

ОРИГИНАЛЬНАЯ СТАТЬЯ

DOI: 10.26794/2226-7867-2024-14-4-89-100
УДК 165.12(045)

Модели объяснения сознания в контексте проблемы определения сильной формы ИИ

В.Н. Белов, С.А. Малофейкин

Российский университет дружбы народов, Москва, Россия

АННОТАЦИЯ

В статье предпринята попытка дополнить современную проблематику выделения критериев сильного ИИ посредством дискуссий в области философии сознания. Получившие популярность представления Д. Деннета («многократных набросков»), Дж. Сёрля (каузально-эмерджентное описание) и Д. Чалмерса (синтетический подход к пониманию сознания) сопоставляются с историей формирования проблемы ИИ. Несмотря на широкое обсуждение проблематики сознания и искусственных форм интеллекта (сильного и слабого), остаются актуальными теории и рассуждения философов о психофизиологической проблеме. Предполагается, что понимание механизмов аналитической работы сознания, раскрытие творческого потенциала личности, а также способность воспринимать множество явлений в их категориальных формах и формировать аксиоматические и синтетические суждения расширят инструментарий машинного обучения. Для дополнения имеющихся представлений о сознании в условиях распространенности информационного подхода (Д.И. Дубровский) и аналитической традиции (В.В. Васильев) приводятся ключевые положения психофизиологической проблемы, обозначенные в истории немецкой и русской философии. Учитывая сложность и многогранность обозначенных проблем (определение сознания, психофизиологическая проблема, определение ИИ, демаркация слабых и сильных форм ИИ, значение языка для выстраивания структур мышления, аналоговое мышление и его возможности), содержание статьи ограничивается анализом наметившихся тенденций в философии и обозначением перспектив для дальнейшего углубления в проблематику.

Ключевые слова: философия сознания; психофизиологическая проблема; искусственный интеллект; аналитическая философия; информационный подход

Для цитирования: Белов В.Н., Малофейкин С.А. Модели объяснения сознания в контексте проблемы определения сильной формы ИИ. *Гуманитарные науки. Вестник Финансового университета*. 2024;14(4):89-100. DOI: 10.26794/2226-7867-2024-14-4-89-100

ORIGINAL PAPER

Models for Explaining Consciousness in the Context of the Problem of Determining the Strong Form of AI

V.N. Belov, S.A. Malofeikin

Peoples' Friendship University of Russia, Moscow, Russia

ABSTRACT

The article attempts to complement the modern problems of highlighting the criteria of strong AI through discussions in the field of philosophy of consciousness. The popular ideas of D. Dennett ("multiple sketches"), J. Searle (causal emergent description) and D. Chalmers (synthetic approach to understanding consciousness) are compared with the history of the formation of the AI problem. Despite the wide discussion of the problems of consciousness and artificial forms of intelligence (strong and weak), the theories and arguments of philosophers about the psychophysiological problem remain relevant. It is assumed that clarifying the mechanism of analytical work of consciousness, the creative potential of the individual, the ability to cover a variety of phenomena in categorical forms, building axiomatic and synthetic judgments will expand the tools of machine learning. To complement the existing ideas about consciousness in the context of the prevalence of information approaches (D.I. Dubrovsky) and the analytical tradition (V.V. Vasiliev), the key provisions of the psychophysiological problem identified in the history of German and Russian philosophy are given. Given the complexity and versatility of the identified problems (definition of consciousness, psychophysiological problem, definition of AI, demarcation of weak and strong forms of AI, the importance of language for building structures of thinking, analog thinking and its capabilities), the content of the article is limited to analyzing emerging trends in philosophy and identifying prospects for further deepening into the problem.

Keywords: philosophy of consciousness; psychophysiological problem; artificial intelligence; analytical philosophy; information approach

For citation: Belov V.N., Malofeikin S.A. Models for explaining consciousness in the context of the problem of determining the strong form of AI. *Gumanitarnye Nauki. Vestnik Finansovogo Universiteta = Humanities and Social Sciences. Bulletin of the Financial University*. 2024;14(4):89-100. DOI: 10.26794/2226-7867-2024-14-4-89-100

ВВЕДЕНИЕ В ПРОБЛЕМАТИКУ: ИСТОРИЧЕСКИЙ ЭКСКУРС

В настоящее время происходит повсеместное внедрение новейших технологий анализа и обработки данных в социальную и даже научную практику. Значение информации и процесса цифровизации становится предметом широких дискуссий, среди которых отметим обсуждение идей Н. Винера, А. Тьюринга, Дж. фон Неймана, М. Гарднера и А. Чёрча. Перспектива искусственных форм анализа массивов данных, близких к возможностям человека, и оптимизация машинного обучения обозначились еще во второй половине XX века. Однако для первооткрывателей технологий машинного обучения и ИИ имплицитным остался вопрос о необходимости воспроизведения инструментария человеческого мышления, специфических особенностей сознания, умения творческого восприятия и охвата многообразия явлений в категориальных формах. В первую очередь рассматривались проблемы воспроизводимости аналитических возможностей человека, а понятия сознания, мозг и интеллект остались без должной апробации. В 1948 г. Н. Винер, один из основателей кибернетики, публикует известный труд «Кибернетика, или управление и связь в животном и машине», в котором обозначилась проблематика сведения искусственных форм обработки и передачи информации к возможностям человеческого разума систематизировать окружающую действительность. Примечательно, что ученый использует уже накопленный опыт не только в рамках математики, но и в областях психологии и логического позитивизма, упоминая парадоксы математической логики Б. Рассела и практические результаты И. П. Павлова [1, р. 127]. Ссылаясь на достижения фундаментальной науки и философии, в том числе на идеи Л. Больцмана, американский математик имплицитно обозначает несводимость аналитических возможностей искусственной обработки информации к потенциалу человека формировать обобщающие системы знания, функционирующие в науке, даже при неполноте представлений о конкретных физических явлениях [1, р. 37]. Иначе говоря, вычислительным машинам не хватает абстраги-

рования от конкретности действительности и от предметности чисел. Во многом ученый оказался прав, и на сегодняшний день есть как минимум две проблемы, с которыми не справляются слабые формы ИИ и нейронные сети — проблема остановки и проблема переобучения. Способность к обучению, но не к качественному изменению стратегии проработки решения, отражает современные пределы интеллектуальных систем в постановке собственных целей при достижении поставленных задач. В свою очередь «проблема остановки», восходящая еще к интерпретации Д. Гилбертом универсальности принципов А. Тьюринга в определении подмены человека машиной и к теории неполноты математики К. Гёделя, остается актуальной в XXI в. В наиболее общем виде суть проблемы выражается в невозможности машиной, что справедливо и для современных форм ИИ, использовать математические подходы обобщения и переноса данных.

О потенциале и ограничениях вычислительных машин Дж. фон Нейман упоминает в своем труде «Вычислительная машина и мозг» (написан в 1956 г. и опубликован после смерти автора в 1958 г.). Обозначив различие между аналоговым, цифровым и химическим способом обработки информации, математик указал на ограничения для использования исключительно одной из них. Но при внедрении принципа «взаимозаменяемости» способов вычисления или «чередования аналоговых и цифровых процессов» повышается общая результативность [2, с. 137]. Особенность и преимущество человеческого мозга перед аналоговыми средствами выражается во взаимозаменяемости цифрового и химического анализа информации, а также в повышении скорости обработки из-за иных конфигураций строения отделов памяти человека при молекулярно-атомарных размерностях в противоположность технологическим возможностям хранения информации. Человеческий мозг, уступая в скорости и точности процессов вычисления, способен охватить «логическую глубину» процессов в параллельной, а не в последовательно-аналитической форме [2, с. 151]. Уступая вычислительной машине в точности и скорости исчисления, человек способен

обобщить более объемные данные, обнаружив связь там, где она не предполагается аналитическим способом. Оставляя вне контекста вопросы о сознании и личности, математик обозначает приоритет развития вычислительных технологий по пути объединения аналоговых и цифровых механизмов обработки данных, а дальше к объединению химических структур нейронных связей человека с цифровыми технологиями.

Рассуждения Н. Винера и Дж. фон Неймана показательны для понимания сути проблемы несводимости сознания к вычислительной машине, т.е. сознание обладает своими преимуществами, а машины воспроизводят и улучшают только отдельные функции мозга, превосходя аналитический потенциал естественного носителя сознания. Подчеркнем, что именно принципы кибернетики и параллелизма процессов вычисления легли в основу создания нейронных сетей, т.е. аналоговых форм воспроизведения вычислительных функций мозга, но не сознания.

После Дартмутского семинара 1956 г. термин ИИ стал набирать популярность и одновременно с этим приобретать размытые очертания, вбирая в объем понятия значительно различающиеся процессы и явления. Очевидно, что, когда Дж. Маккарти на Дартмутском семинаре обозначил термин ИИ («artificial intelligence» или AI) и его наиболее широкую формулировку, когнитивные науки и в целом проблематика сознания не были оформлены. Под определение лауреата премии Тьюринга подходит все многообразие вычислительных машин и программно-обеспечения, поэтому дальнейшая история проблемы связывается не только с развитием информационных технологий и функционального программирования, в том числе машинного обучения, но и с методологией науки и философией. Хотя Н. Винер, Дж. фон Нейман, а также А. Тьюринг и А. Чёрч вполне отчетливо в своих докладах и публикациях обозначили различие между воспроизведением структур мозга и мышления с искусственной формой сознания, только к концу XX в. появились авторитетные заявления по данному вопросу.

Под влиянием нарастающего количества неточностей в определениях и используемой терминологической базы, провоцирующих методологические и проектные упущения, ученые и специалисты в областях математики, информатики и машинного обучения, в частности, обращаются к накопленному потенциалу философии.

В том числе оформляется проблема этического применения интеллектуальных систем, которая становится актуальной [3, р. 937]. Именно в философии обозначаются наиболее неудобные вопросы, а очевидные представления и общеизвестные мнения ставятся под сомнение и последовательно реконструируются, уточняя основы объективного мировоззрения и корректируя границы здравого смысла.

История философии показывает фундаментальность вопроса об истоках мышления, сознательной жизни и когнитивных процессов в целом. Несмотря на то, что когнитивные науки после 1960-х гг. (в особенности благодаря «теории Сантьяго» и распространению междисциплинарного подхода в науке и философии) базируются на естествознании, но в процессе концептуализации достигнутых результатов неизменно отсылают к философии, к проблематике эпохи Нового времени, философии немецкого классического идеализма и к аналитической традиции. Особую популярность в среде философов, занимающихся проблематикой сознания, соотношения сознания и мозга, и в частности «теорией тождества», приобретает именно аналитическая традиция. Несомненно, что Г. Фреге и Р. Карнап, Дж. Мур и Б. Рассел, а также Л. Витгенштейн обозначили ряд важнейших идей для развития современных представлений о функциях и потенциале когнитивных способностей. Однако исследователи новейшего времени отсылают в своих трудах и публичных выступлениях не только к аналитической традиции, но и к классическим вопросам XVII–XVIII вв. (дуализму Р. Декарта, ментальной географии Д. Юма, трансцендентальному субъекту И. Канта) и системным формам философии XX в. (например, теории коммуникации Ю. Хабермаса [4, 5]). Данное обстоятельство подчеркивает потенциал философии в освоении сложнейших вопросов современности.

ПРИРОДА СОЗНАНИЯ И ИНТЕЛЛЕКТ: ПОСЛЕ 1994 ГОДА

Конференция «На пути к науке о сознании» в городе Туссан (штат Аризона) 1994 г. стала отправной точкой для изменения ракурса рассмотрения проблемы сознания в преломлении к вопросу о возможности создания искусственного интеллекта. Наиболее широко обсуждаемым стал доклад Д. Чалмерса о сложных и легких проблемах сознания. Воспроизведение отдельно взятых механизмов обработки и передачи информа-

ции не являются сложной проблемой сознания и в полной мере могут быть решены в рамках эмпирических исследований. Трудная проблема выражается в поиске ответа на природу и отличительную сущностную характеристику сознания, отличающую его от любых аналитических способов работы с информацией. Напомним, что уже Дж. фон Нейман указал на параллелизм обработки информации как на принципиально отличную характеристику человеческого мозга. Аналитический подход, использующий строгий детерминизм и равномерное линейное распределение потоков информации, не может объяснить высокий функционал мозга при очевидной более низкой производительности вычислений. Теперь стало очевидным, что ранее предложенное определение ИИ Дж. Маккарти не удовлетворяет исследователей в областях когнитивных наук. Дж. Сёрль усиливает зарождающиеся тенденции и проводит демаркацию между слабой и сильной формой ИИ, а для ученых становится актуальной проблематика создания не искусственного интеллекта, а моделирования именно сознания, когнитивной целостности механизмов человеческого разума. В свою очередь в практико-прикладных исследованиях все чаще начинают отказываться от термина ИИ в интерпретации 1956 г., обозначая новые технологические решения как интеллектуальные системы или методы искусственного интеллекта, подчеркивая несводимость вопроса о методах обработки информации к проблеме моделирования когнитивной целостности сознания. Как итог, философия заняла господствующую позицию в комплексном объяснении природы сознания, объединяя исследования когнитивных процессов, проблематику создания сильного ИИ, область информационных и цифровых процессов, структур мышления (в том числе логика и языкознание) сквозь призму соотношения «сознание — личность», «сознание — мозг» и «мозг — тело».

Одну из наиболее оригинальных позиций, объясняющих сущность сознания, предложил Д. Деннет. В противоположность подходу Дж. Сёрля и его «китайской комнаты», рождается «бостонский зомби» и «картезианский театр» [6, с. 202]. Основываясь на функционализме и исходя из достижений естествознания, американский философ объясняет сознание как производный механизм от физиологического устройства человека, но подчеркивает наличие особых форм организации психики. Во-первых, функционализм Д. Деннета выстраивается на компромиссной

позиции между классическим эволюционизмом и детерминированным монизмом. Человеческое сознание обусловлено особенностью устройства мозга, который является продуктом эволюции и культурного развития. Язык и формы логического мышления, формируемые в опыте истории и постоянно совершенствующиеся благодаря научному прогрессу, влияют на усложнение сознания и когнитивных практик. Во-вторых, сознание не исходит из какой-либо заданной точки или области в строении мозга, а является результатом целостности функционирования человека как биологического существа. Процессы в мозге осуществляются не по принципам прямолинейной аналитики и строгой детерминации, что характерно для искусственных аналитических систем. Человек обладает сознанием, так как все когнитивные процессы, которые могут быть описаны как информационные, в том числе взаимодействие «сознание — тело» и «мозг — тело» распределены параллельно, перекрещиваясь и многослойно. Здесь обнаруживается отсылка к исследованиям Дж. фон Неймана и, в свою очередь, критика А. Тьюринга и Н. Винера в понимании устройства мозговой активности. В 1990-х гг. ученый вводит понятие «интенциональных установок», объясняющих мотивацию к принятию и исполнению решений. Уровень развития языковых средств, используемых человеком, играет первостепенную роль в процессе образования сознания, формируя нарративы и первичность «интенциональных установок». Добавим, что «интенциональность» Д. Деннета отличается по содержанию и качеству от идей Э. Гуссерля и его трансцендентальной феноменологии. Интенциональность (intentionality) определяет мотивацию к действию, а общей моделью (программой) становится интенциональная установка, присущая всем живым объектам: «по определению, интенциональными системами являются все те и только те объекты, поведение которых предсказуемо или объяснимо из интенциональной установки. Самореплицирующие макромолекулы, термостаты, амебы, растения, крысы, летучие мыши, люди и компьютеры, играющие в шахматы, все это интенциональные системы, одни более интересные, другие менее» [7, с. 41]. Человек является объектом, обладающим «более интересной системой» установок «намеренностей» что-либо сделать, т.е. активность и сложность человеческого поведения обусловлена исключительно физиологическими установками организации мозга.

В рассуждениях Д. Деннета можно обнаружить аналогию с французским материализмом эпохи Просвещения. Однако философ дополняет свои представления на природу сознания «моделью многократных набросков» (Multiple Drafts Model) и информационным подходом. Данная модель не только оригинально описывает происхождение когнитивных актов, но и отражает современные тенденции в проектировании наиболее результативных методов искусственного интеллекта. Весь полученный опыт, в том числе и в процессе коммуникации преобразуется в информацию. Данное положение можно раскрыть через теорию коммуникации Ю. Хабермаса и его представления о различных уровнях передачи информации. Далее в рассуждениях Д. Деннета сознание не фокусируется на точности и детализации воспроизведения полученных данных, а стремится многократно предложить решение или мотивацию к действиям на основе той информации, которая поступила, но еще не обработана. При этом новый опыт постоянно добавляется к уже обрабатываемым процессам, постоянно расширяя область предлагаемых решений. Проблема в том, как и каким способом происходит выбор. В попытке преодолеть затруднение Д. Деннет активно использует объяснения мотивации поведения биологических видов из общей теории эволюции и научно-популярные сведения из биологии, генетики, нейронаук. В итоге теория американского философа была дополнена критическим обзором дарвинизма в труде «Опасная идея Дарвина: Эволюция и смысл жизни» (Darwin's Dangerous Idea. Evolution and the Meaning of Life, 1995).

Отличительной особенностью и, одновременно, противоречием является отрицание каких-либо форм интуиции, интроспекции и апелляции к внутреннему содержанию опыта человека. При отрицании подобного уровня восприятия или взаимодействия с окружающей средой бессмысленным становится введение «интенциональной установки», обуславливающей интенциональность. С другой стороны, американский ученый предлагает схематическое представление сознательных процессов и открывает возможность для улучшения антропоморфности интеллектуальных систем, придавая им внешние человеческие проявления, в том числе в осуществлении коммуникации. Безусловно, исследования Д. Деннета имеют значение для оптимизации интеллектуализированных систем и для развития методов искусственного интеллекта. Важно отметить, что

для интеллектуальных систем в целом придание им человеческих черт не является главным. Более важным является расширение функционала для решения конкретных задач.

Модель сознания Дж. Сёрля построена на разделении синтаксических и семантических структур. Наиболее ярко данное разделение выражено в схеме «китайской комнаты», где отчетливо показана демаркация простых операций алгебраической последовательности и сложных семантических структур. С его позиции сознание определяется тремя аспектами, составляющими единство: качественность, субъективность и единство [8, с. 301]. Рациональность поступка зависит от согласованности трех проявлений сознания. В свою очередь личность «не то же самое, что поле сознания, но ее деятельность, поскольку в силу оснований она принимает решения и приводит их в исполнение, нуждается в едином поле сознания, содержащем такие познавательные элементы, как ощущения и воспоминания, и, кроме того, такие волевые элементы, как размышления и действия» [8, с. 303]. Критикуя различные формы физикализма и картезианские формы дуализма, Дж Сёрль формирует модель сознания как каузально-эмерджентного объекта, который может быть субъективным в онтологическом понимании.

Особая заслуга философа состоит в разделении интеллектуальных систем на две формы — сильный и слабый ИИ. Возможность создания искусственного сознания Дж. Сёрль отрицает, так как никакая форма интеллекта, построенная на синтаксически ограниченном поле, не способна к безграничности формирования семантических систем. Иначе говоря, для ИИ всегда потребуется что-то зависящее от внешнего оператора, в том числе контекст решения проблемы, массив данных, поставленная цель или даже простая загрузка программного обеспечения. Отличительным признаком сильной формы ИИ в контексте исследований Дж. Сёрля будет являться способность к самостоятельной организации деятельности и выстраивания семантических структур. Но даже в данном случае форма ИИ не будет приближена к обладателю естественного сознания, так как необходимо удвоение системы посредством обратного взаимодействия между психической и непосредственно сознательной активностью человека. Первая во многом обусловлена физиологическими процессами, а психическое может быть субъективным явле-

нием, но у носителя естественного сознания они представлены неразрывно. Если воспроизвести только часть когнитивной активности человека, то целостность не будет достигнута. Подобная позиция очень близка холизму, когда целое превосходит совокупность частей.

В качестве третьей модели выделим подход Д. Чалмерса, который критически описывает существующие теории сознания, в том числе физикализм в различных формах и модели феноменализма. Функциональный анализ оказывается близок общим представлениям философа, но в то же время подвергается особой критике, поскольку с его позиции редуцирует сложность сознания к простым актам «стимул — реакция». Человеческое сознание явно более сложная модель, чем упрощенная система рефлекторного проявления на внешние раздражители.

Основой для описания идей Д. Чалмерса можно обозначить раннюю работу «Сознающий ум: В поисках фундаментальной теории». В ней позиция австралийско-американского ученого определяется как натуралистический дуализм, но в дальнейшем изменяется на плюрализм в контексте одновременности признания объективности существования сознания как независимого феномена и утверждения, что мозг влияет на когнитивные процессы. Однако вне зависимости от принадлежности к установившейся традиции в философии сознания Д. Чалмерс обозначил эпистемологически значимую демаркацию. Комплекс вопросов о сознании человека и сознании в целом могут быть разделены на две группы: трудную и легкую проблему. К «легким проблемам» относятся исследования механизма и функционирования когнитивных процессов. Легкую часть вопроса о сознании составляют «обучение и память — функциональные свойства, характеризующиеся каузальными ролями, и поэтому вопрос «как физическая система могла бы обладать психологическим свойством P?» тождествен вопросу «как некое состояние физической системы могло бы играть такую-то каузальную роль?». Это вопрос для наук о физических системах» [9, с. 44]. Подчеркнем, что подобные исследования становятся основой для разработки и моделирования интеллектуальных систем, методов искусственного интеллекта и т.д.

Сложная или трудная проблема лежит за пределами усложненных форм механизма, детерминизма физических явлений и бихевиоризма, т.е. «совершенно иначе обстоит дело с феноме-

нальными аспектами ментального. Здесь проблема соотношения ментального и телесного столь же тяжела, как и раньше. Впечатляющий прогресс физических и когнитивных наук не пролил большого света на вопрос о том, как и почему когнитивные функциональные процессы сопровождаются сознательным опытом. Прогресс в понимании ментального сводился почти исключительно к объяснению поведения. Он не коснулся вопроса о сознательном опыте» [9, с. 45]. Поскольку Д. Чалмерс приобрел значительную популярность, его идеи воспроизводятся в различных контекстах и в отношении к взаимоисключающим концепциям. Однако очевидно, что обозначение несогласованности апостериорного и субъективного опыта с целостностью сознания как личности сквозь призму волновой функции Луи де Бройля (в преломлении к теории «множественных миров» Х. Эверетта) и квантовую теорию реальности, значительно более сложная попытка описать сознание, чем простой натуралистический дуализм (Глава 10–11. «Сознающий ум: В поисках фундаментальной теории»). В Главе 10 Д. Чалмерс указывает, что позиция Х. Эверетта хотя и не ясна эпистемологически, но имеет общее основание с его гипотезой о сознании, которую «точнее назвать интерпретацией одного большого мира. Есть только один мир, но он больше, чем мы могли бы подумать» [9, с. 430].

Если упрощенно изложить позицию Дэвида Чалмерса, можно выделить одно важное следствие, касающееся проблемы определения критериев сильного искусственного интеллекта. Использование моделей аналитической последовательности и изначальная предопределенность процессов обработки данных не позволят достичь функций, аналогичных сознанию. Сильная форма ИИ может быть неотличима от подавляющего большинства проявлений когнитивной активности человека, но достичь «взаимозаменяемости» возможно только при устранении индивидуальных различий. Однако в подобном случае устранялось бы и различие между личностями как носителями индивидуальности, т.е. сознаниями, обогащенными уникальным опытом. Если отбросить представления об уникальности каждого носителя сознания, согласившись с общими положениями информационного подхода, то возникает трудность крайней формы эпифеноменализма.

Хотя исследования Д. Чалмерса учитывают достижения квантовой физики и естествозна-

ния, стоит отдельно указать на понимание проблемы ИИ выдающегося английского ученого Р. Пенроуза, специалиста в различных областях математики, общей теории относительности и квантовой теории. Обобщая работы Дж. Сёрля и Д. Деннета с уже установившимися фактами науки, английский ученый в своей известной работе «Новый ум короля» пишет: «я считаю, что проблема понимания свойств сознания гораздо более многогранна, чем можно извлечь из экспериментов с ИИ. Тем не менее, я уверен в необходимости признания этой области исследований и уважительно отношусь к ней» [10, с. 57]. Далее обозначается ряд критических замечаний о возможности моделирования процессов, которые не ясны и не имеют однозначных исходных точек анализа, кроме субъективных и логических предпосылок. Если будет доказана алгоритмическая основа работы мозга и операциональная природа сознания, то результатом станет достижение только «эквивалентной машины Тьюринга» [10, с. 68]. Р. Пенроуз очень точен в своих формулировках, подчеркивая, что Тест Тьюринга обозначает «эквивалентность» когнитивных функций в алгоритмической форме, близкой к естественному сознанию, но им не тождественные. После дополнений А. Чёрча, в рамках Теста А. Тьюринга появляются усложняющие обстоятельства, но при этом упрощается специфика определения особых семантических форм человека — эмоций, чувств, симпатии и антипатии, иррациональности мотивации и наличия изменяемых состояний сознания. В дальнейшем из-за чрезмерного редукционизма понятие ИИ с момента его обозначения в 1956 г. в областях математики, кибернетики и информатики вытесняется термином «интеллектуальные системы», что прослеживается вплоть до конца XX в.

Критерием сильной формы ИИ с данной позиции становится возможность преодоления вычислительных ограничений, обусловленных непрерывностью и бесконечностью некоторых физических и математических явлений. В 1994 г. Дж. Сёрль, проведя различие между сильными и слабыми формами искусственного интеллекта, сформулировал для философов понятие «общего положения вещей» в сфере инженерных и естественных наук. В результате дискуссии о создании искусственного сознания окончательно вышли за рамки математических обсуждений и перешли в сферу философии.

ОПЫТ РУССКОЙ ФИЛОСОФИИ

В контексте истории российской философии представлены интересные и во многом оригинальные решения отдельных вопросов сознания. Преимущественно внимание философов и ученых было направлено на рассмотрение значения разума, высших познавательных способностей, связи ментальных и физических процессов. Одной из наиболее ярких дискуссий в конце XIX — начале XX в. стало обсуждение психофизиологической проблемы в контексте развития трансцендентального идеализма и распространения идей позитивистской философии [11]. Другим значимым событием, продолжившим наметившиеся тенденции, становится оформление основ общей теории психологии С.Л. Рубинштейном, где обозначились контролируемые и воспроизводимые свойства сознания. Русский психолог и философ, раскрывая соотношение между онтологическими свойствами реальности и гносеологическими способностями ее восприятия, представил позицию близкую к конструктивизму, но отличающуюся умеренностью и согласованностью с базовыми принципами естествознания. Активность сознания и его качества проявляются в осознаваемой деятельности посредством структур мозга и нервной системы, где исключительно рефлекторная теория не способна объяснить внутренние психические явления [12, с. 173]. Но сознание становится активным субъектом, только если осуществляется возможность рефлексии собственной деятельности (в терминологии С.Л. Рубинштейна «осознавание»). Позиция ученого включает сложную структуру описания осознаваемых и исключительно деятельностных актов человека, позволяя разграничить осознанные поступки от дееспособного поведения и состояний «измененного сознания». Во многом благодаря влиянию более ранних представлений в русской философии и немецкой традиции «новокантианства» [13] С.Л. Рубинштейн смог выделить структуру познавательного процесса в категориях, поддающихся моделированию и, что наиболее значимо, — оцениванию со стороны внешнего наблюдателя. Таким образом, его системное описание сознания во взаимосвязи с объективной реальностью стало основой и отправной точкой для развития общей теории психологии, объединяющей не только теоретические, но и практико-прикладные подходы в исследовании сознания.

Развитие общей теории психологии и раскрытие особенностей психофизиологической

проблемы позволили смоделировать основные механизмы принятия решения. Схематизм когнитивных способностей человека в контексте прикладных достижений психологии и общетеоретических описаний системы «мозг — тело» стали макетом для создания первых алгоритмов обработки информации. Не случайно в трудах Н. Винера и первых теоретиков машинного обучения присутствуют отсылки к исследованиям в областях психологии и философии. В частности, Н. Винер использует результаты И. П. Павлова, а Дж. фон Нейман ссылается на исследования нейронов и строения мозга человека в терминологии логического позитивизма (в частности, на Б. Рассела).

В современной традиции российской философии можно выделить как минимум два различающихся, но аргументированных взгляда на природу и системную организацию сознания, структур мышления, соотношения «мозг — сознание»: информационный подход к психофизиологической проблеме и натуралистическая версия интеракционизма. Если учитывать хронологическую последовательность, то стоит выделить Д. И. Дубровского и его информационный подход в первую очередь. Уже в конце 1960-х гг. им была оформлена оригинальная теория соотношения феномена сознания и физиологической организации когнитивных процессов. Учитывая тенденции в нейронауках, в кибернетике и появление теории самоорганизующихся систем, советско-российским ученым было предложено разрешение психофизиологической проблемы посредством объяснения психической активности как информационных потоков. Примечательно, что если Н. Винер и Дж. фон Нейман избрали отправной точкой для проектирования вычислительной техники и схематизации информационных потоков естественное строение мозга и механизмы восприятия человека, то Д. И. Дубровский идет в обратном порядке.

На основании устройства информационных потоков, т.е. приема-передачи сигнала, поступления и обработки информации, процесса вычисления ученым объясняется функционирование сознания как обработки кода. В свою очередь мозг выступает как носитель и источник для обработки предзаданного кода, который разворачивается в информационные потоки. Источником для наполнения кода информацией выступает система рефлексов и результаты восприятия органов чувств, что в полной мере согласовано с общей

теорией психологии. Однако теория рефлексов и внешних раздражителей не описывает обратный процесс, т.е. влияние сознательных усилий на формирование устойчивых моделей реакции на окружающие факторы. Более того, И. П. Павлов справедливо оставлял данные вопросы вне контекста своих практических исследований. Тем не менее информационный подход позволяет расширить моделирование естественных психических процессов для целей машинного обучения и смежных дисциплин.

В труде «Проблема идеального. Субъективная реальность» Д. И. Дубровского представлены основные положения информационного подхода: 1) информация есть результат отражения; 2) информация не существует вне материального носителя; 3) информация не существует вне кодовой формы; 4) информация обладает совокупностью синтаксических, семантических и прагматических (информация всегда подразумевает целеполагание) характеристик; 5) информация может служить фактором управления [14, с. 137]. Далее утверждается, что «всякое явление сознания (как явления субъективной реальности) есть определенная информация, присущая определенному социальному индивиду» [14, с. 139]. Как видим, сознание фундируется на основании информационных потоков, представленных как форма кода и обусловленных особенностями организации мозга человека. В более поздних работах развивается представление о возможности человека оперировать информацией «по своей воле», что и составляет основу субъективной реальности [15, с. 94]. При этом Д. И. Дубровский справедливо замечает, что «надо четко различать классическую психофизическую проблему — философскую по своей сути — и современную психофизиологическую проблему, сердцевина которой состоит в вопросе об отношении качества субъективной реальности к мозговым процессам», так как «последняя является научной проблемой» [15, с. 96]. Обратная сторона данной цитаты выражается в несводимости целостности личности к исключительно нейрофизиологическим объяснениям функционирования мозга и нервной системы. Добавим, что человек как личность буквально остается «вещью в себе» в кантовском смысле слова, но процессы когнитивной активности могут быть моделированы в искусственных формах.

В контексте развития представлений об информации как особом феномене, организующем не только «социального индивида», но и соци-

альные группы, раскрывается проблематика виртуализации реальности. Здесь стоит выделить работы Д.В. Иванова, который обосновал установление новых социальных норм и трансформацию восприятия под влиянием тесной интеграции информационных и интеллектуальных технологий в повседневную практику. В труде «Виртуализация общества. Версия 2.0» (2002) описывается, как искусственные формы моделирования, и в том числе информационные технологии, постепенно вытесняют естественное сознание в областях науки, искусства и института семьи. Как итог информационный продукт не отличается от естественных форм, например, в области искусства. Так как все процессы сводятся к закономерностям функционирования смоделированных систем, т.е. к синтетическим структурам, не предполагающим появление и активность творческого индивидуального объекта, то информационный продукт по внешним признакам обладает качествами результата деятельности человека. Таким образом, информация становится отправной точкой для выявления критериев объективности, но сама по себе она является лишь искусственным технологическим потоком данных [16, с. 200]. Таким образом, человеческая субъективность оценивается с позиции информационного обезличенного потока данных.

Экстраполировав подобные суждения на историю объяснения сознания, можно выделить следующую последовательность: первоначально алгоритмы и схематизм вычислительных машин формируются по аналогии со строением мозга и упрощенного представления о когнитивных процессах, далее развивается моделирование ключевых функций сознания (память; анализ, обобщение и концептуализация информации). Теперь с помощью моделей обмена потоков информации и механизмов анализа больших массивов данных объясняется строение мозга и функционирование когнитивных процессов человека. Позиция Д.В. Иванова позволяет эксплицитно обнаружить оформившийся герменевтический круг в процессе описания строения и функционирования мозга.

На методологической базе английской аналитической традиции сквозь призму накопленного в истории философии опыта оригинальное описание проблемы «сознания» предлагает В.В. Васильев. Его модель натуралистической версии интеракционизма в большей степени задает вопросы, чем предлагает ответы. Труд «Сознание

и вещи. Очерк феноменалистической онтологии» наиболее ярко отражает позицию российского философа. Синтезируя представления о фундаментальных основаниях когнитивной деятельности человека, в том числе критически анализируя идеи Д. Деннета, Дж. Сёрля, Д. Чалмерса и других не менее известных ученых, В.В. Васильев обозначает ключевой критерий в описании дуализма «сознание — тело», — согласованность со «здравым смыслом». Не углубляясь в особенности терминологического разграничения интеракционизма и позиции локальной супервентности сознания, укажем цитату философа, отражающую его позицию: модель локального интеракционизма «отмежевывается от либертарианства. Показывая же необходимость ментальных состояний для реализации физической причинности, она позволяет говорить о реальной действительности подобных состояний. Согласно этой концепции, мои намерения оказываются необходимым условием моих поведенческих реакций и поэтому не замещаются и не могут замещаться физическими причинами. Эти причины не являются тем, что автоматически порождает мои действия. Они могут формировать «сознание — тело» лишь при участии субъекта, который, таким образом, может рассматриваться как их подлинный соавтор, несущий ответственность за то, что он совершает (в данном случае можно говорить не только о Я, но и о субъекте в целом)» [17, с. 197]. Психическая активность побуждает к действию, в свою очередь состояния сознания обуславливаются физиологической активностью (не действиями, а именно происходящими явлениями в физиологической активности человека). В наиболее общем виде «локальный интеракционизм» означает «модель ментальной каузальности, обосновывающая каузальную действительность сознания на уровне локальных событий на основе положений феноменалистической онтологии и опровержения тезиса локальной супервентности сознания» [18, с. 148]. Иначе говоря, физическая организация тела человека оказывает влияние на то, как сознание комплексно оценивает ситуацию своего существования. Сознание, формируя приоритеты, цели и задачи, воплощает их посредством физиологической активности, которая уже в свою очередь оказывает влияние на окружающую действительность. При значительном упрощении рассуждений В.В. Васильева и А.В. Кузнецова, эксплицитно прослеживается позиция менталь-

ной географии Д. Юма, дополненная физикализмом начала XX в.

Отметим, что модель локального интеракционизма лишь косвенно относится к проблематике выявления критериев сильных форм ИИ, так как преимущественно находится в плоскости онтологии, не предоставляя материалов для схематизации когнитивных процессов. Однако здесь отчетливо прослеживается невозможность сведения сильной формы ИИ к искусственному сознанию из-за специфики процесса восприятия собственного «Я». Для того, чтобы сильная форма ИИ могла считаться сознанием, потребуется формирование у интеллектуальной системы представления о собственном отличии от человека. Далее под влиянием данного отличительного показателя должно формироваться или изменяться представление о собственном «Я». В качестве «Другого» может выступать и иная сильная форма ИИ, но тогда потребуется создание как минимум двух и более различных по функционированию искусственных форм сознания.

ВЫВОДЫ

В настоящее время стали очевидны проблемы, связанные с интеграцией интеллектуальных систем в повседневную практику, структуры высшего образования и в научно-исследовательскую деятельность. Кроме того, обнаружилось затруднение в воспроизведении когнитивных возможностей человека в более узкоспециализированных практиках, в частности обобщение многообразия полученного опыта и поиск решений вне заданного контекста поставленных задач со стороны оператора.

В то же время человек способен находить неожиданные решения, привлекая первоначально несвязанные контексты и семантические системы. Учитывая доступные формы интеллектуальных систем, нет оснований утверждать о создании сильной формы ИИ, способной к воспроизведению потенциала человеческого мышления. С другой стороны, интеллектуальные системы как комплекс различных инструментов, в том числе искусственных нейронных сетей, значительно более эффективно справляются с анализом больших данных, обработкой информации и поиском решений, но только при синтаксически правильно поставленном запросе от человека или социальной группы. Машинное обучение на сегодняшний день способно оптимизировать интеллектуальные системы под конкретные отра-

слевые или узкоспециализированные задачи, но последующее развитие методов искусственного интеллекта требует новых концептуальных моделей устройства когнитивных процессов. Таким образом, дальнейшее обсуждение вопросов развития машинного обучения и совершенствования интеллектуальных систем в целом становится тесно сопряжено с философией, в особенности с проблематикой выявления и описания отличительных особенностей сознания и структуры мышления.

После 1994 г. Дж. Сёрлем критически оценивалось понимание ИИ как некоторой формы когнитивной активности, что справедливо с позиции специалистов по методам искусственного интеллекта. Еще ранее Дж. фон Нейман вполне отчетливо указал на причину непродуктивности воспроизведения всей целостности человеческого мышления в искусственной форме. До конференции в городе Туссане термин ИИ со времен его определения Дж. Маккарти приобретает некорректность использования, что спровоцировало появление терминов «интеллектуальные системы», «методы искусственного интеллекта», «методы обработки больших массивов данных» и т.д. [19]. Несмотря на требующуюся конкретизацию того, что именно называть искусственным интеллектом и сознанием, вопрос о развитии интеллектуальных систем остается открытым. Если обобщить позиции «многократных набросков», каузально-эмерджентного описания и умеренного физикализма с достижениями общей психологии в рамках психофизиологической проблемы, то в контексте философии сознания и смежных направлений можно выделить критерии сильной формы ИИ. Во-первых, сильный ИИ или его формы не являются искусственным сознанием и не эквивалентны личности в собственном смысле слова. Во-вторых, сильная форма ИИ способна к самостоятельной постановке цели без прямого или косвенного участия оператора, задающего массив данных и область самообучения. В-третьих, должно быть преодолено ограничение современных нейронных сетей, в частности проблематика решения задач, требующих «глубинной логики» (в терминах Дж. фон Неймана), математического обобщения и переноса. В заключение отметим наиболее значимый, с нашей позиции, критерий сильной формы ИИ, близкой к искусственному сознанию — самостоятельный поиск массивов данных и адаптация в процессе принятия решения посредством параллельно-

го, а не последовательного достижения цели. Данные критерии могут стать целеполагающим основанием для специалистов машинного обучения, работающих с общими интеллектуальными системами и с наиболее перспективными интеллектуализированными системами.

СПИСОК ИСТОЧНИКОВ

1. Wiener N. Cybernetics: or control and communication in the animal and the machine. NY: M.I.T. Press; 1985. 212 p.
2. Нейман Дж. фон. Вычислительная машина и мозг. М.: АСТ; 2023. 192 с.
3. Tsvyk V.A., Tsvyk I.V., Pavlova T.P. The problematic area of philosophical discourses on the application of artificial intelligence systems in society. *RUDN Journal of Philosophy*. 2023;4(27):928–939. DOI: 10.22363/2313-2302-2023-27-4-928-939
4. Тетюев Л.И. Философия языка Юргена Хабермаса: рациональная теория универсальной прагматики. *Известия Саратовского ун-та. Серия Философия. Психология. Педагогика*. 2019;2(19):171–175. DOI: 10.18500/1819-7671-2019-19-2-171-175
5. Чекушкина Е.Н. Коммуникативная теория Ю. Хабермаса и культура информационного общества. *Теория и практика общественного развития*. 2014;1:25–27.
6. Волков Д.Б. Бостонский зомби: Д. Деннет и его теория сознания. М.: ЛЕНАНД; 2019. 320 с.
7. Деннет Д. Виды психики: на пути к пониманию сознания. М.: Идея-Пресс; 2004. 184 с.
8. Сёрль Дж. Рациональность в действии. М.: Прогресс-Традиция; 2004. 336 с.
9. Чалмерс Д. Сознательный ум: В поисках фундаментальной теории. М.: УРСС; 2013. 512 с.
10. Пенроуз Р. Новый ум: О компьютерах, мышлении и законах физики. М.: УРСС; 2022. 416 с.
11. Владимиров П.А. Значение критики «Нового психофизиологического закона» А.И. Введенского в русской философии. *Кантовский сборник*. 2017;1(36):52–65. DOI: 10.5922/0207-6918-2017-1-5
12. Рубинштейн С.Л. Бытие и сознание. СПб.: Питер; 2012. 288 с.
13. Белов В.Н. Философские теории В.Э. Сеземана и С.Л. Рубинштейна. *Личность. Культура. Общество*. 2023;1-2(25):61–75. DOI: 10.30936/1606_951X_2023_25_1_2_61_75
14. Дубровский Д.И. Проблема идеального. Субъективная реальность. М.: Канон+; 2002. 368 с.
15. Дубровский Д.И. Проблема «сознание и мозг»: информационный подход. *Знание. Понимание. Умение*. 2013;4:92–98.
16. Иванов Д.В. Виртуализация общества. Версия 2.0. СПб.: «Петербургское Востоковедение»; 2002. 224 с.
17. Васильев В.В. Сознание и вещи. Очерк феноменалистической онтологии. М.: Книжный дом «ЛИБРОКОМ»; 2014. 240 с.
18. Кузнецов А.В. Локальный интеракционизм. Новые решения проблемы ментальной каузальности. *Вопросы философии*. 2016;1:148–161.
19. Albus J., Meystel A. Behavior generation in intelligent systems. Gaithersburg (USA); 1997. 219 p.

REFERENCES

1. Wiener N. Cybernetics: or control and communication in the animal and the machine. NY: M.I.T. Press; 1985. 212 p.
2. Neiman J. von. Computer and brain. Moscow: AST; 2023. 192 p. (In Russ.).
3. Tsvyk V.A., Tsvyk I.V., Pavlova T.P. The problematic area of philosophical discourses on the application of artificial intelligence systems in society. *RUDN Journal of Philosophy*. 2023;4(27):928–939. DOI: 10.22363/2313-2302-2023-27-4-928-939
4. Tetyuyev L.I. Jurgen Habermas's philosophy of language: A rational theory of universal pragmatics. *Izv. Saratov Univ. (N. S.), Ser. Philosophy. Psychology*. 2019;2(19):171–175. (In Russ.). DOI: 10.18500/1819-7671-2019-19-2-171-175
5. Chekushkina E.N. The communicative theory of Yu. Habermas and the culture of the Information Society. *Teoriya i praktika obshhestvennogo razvitiya = Theory and practice of social development*. 2014;1:25–27. (In Russ.).
6. Volkov D.B. The Boston zombie: D. Dennett and his theory of consciousness. M.: LENAND; 2019. 320 p. (In Russ.).
7. Dennett D. Types of the psyche: On the way to understanding consciousness. M.: Idea Press; 2004. 184 p. (In Russ.).

8. Serle J. Rationality in action. M.: Progress-Tradition; 2004. 336 p. (In Russ.).
9. Chalmers D. The conscious mind: In search of a fundamental theory. Moscow: URSS; 2013. 512 p. (In Russ.).
10. Penrose R. The new mind: About computers, thinking and the laws of physics. M.: URSS; 2022. 416 p. (In Russ.).
11. Vladimirov P. A. The significance of criticism of the “New psychophysiological law” by A. I. Vvedensky in Russian philosophy. *Kantovskij sbornik = The Kantian collection*. 2017;1(36):52–65. (In Russ.). DOI: 10.5922/0207–6918–2017–1–5
12. Rubinstein S. L. Being and consciousness. St. Petersburg: Peter; 2012. 288 p. (In Russ.).
13. Belov V. N. Philosophical theories of V. E. Seseman and S. L. Rubinstein. *Lichnost` . Kul`tura. Obshchestvo = Personality. Culture. Society*. 2023;1–2(25):61–75. (In Russ.). DOI: 10.30936/1606_951X_2023_25_1_2_61_75
14. Dubrovsky D. I. The problem of the ideal. Subjective reality. M.: Canon+; 2002. 368 p. (In Russ.).
15. Dubrovsky D. I. The problem of “consciousness and the brain”: An information approach. *Znanie. Ponimanie. Umenie = Knowledge. Understanding. Ability*. 2013;4:92–98. (In Russ.).
16. Ivanov D. V. Virtualization of society. Version 2.0. St. Petersburg: “Peterburgskoye Vostokovedenie”; 2002. 224 p. (In Russ.).
17. Vasiliev V. V. Consciousness and things. An essay on phenomenalist ontology. Moscow: Knizny Dom “LIBROCOM”; 2014. 240 p. (In Russ.).
18. Kuznetsov A. V. Local interactionism. New solutions to the problem of mental causality. *Voprosy` filosofii = Questions of philosophy*. 2016;1:148–161. (In Russ.).
19. Albus J., Meystel A. Behavior Generation in intelligent systems. Gaithersburg (USA); 1997. 219 p.

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Владимир Николаевич Белов — доктор философских наук, профессор, заведующий кафедрой онтологии и теории познания факультета гуманитарных и социальных наук, Российский университет дружбы народов, Москва, Россия

Vladimir N. Belov — Dr. Sci. (Philos.), Prof., Head of the Department of Ontology and Theory of Cognition, Faculty of Humanities and Social Sciences, Peoples’ Friendship University of Russia, Moscow, Russia
<https://orcid.org/0000-0002-7741-4653>

Автор для корреспонденции / Corresponding author:
belov_vn@pfur.ru

Сергей Александрович Малофейкин — аспирант кафедры онтологии и теории познания факультета гуманитарных и социальных наук, Российский университет дружбы народов, Москва, Россия

Sergey A. Malofeikin — Postgraduate Student at the Department of Ontology and Theory of Cognition, Faculty of Humanities and Social Sciences, Peoples’ Friendship University of Russia, Moscow, Russia
pfur_aspirant_mca@mail.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.
Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 11.09.2024; принята к публикации 11.10.2024.
Авторы прочитали и одобрили окончательный вариант рукописи.
The article was received on 11.09.2024; accepted for publication on 11.10.2024.
The authors read and approved the final version of the manuscript.