дубровский 1971; Дубровский 1980] и с тех пор развивались и уточнялись в связи с достижениями нейронаучных исследований [Дубровский 2015; Дубровский 2019]. В наиболее развитом и сжатом виде теоретическое решение основных вопросов «Трудной проблемы сознания» изложено в статье, опубликованной в ведущем международном *нейронаучном журнале* [Dubrovsky 2019a]. Отношение между явлением субъективной реальности, например, переживаемым мной в данном интервале зрительным образом определенного объекта (обозначим его **A**), которое рассматривается в качестве информации о нем, и соответствующим ему мозговым процессом (обозначим его **X**), есть отношение между информацией и ее носителем. Анализ показывает, что связь между **A** и **X** является функциональной, возникшей на основе сложившейся в процессе филогенеза и онтогенеза *кодовой зависимости*; **A** и **X** – явления одновременные и однопричинные, находятся в отношении взаимнооднозначного

соответствия, а это позволяет рассматривать ряд существенных свойств **A** для построения моделей **X**, и, наоборот, использовать полученные нейронаукой результаты исследований свойств **X** для объяснения осознаваемых когнитивных процессов человека, а тем самым и для построения их компьютерных моделей. Создание все более эффективных нейрокомпьютерных интерфейсов позволяет использовать их в целях развития ИИ, осмысления продуктивных способов разработки AGI.

Здесь, прежде всего, представляют интерес результаты нейронаучного исследования организации психических процессов, тех операциональных динамических структур, которые специфичны для осуществления сознательной деятельности. За последние десятилетия в этом направлении нейронаукой достигнуты значительные успехи, связанные с методами картирования и визуализации мозговых процессов. Они имеют прямое отношение к разработке когнитивных архитектур AGI и интеллектуальной робототехники. Рассмотрим некоторые наиболее важные достижения нейронауки под углом интересующих нас вопросов.

Результаты исследований зеркальных нейронов и их систем

Это направление позволило пересмотреть классические представления о связи сенсорных и моторных функций в деятельности головного мозга. Было показано, что зеркальные нейроны одновременно осуществляют «кодирование сенсорной информации в моторных терминах», то есть создают единство зрительного восприятия и его вероятной моторной репрезентации [Риццолатти, Синигалья 2012]. Это свидетельствует о том, что «в то время как индивид наблюдает за выполнением действий другими, в его мозге выполняется потенциальный моторный акт, по всем своим характеристикам сходный с тем, который спонтанно активируется при подготовке и выполнении реального действия. Различие лишь в том, что в одном случае это действие остается потенциальным, а в другом переводится в последовательность конкретных движений» [Риццолатти, Синигалья 2012, 97]. Благодаря нашим зеркальным системам, мы способны незамедлительно распознавать определенные типы действий, производимых другими, достигая понимания другого. Это указывает на особенности организации зрительного образа (объединения в нем *отображения* предмета и *вероятного способа действия с ним*, а также санкционирующего *акта понимания*, «принятия» смысла этого действия), Такое единство трех операций способно обеспечить весьма эффективное взаимодействие между людьми, и это согласуется с результатами феноменологических исследований динамической структуры субъективной реальности. Все это побуждает к поискам возможности создания подобных трехмерных функциональных структур у робота.

Особенно важно иметь в виду, что именно на моторных репрезентациях наблюдаемых действий, создаваемых зеркальными нейронами, «основана *способность обучения при помощи подражания*» [Риццолатти, Синигалья 2012, 97] (курсив мой – Д.Д.). Это обстоятельство может служить стимулом для осмысления пути построения моделей расширения диапазона «навыков» интеллектуального робота, его способности умножать свои «навыки» в данном классе действий, которая столь существенна для построения AGI. Но при этом необходимо учитывать особенности операций категоризации, производимых зеркальными системами. Они осуществляют не только семантическую категоризацию (категоризацию объектов), но и *моторную категоризацию* (категоризацию действий). Проблема категоризации действий в связи с категоризацией объектов восприятия имеет принципиальное значение для расширения деятельных способностей интеллектуального робота, а вместе с этим для создания таких когнитивных архитектур AGI, которые будут способны к «пониманию» причинно-следственных отношений.

Рассматривая процессы обучения ребенка, Дж. Риццолатти и его сотрудники подчеркивают *«первичность» моторных категоризаций*, более того, говорят о том, что «моторная категоризация обеспечивает фундамент, на котором в дальнейшем будет возведено здание перцептивного опыта, источника семантической категоризации» [Риццолатти, Синигалья 2012, 56]. В связи с этим идет речь о «словаре моторных актов», который регулирует способность выполнения действий. Эти вопросы также имеют прямое отношение к задачам моделирования когнитивных архитектур AGI.

Наряду с исследованием зеркальных механизмов сенсорномоторных процессов в последние годы открыты и изучены те мозговые структуры, которые ответственны за эмоциональные состояния, их восприятие и понимание другими людьми. Это относится ко всем так называемым первичным эмоциям (страх, гнев, боль, удивление, радость и т.д.), выработанных в ходе эволюции и сохраняющих свою адаптивную функцию. Акт мгновенного понимания эмоций другого обусловлен тем, что в зеркальной системе у наблюдающего возникают «потенциальные висцеромоторные акты», которые в актуальной своей форме как раз и вызывают переживание соответствующих эмоций [Риццолатти, Синигалья 2012, 159]. Таким образом, зеркальные системы образуют фундаментальный механизм коммуникации, социальных взаимодействий между индивидами и могут служить для поисков создания более эффективных *способов коммуникации* между людьми и роботами и между самими роботами. Обобщая результаты исследования зеркальных нейронов и их систем, можно видеть, что они конкретизируют многие существенные аспекты ключевого принципа сознательной

деятельности – единства иноотображения и самоотображения, причем на всех уровнях: от сенсорного до интеллектуально–эмоционального и целостной структуры нашего Я. При этом весьма важно учитывать результаты эпистемологических исследований познавательной деятельности субъекта, взятого не только в классическом, но и в постклассическом ключе [Лекторский (ред) 2017; Лекторский 2019].

«Чтение мозга» (Brain-Reading).

Это сравнительно новое направление нейронауки, которое использует методы картирования и визуализации мозговых процессов, формирует и исследует разнообразные нейродинамические корреляты психических явлений. Оно нацелено на расшифровку их мозговых кодов и достигло уже существенных результатов в области изучения явлений субъективной реальности. Еще более десяти лет тому назад японские исследователи Ё. Мияваки, Ю. Камитани и их сотрудники расшифровали нейродинамические эквиваленты зрительного восприятия статичных черно-белых объектов (отводя сигналы от мозга испытуемых, воспроизводили на экране компьютера переживаемые ими в данном интервале зрительные образы). При этом они смогли расшифровывать не только образы непосредственно воспринимаемого объекта, но и воспоминание о нем. Вслед за ними нейрофизиологи из Института Беркли в Калифорнии (Ш. Нишимото, Э. Гэлбрейт и др.) получили близкий результат с цветными движущимися образами (воспроизводя на экране компьютера фрагменты воспринимаемого человеком цветного фильма).

За последние годы на основе технологии глубоких нервных сетей расшифровка мозговых кодов психических явлений значительно усовершенствовалась. Нейробиологи и информатики из Киотского университета в Японии предложили новый способ непосредственно визуализировать по данным фМРТ изображение, которое видит человек в момент сканирования мозга [Shen et al. 2017]. Впечатляющий результат получен в области прямого перевода из мозга (то есть декодирования) текста из 30-50 предложений на английском языке с помощью отведения сигналов примерно от 250 пунктов перисильвиевой коры мозга испытуемого. При этом средняя частота ошибок в словах не превышала 3% [Makin et al. 2020].

К настоящему времени установлено большое число достаточно четких корреляций между определенными психическими явлениями и соответствующими нервными процессами. Они широко используются для создания новых нейротехнологий в медицине, различных интерфейсах «мозг-компьютер-машина», применяются в робототехнике, где их перспективы могут быть особенно значительными в области создания гибридных

человеко-робототехнических систем с прямыми интерфейсами «мозг-интеллектуальный робот».

«Чтение мозга» позволяет строить модели когнитивных процессов. Но для этого требуется решение ряда теоретико-методологических вопросов. И здесь опять же на первом плане феноменологический анализ, задача четкого выделения и описания того мыслительного акта, то есть определенного явления субъективной реальности, нейродинамический коррелят которого мы хотим установить. Установление коррелята психического явления ведет к расшифровке его мозгового кода, то есть к пониманию содержащейся в нем определенной информации. Накопленный в нейронауке опыт расшифровки кодов может быть полезен при разработке способов коммуникации между человеком и роботом и между самими роботами, для создания в этих целях эффективных методов кодирования и декодирования информации.

Эго-система головного мозга и нейронаучные исследования коррелятов «Я»

С позиций современной нейронауки наше Я представляет особую структурно-функциональную организацию, которую называют Эго-системой головного мозга или Самостью. Она воплощает личностные свойства индивида, включая их генетический и биографический уровни. Она служит объектом многообразных нейронаучных исследований. Наиболее значительные их результаты представлены в фундаментальной монографии А. Дамасио [Дамасио 2018]. В плане неврологии, при анализе патологических нарушений исключительно ценный материал для понимания Самости содержится в широко известной книге Вилейанура Рамачандрана [Рамачандран 2014]. Эти вопросы служили в последние десятилетия предметом изучения для ряда крупных представителей нейронауки (Дж. Эдельман, К. Кох, Дж. Тонони и др.). Среди советских нейрофизиологов представляют интерес работы Д.П. Матюшкина [Матюшкин 2007].

А. Дамасио выделяет в структуре самости три уровня: 1) протосамость, 2) базовую самость, 3) автобиографичную самость.

Протосамость – это первичные телесные ощущения, которые спонтанно возникают в бодрствующем состоянии и составляют основу всех прочих ощущений и переживаемых эмоций. Она включает, по терминологии А. Дамасио: 1) «ведущие интероцептивные карты» (содержание которых строится совокупностью сигналов, поступающих из внутренней среды организма и внутренних органов); 2) «ведущие карты организма», отображающих схему всего тела и его основных частей – головы, туловища и конечностей; 3) карты «сенсорных порталов, управляемых извне», совокупность тех телесных структур, которые участвуют в произведении сенсорного акта. Все эти аспекты

протосамости тесно взаимосвязаны, они определяют рамки необходимой стабильности биологических процессов в организме, его уникальность и целостность, его «позицию наблюдателя» в мире, что служит основанием базовой и автобиографической самости.

Базовая самость, по словам А. Дамасио, «возникает за счет соединения видоизмененной протосамости с объектом, который вызвал изменения и теперь особо отмечен ощущениями и выделен вниманием» [Дамасио 2018, 231]. На этом уровне возникает качество «субъективности», ощущение собственного существования, собственного знания, определение его ценностного значения. «Из этого ощущения знания рождается ощущение присвоения образов и ощущение активного участия» [Дамасио 2018, 237].

Автобиографичная самость – это интегральная самость, включающая ее нижележащие уровни путем постоянного взаимодействия с ними. Она воспроизводит все наши воспоминания, сохраняя единство и идентичность Я, создает способность его проекций в будущее. Возникновение автобиографической самости у человека ознаменовало качественно новый этап развития субъективной реальности по сравнению с животным миром, Но это нисколько не умаляет ее фундаментальной зависимости от протосамости и базовой самости. Более того, осмысление интегральной самости (Самости в целом, Эго-системы) требует тщательного анализа ее глубинных биологических основ.

А. Дамасио подробно исследует те морфологические структуры мозга и его нейродинамические функциональные системы, сложные взаимосвязи между верхними отделами мозгового ствола, таламусом и корой мозга которые лежат в основе Самости. С этой целью он выделяет фундаментальную операцию *конвергенции-дивергенции* и те зоны мозговых процессов, где они осуществляются. «Зоны конвергенции-дивергенции (ЗКД) представляют собой совокупность нейронов, в границах которой контактируют между собой множество петель прямой и обратной связи. ЗКД поддерживает «прямую» связь с сенсорными областями, расположенными «ниже» в цепи обработки сигналов. Цепь эта начинается в точке, где сенсорный сигнал попадает в кору головного мозга. В эти области ЗКД отправляет собственные проекции, осуществляя таким образом обратную связь. Кроме того, ЗКД посылает проекции, находящиеся на следующем уровне цепочки, и получает от них обратную связь» [Дамасио 2018, 167]. ЗКД являются микронеонами, образующими в мозгу обширные ОКД (области конвергенции-дивергенции). Они «играют важную роль в создании и упорядочивании важнейших аспектов содержимого психики, наделенной сознанием, и в том числе тех фрагментов, из которых складывается автобиографическая самость» [Дамасио 2018, 170].

Следует подчеркнуть, что указанные операции и функционирование ЗКД и ОКД служат необходимыми факторами *самоорганизации Самости*, а, следовательно, нашего Я. Таким образом, А. Дамасио рассматривает Самость и в качестве объекта – мозгового носителя сознания, и в качестве субъекта – самого сознания, нашего Я, во всех его проявлениях (то есть Я тоже является объектом познания Я). На это важно обратить особое внимание. Анализ содержания и динамической структуры Я отчетливо показывает широкие *способности Я к саморегуляции и самоорганизации*, касающиеся как способности поддержания своей идентичности, так и оптимизации функций познавательной и практической активности. Эти весьма актуальные вопросы тесно связаны с проблемой свободы воли, которая широко обсуждается в последнее время представителями нейронауки. Я опускаю здесь эту важную тему, так как она специально рассматривается мной в статье «Проблема свободы воли и современная нейронаука» [Dubrovsky 2019b].

Учет операций конвергенции-дивергенции и способов саморегуляции Эго-системы головного мозга (соответственно, Я) может быть полезен при моделировании когнитивной архитектуры интеллектуального робота в целях расширения его «субъектности», способов его самооптимизации, то есть самообучения, по крайней мере, в зоне разных вариантов задачи одного типа. Это относится и к поискам новых подходов к разработке AGI.

Исследования представителей российской нейронауки

Значительный интерес представляют, на мой взгляд, прежде всего исследования К.В. Анохина и В.Я. Сергина.

К.В. Анохин указывает на принципиальные недостатки программ, предложенных DARPA и IBM, рассчитанных на преодоление экстенсивного развития архитектуры ИИ и поиски нового пути создания машины не фон-неймановской архитектуры, работающей на принципах операций, производимых головным мозгом. Он убежден, что для создания «биологоподобного ИИ» необходима нейронаучная теория естественного интеллекта, «она должна решать фундаментальный вопрос о том, откуда из операций нервной системы берутся свойства разума и сознания» [Анохин 2017, 77]. Без этого, как считает Анохин, теория не может быть эффективной.

Когнитом – это нейронная гиперсеть мозга. Автор вводит «три ключевые понятия»:
1) «когнитивная группа» («ког»), это – группа нервных клеток, представляющая собой в то же время «элемент опыта, который кодирует тот или иной аспект соотношения когнитивного агента с теми или иными аспектами предметной среды» [Анохин 2017, 81];
2) связь этих когов, образующих более широкие устойчивые группы («локи»), которые

приобретают системные свойства более высокого порядка; 3) в результате «у когнитивных систем появляются новые сетевые свойства, которые могут быть исследованы, формализованы, просчитаны, измерены и описаны в рамках современных теорий сложных сетей и теории графов» [Анохин 2017, 81]. Хотя концепция К.В. Анохина и вызывает ряд существенных вопросов, особенно в отношении возникновения качества субъективной реальности, она, в отличие от других известных нейронаучных концепций, делает заметный шаг на пути исследования функционирования нейронной гиперсети мозга, построения ее математизированных моделей, а постольку может быть особенно полезна для разработок AGI. Эти вопросы подробнее обсуждаются в его недавно вышедшей обширной статье [Анохин 2021].

Весьма ценными в интересующем нас плане являются результаты многолетних исследований В. Я. Сергина, предложившего концепцию иерархической модели восприятия, в которой центральную роль играет функция, названная автором «объемлющей характеристикой» [Сергин 2009; Сергин, Сергин 2019]. Процесс формирования восприятия проходит ряд уровней от низшего к высшим, в результате чего в итоге формируется целостный сенсорный образ объекта. Однако при построении иерархической модели восприятия возникает острая проблема так называемого «комбинаторного взрыва», то есть экспоненциального роста числа нейронов, необходимых для отображения сенсорного объекта при увеличении числа его признаков. Проблема такого рода возникает и в области ИИ, особенно при создании систем распознавания образов, управления роботами и др. [Рассел, Норвиг 2015]. Проблема комбинаторного взрыва, как полагают специалисты, остается весьма существенной в области применения поисковых алгоритмов, к которым могут быть сведены практически все задачи интеллектуального распознавания, принятия решений и управления [Нагоев 2016]. В концепции В.Я. Сергина показано, что сенсорные системы мозга снимают эту проблему путем формирования объемлющей характеристики на каждом уровне организации восприятия, так как она представляет собой не сумму признаков, а их синтез, определяемый адаптивными целями субъекта. Иерархия объемлющих характеристик – это иерархия абстрагирования вследствие исключения множества нерелевантных и мелкомасштабных деталей на каждом следующем уровне обработки. Были проведены специальные компьютерные эксперименты, которые убедительно показали адекватность модели объемлющих характеристик [Сергин 2015].

Это открывает новые возможности исследований процессов восприятия, но в то же время, может служить задачам построения моделей нейроморфных когнитивных архитектур для развития интеллектуальной робототехники и AGI.

Новейшие исследования нейронных коррелятов сознания

В последнее десятилетие был проведен цикл обширных систематических исследований нейронных коррелятов сознания (НКС), результаты которых могут представлять существенный интерес для наших целей. Исследования проводились под углом определения тех структур головного мозга, которые непосредственно ответственны за переживаемое в данном интервале сознательное состояние, то есть за само качество субъективной реальности (как без учета его конкретного содержания, но также и при его выяснении). Обобщение результатов исследований, проведенных с помощью различных методов, в том числе с учетом анализа патологических случаев и следствий нейрохирургических операций по удалению отдельных участков мозга, позволило считать, что анатомические нейронные корреляты сознания локализируются преимущественно в задней, так называемой «горячей» зоне коры (posterior cortical hot zone), охватывающей сенсорные области, а не в лобно-теменной коре [Koch et al. 2016]. Некоторые префронтальные регионы проявляют участие в реализации таких аспектов активности сознания как внимание, чувство рефлексии, оценка эмоций, тогда как другие непосредственно ничего не вносят в переживание состояний сознания.

При этом выделяется «полная НКС», представляющая любое состояние сознания, и «контент-специфическая НКС», выражающая определенное его содержание. В последнем случае имеют в виду расшифровку нейродинамического кода конкретного явления субъективной реальности в данном интервале, то есть результаты исследований, которые выше обозначались нами как «чтение мозга». Здесь важно отдавать себе отчет в различии между установленным коррелятом данного конкретного по содержанию явления субъективной реальности и его действительной мозговой кодовой структурой. Коррелят представляет собой чаще всего лишь вход в эту кодовую структуру, способ ее активации. Это не более чем прелюдия к расшифровке кода, имеющего чрезвычайно сложную динамическую структуру, охватывающую многие корковые и подкорковые участки мозга, как в его вертикальном, так и в горизонтальных измерениях. В этой связи исследователи отмечают необходимость более основательного изучения «меры дифференциации и интеграции» участия разных мозговых структур в производстве явлений сознания, подчеркивают, что новые данные о роли задней горячей зоны коры требуют дальнейших исследований.

Такие исследования были продолжены с участием большинства авторов рассмотренной выше статьи. Они показали, что многие данные противоречат друг другу, не позволяют в развернувшейся дискуссии прийти к общепринятым заключениям.

Дискуссия приобрела очень широкий характер. Авторы в новой статье реферируют и оценивают около 100 публикаций на эту тему [Boly et al. 2017]. Они склоняются к признанию все же большей роли функций фронтальной коры в реализации сознательных психических процессов. Особенно это относится к свойствам процессов мышления.

Такое положение дел можно истолковать в том смысле, что реальный мыслительный процесс при решении нетривиальных задач является многомерным, включает такие весьма различные состояния субъективной реальности, которые не могут быть выражены четкими, рационально определенными характеристиками, но играют очень важную роль в осознаваемом процессе решения задачи. Имеются в виду интуитивные факторы, слабо выраженные эмоциональные состояния, феномены поддержания целеполагания, интенционального вектора, волевого усилия и, что особенно важно, проявления функций саморегуляции, тонких поправок в ходе реализации мыслительного процесса. Именно эти его недостаточно явные компоненты и выступают в качестве функций соответствующих нейронных комплексов *фронтальной коры*, которые «не видны» в получаемых коррелятах или же крайне слабо и противоречиво отображаются в них. Это обусловлено ограниченной разрешающей способностью используемых методов, а, с другой стороны, недостаточной феноменологической проработкой тех явлений субъективной реальности, которые берутся в качестве объектов для определения их нейродинамических коррелятов (в первую очередь это относится к «контент-специфическим НКС», которые требуют профессионального феноменологического описания). Приведенные обширные материалы исследований нуждаются в тщательном анализе, могут быть весьма полезными для целей разработки AGI.

Разумеется, эти и другие, отмеченные выше концептуальные описания результатов нейронаучных исследований сознания, лишь начало большой работы, требующей участия аналитика, системного архитектора, программиста ряда других специалистов, создающих технический проект и обеспечивающих его реализацию. Организация систематического сотрудничества специалистов в области эпистемологии и методологии, искусственного интеллекта и нейронауки – необходимое условие успешной разработки Общего искусственного интеллекта.

Ссылки – References in Russian

Анохин 2017 – Анохин К.В. Когнитивные вычисления на основе нейронных гиперсетей // Философия искусственного интеллекта. Под ред. В. А. Лекторского, Д. И. Дубровского, А.Ю. Алексеева. М.: ИИнтелл, 2017. С. 70–86.

Анохин 2021 – *Анохин К.В.* Когнитом: В поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности, 2021. том 71. № 1. С. 39–71.

Ведяхин (ред) 2021 – Сильный искусственный интеллект: На подступах к сверхразуму / Ведяхин А. [и др.]. М.: Интеллектуальная Литература, 2021.

Дамасио 2018 – *Дамасио А.* Так начинается «я». Мозг и возникновение сознания. М. Карьера-Пресс, 2018.

Дубровский 1968 – *Дубровский Д. И.* Мозг и психика // Вопросы философии. 1968, № 8. С. 119–135.

Дубровский 1971 – *Дубровский Д. И.* Психические явления и мозг. Философский анализ проблемы в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики. М: Наука, 1971. С. 279–299. (Изд. 2-е, доп., М.: ЛЕНАНД, 2020).

Дубровский 1980 – *Дубровский Д. И.* Информация, сознание, мозг. М.: Мысль, 1980. (Изд. 2-е, доп., М.: ЛЕНАНД, 2020).

Дубровский 2015 – *Дубровский Д. И.* Проблема «Сознание и мозг»: Теоретическое решение. М.: Канон+, 2015.

Дубровский 2019 – *Дубровский Д. И.* Проблема сознания. Теория и критика альтернативных концепций. М.: ЛЕНАНД, 2019.

Кузнецов 2016 – *Кузнецов О. П.* Избранные труды: автоматы, языки и искусственный интеллект. М.: Наука, 2016.

Лекторский (ред.) 2017 – Субъективный мир в свете вызовов современной когнитивной науки. / Общ. ред. и сост. В.А. Лекторского (Отв. редактор), Е.О. Труфановой, А.Ф. Яковлевой. М.: Аквилон, 2017.

Лекторский 2019 – *Лекторский В.А.* Эпистемология классическая и неклассическая. М.: URSS, 2019.

Матюшкин 2007 – *Матюшкин Д.П.* О возможных нейрофизиологических основах природы внутреннего «Я» человека // Физиология человека. 2007. Т. 33, № 6. С. 1–9

Нагоев 2016 – *Нагоев З.В.* Нейроподобные мультиагентные рекурсивные когнитивные архитектуры в задачах интеллектуальной обработки информации и управления // Нейроинформатика–2016. Лекции по нейроинформатике./ Ю.В. Тюменцев (ред.). М.: МИФИ, 2016. С. 12–45.

Рамачандран 2014 – *Рамачандран В. С.* Мозг рассказывает. Что делает нас людьми. М.: Карьера-Пресс, 2014.

Рассел, Норвиг 2015 – *Рассел С., Норвиг Р.* Искусственный интеллект: современный подход. 2-е изд. М.: Вильямс, 2015.

Риццолатти, Синигалья 2012 – *Риццолатти Д., Синигалья К.* Зеркала в мозге. О механизмах совместного действия и сопереживания / перевод с англ. О.А Кураковой и М.В. Фаликман. М.: Языки славянских культур, 2012.

Сергин 2009 – *Сергин В. Я.* Психофизиологические механизмы восприятия: концепция объемлющих характеристик // *Успехи физиологических наук.* 2009. Т.40 (4). С. 42–63.

Сергин 2015 – *Сергин А.В.* Компьютерная модель распознавания: концепция объемлющих сенсорных характеристик // *Нейрокомпьютеры: разработка, применение.* 2015. № 8. С. 33–39.

Сергин, Сергин 2019 – *Сергин В. Я., Сергин А. В.* Иерархическая модель восприятия без комбинаторного взрыва // *Журнал высшей нервной деятельности.* 2019. Том 69. № 5. С. 633–658.

References

Anokhin, Konstantin V. (2017) Cognitive Computing Based on Neural Hypernets, Lektorsky, Vladislav A., Dubrovsky, David I., A. Yu.. Alekseev, Andrei Yu. (eds.), *The Philosophy of Artificial Intelligence*, Iintell, pp. 70–86 (in Russian).

Anokhin, Konstantin V. (2021) Kognitom: In Search of a Fundamental Neuroscientific Theory of Consciousness, *Journal of Higher Nervous Activity*, Volume 71, No. 1. P. 39–71 (in Russian).

Boly, Melanie, Massimini, Marcello, Tsuchiya, Naotsugu, Postle, Bradley et al. (2017) Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence, *The Journal of Neuroscience*, Oct. 4, 2017. 37(40), 9603–9613.

Damasio, Antonio (2010) *Self Comes to Mind. Constructing the Conscious Brain*, Pantheon Books, New York (Russian translation).

Dubrovsky, David I. (1968) Brain and Psyche, *Voprosy Filosofii*, Vol. 8 (1968), pp. 119–135 (in Russian).

Dubrovsky, David I. (1971) *Mental Phenomena and the Brain. Philosophical Analysis of the Problem in Connection with Some Urgent Problems of Neurophysiology, Psychology and Cybernetics*, Nauka, Moscow, pp. 279–299 (in Russian).

Dubrovsky, David I. (1980) *Information, Consciousness, Brain*, Mysl, Moscow (in Russian).

Dubrovsky, David I. (2015) *The Problem “Consciousness and the Brain”*: Theoretical solution, Kanon +, Moscow (in Russian).

Dubrovsky, David I. (2019) *The Problem of Consciousness. Theory and Criticism of Alternative Concepts*, LENAND, Moscow (in Russian).

Dubrovsky, David I. (2019a) “‘The Hard Problem of Consciousness’”. Theoretical Solution of Its Main Questions’, *AIMS Neuroscience*, 2019, Vol. 6, No. 2, pp. 85–103.

Dubrovsky, David I. (2019b) The Problem of Free Will and Modern Neuroscience, *Neuroscience and Behavioral Physiology*, 2019, Vol. 49, No. 5, pp. 629–639.

Efimov, Albert (2020) Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence, *Artificial General Intelligence (AGI)*, 2020, Vol. 12177, pp. 83–94.

Koch, Christof, Massimini Marcello, Boly, Melanie, Tononi, Giulio (2016) Neural Correlates of Consciousness: Progress and Problems, *Nature Reviews Neuroscience*, Volume 17 (5), pp. 307–321.

Kuznetsov, Oleg P. (2016) Selected Works: Automata, Languages and Artificial Intelligence, Nauka, Moscow (in Russian).

.Lektorskiy, Vladislav A. (ed.). (2017) The Subjective World in the Light of the Challenges of Modern Cognitive Science, Akvilon , Moscow (in Russian).

Lektorsky, Vladislav A. (2019) *Classical and Non-classical Epistemology*, URSS, Moscow (in Russian).

Makin, Joseph G., Moses, David A. and Chang, Edward F. (2020) “Machine Translation of Cortical Activity to Text with an Encoder–Decoder Framework”, *Nature Neuroscience*, Vol. 23, pp. 575–582.

Matyushkin, Dmitry P. (2007) The Possible Neurophysiological Basis of the Inner Self, *Human Physiology*, Vol. 33, No. 6. pp. 1–9 (in Russian).

Nagoev, Zalimkhan V. (2016) Neuro-like Multi-agent Recursive Cognitive Architectures in the Tasks of Intelligent Information Processing and Control, Tyumentsev, Yuri V. (ed.) *Neuroinformatics–2016. Lectures on Neuroinformatics*, MIFI, Moscow, pp. 12–45 (in Russian).

Ramachandran, Vilayanur (2011) *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human*, W.W. Norton & Co (Russian translation).

Rizzolatti, Giacomo, Sinigaglia, Corrado (2008) *Mirrors in the Brain – How Our Minds Share Actions and Emotions*, Oxford University Press (Russian translation).

Russell, Stuart, Norvig, Peter (2010) *Artificial Intelligence: A Modern Approach*, University of California, Berkeley (Russian translation).

Sergin, Vladimir Ya. (2009) Psychophysiological Mechanisms of Perception: the Concept of Encompassing Characteristics, *Advances in Physiological Sciences*, Vol. 40 (4). pp. 42–63 (in Russian).

Sergin, Alexei V. (2015) Computer Model of Recognition: the Concept of Enveloping Sensory Characteristics, *Neurocomputers: Development, Application*, Vol. 8 (2015), pp. 33–39 (in Russian).

Sergin, Vladimir Ya., Sergin, Alexei V. (2019) Hierarchical Model of Perception without a Combinatorial Explosion, *Journal of Higher Nervous Activity*, Volume 69, No. 5. pp. 633–658 (in Russian).

Shen, Guohua, Horikawa, Tomoyasu, Majima, Kei, Kamitani, Yukiyasu Deep Image Reconstruction from Human Brain Activity (2017 web), URL: <https://www.biorxiv.org/content/biorxiv/early/2017/12/30/240317.full.pdf>

Vedyakhin Alexander (ed.) (2021) *Strong Artificial Intelligence: Approaching the Superintelligence*, Intellectual Literature, Moscow (in Russian).

<i>Сведения об авторе</i>	<i>Author's Information</i>
ДУБРОВСКИЙ Давид Израилевич – доктор философских наук, профессор, главный научный сотрудник сектора теории познания Института философии РАН.	DUBROVSKY David Izrailevich– DSc in Philosophy, Professor, Chief Researcher of the Sector of Theory of Knowledge of the Institute of Philosophy of the Russian Academy of Sciences.