

**СОЗНАНИЕ, ТЕЛО И КУЛЬТУРА
СКВОЗЬ ПРИЗМУ ЭНАКТИВИЗМА
В КОНТЕКСТЕ РАЗВИТИЯ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

А. Е. Уланова

*Московский государственный институт международных отношений
(университет) МИД России*

Поступила в редакцию 1 октября 2019 г.

Аннотация: *в статье рассмотрены ограничения, возникающие в процессе создания искусственного интеллекта, проанализированы основные положения энактивизма, а также сходства и различия его аутопоэтического, сенсомоторного и радикального направлений. Предложен подход, позволяющий использовать представление о процессуальности сознания и о связи социокультурного контекста с появлением когнитивных функций, чтобы сформулировать условия возникновения «сильного» искусственного интеллекта.*

Ключевые слова: *энактивизм, искусственный интеллект, психофизическая проблема, трудная проблема сознания, познание.*

Abstract: *the article deals with the limitations of artificial intelligence development. The author analyses the main features of enactivism, similarities and differences of its autopoietic, sensorimotor and radical currents. As a result, an approach to a strong artificial intelligence based on the process of consciousness and on the connection between sociocultural context and cognitive functions is suggested.*

Key words: *enactivism, artificial intelligence, mind-body problem, hard problem of consciousness, cognition.*

Происходящее в настоящее время развитие технологий машинного обучения приводит к возобновлению существующих и появлению новых дискуссий о возможности создания «сильного» искусственного интеллекта (ИИ). Данный термин был введен Дж. Серлом, который подразумевал под «сильным» ИИ определенным образом запрограммированный компьютер, способный на «понимание и осуществление других мыслительных процессов» [1, р. 417]. И хотя существующие сегодня формы искусственного интеллекта можно охарактеризовать как «слабые», т. е. решающие задачи, в том числе творческого характера, в одной или нескольких смежных областях, прогресс в данной области заставляет задуматься о вероятности появления ИИ, обладающего сопоставимыми или превосходящими человека когнитивными способностями. Так, в книге шведского философа Н. Бострома «Искусственный интеллект: этапы, угрозы, стратегии» автор описывает несколько возможных путей возникновения сверхразума, в том числе стремительное усложнение компьюте-

ров, создание полноценной модели головного мозга, а также совершенствование сетей, связывающих людей и программы (например, Интернета), которое приведет к появлению коллективного сверхразума [2].

Но возможно ли в принципе создание «сильного» искусственного интеллекта? Многие исследователи и философы склонны отвечать на данный вопрос положительно, однако у такого подхода есть и противники. Одним из самых известных является представитель феноменологии Х. Дрейфус. Он вывел четыре допущения (биологическое, психологическое, эпистемологическое и онтологическое), на которых основывается позиция сторонников появления искусственного разума. Биологическое допущение заключается в том, что мозг и компьютер используют в работе схожие механизмы. Психологическое допущение, вытекающее из биологического, предполагает, что разум, как и машина, обрабатывает дискретные данные в виде символов, опираясь на формальные правила. Эпистемологическое допущение гласит, что все знания и действия человека потенциально поддаются формализации и последующему воспроизведению искусственным разумом. Наконец, онтологическое допущение основано на предположении, что всё в мире, «являющееся существенным для разумного поведения, может быть представлено в терминах множества четко определенных независимых элементов» [3, с. 166]. Американский философ утверждал, что данное правило красной нитью проходит через всю западную философскую традицию, начиная с Платона и заканчивая Л. Витгенштейном, который в «Логико-философском трактате» пришел к мысли описать мир как множество атомарных фактов в форме логически независимых предложений. Идея, что «мы живем в мире, в котором гарантирована ясность, определенность и управляемость» [там же, с. 174], легла в основу естественных наук, стала толчком для научно-технического прогресса и вывела человечество на принципиально новый уровень качества жизни. По мнению Х. Дрейфуса, именно онтологическое допущение делает появление «сильного» ИИ принципиально возможным, ведь при развитии имеющихся технологий искусственный разум должен будет оперировать четко структурированной моделью мира.

Данный подход лежит в основе как классических алгоритмов, так и искусственных нейронных сетей, и результаты его применения впечатляющи: современный ИИ обыгрывает человека в го, создает музыкальные композиции и ставит медицинские диагнозы. Он достигает вершин при узкой специализации, но бессилён, когда речь идет о проявлении универсальных когнитивных функций. В чем же причина этих неудач? Дрейфус полагал, что виной всему «телесность» разума и «контекстуальность» поведения, то, что сближает нас с животными. В частности, он приводил в пример способность распознавать образы. И, хотя американский философ ошибался в прогнозах возможностей искусственного интеллекта (так, благодаря развитию технологий компьютерного зрения уровень распознавания образов программами настолько высок, что позволяет развивать индустрию беспилотного транспорта), представляет

ся необходимым, несколько переосмыслив, развить его предположение о «телесном» разуме и «контекстуальном» поведении в рамках исследования ИИ. Для этого следует обратиться к идеям энактивизма.

Основные принципы энактивизма

Х. Дрейфус представляет феноменологическое направление, поэтому его взгляды во многом основаны на работах М. Мерло-Понти. Своим идейным вдохновителем считают французского философа и сторонники энактивизма – направления в когнитивных науках и философии, которое появилось в 90-е гг. XX в. Если Х. Дрейфус писал о «телесности» разума, т. е., фактически, о стирании границы между сознанием и телом, об отрицании необходимости их противопоставления, то энактивисты идут дальше и предлагают снять дихотомию не только сознания и тела, но и субъекта и объекта, организма и окружающей среды и др., развивая концепцию «воплощенного разума». Первой работой, относящейся к энактивизму, принято считать монографию «Воплощенный разум: когнитивная наука и человеческий опыт», написанную Ф. Варелой, Э. Томпсоном и Э. Рош. В ней опровергается классическое представление о познании как отражении внешнего мира в субъективных репрезентациях. «Воплощенность» разума заключается в том, что «когнитивные процессы зависят от того вида опыта, который происходит в результате обладания телом с различными сенсорно-моторными навыками», а они «сами являются встроенными в более обширный биологический, психологический и культурный контекст» [4, p. 172].

Одним из наиболее ярких примеров, которые авторы приводят для подтверждения своей позиции, является восприятие цвета. Если в рамках традиционного подхода предполагается, что цвет – свойство какого-либо объекта, на который смотрит человек или другое существо, обладающее зрительной системой, то с позиции энактивизма цвет – это переменная, зависящая от времени суток, освещения, окружающей среды, работы зрительного нерва или головного мозга.

По мнению Е. Н. Князевой, в основе энактивизма лежат следующие понятия: «холизм» (целое больше суммы частей, но и в части может быть свернуто целое), «динамизм» (отсутствие дискретности, явления описываются не как структуры, а как процессы), «процессуализм» (схожий с философией процесса А. Н. Уайтхеда, согласно которой «действительный мир есть процесс», а «процесс есть становление актуальных сущностей» [5, p. 33], «эмерджентизм» (появление более сложных свойств у целого по сравнению с его частями), «аутопоэзис», т. е. самовоспроизведение и самоподдержание сложных организаций, действенность человека (мир не статичен, всё обретает смысл в движении) и натурализм («жизнь, которая тождественна познанию, – это не отражение мира, а извлечение смысла») [6]. Энактивизм отрицает репрезентационизм и предлагает рассматривать сознание, действующее с помощью познания, и тело как целостность, конструирующую мир вокруг себя, в полном соответствии с идеями Мерло-Понти, который писал: «Мое тело является <...> нуле-

вой точкой отсчета системы координат, с помощью которой я организую окружающие меня объекты как находящиеся справа и слева, спереди и сзади, ниже и выше» [7, с. 130].

В энактивизме выделяют три направления: аутопоэтический, сенсомоторный и радикальный. В основе аутопоэтического энактивизма лежит идея самоподдержания разумной системы, позволяющая ей отделять себя от окружающей среды и создавать между собой и внешним миром область взаимодействия. Простейшая подобная структура – бактерия, которая определяет свою организацию «с помощью деятельности в качестве системы производства собственных компонентов» [8, р. 79]. Сторонники данного подхода полагают, что жизнь и разум неразрывно связаны, а сенсомоторные взаимодействия между организмом и средой носят телеологический характер, благодаря которому возникают когнитивные способности и формируется процесс, который энактивисты иногда называют «выработкой смысла» [9, р. 73].

Исследователи, работающие в рамках сенсомоторного энактивизма, изучают восприятие, в особенности зрительную чувствительность, и пытаются ответить на вопрос, как при ограниченности нашего зрительного опыта (наличие слепых пятен, невозможность видеть объекты одновременно со всех сторон и т. д.) мы способны создавать вокруг себя законченную, целостную и непрерывную реальность. Последователи этого энактивистского направления полагают, что восприятие – динамический процесс, в ходе которого человек накапливает опыт контроля своего тела и взаимодействия с окружающей средой, благодаря чему он получает навык «доставивания» и прогнозирования свойств внешнего мира. Более того, сам мир служит своего рода хранилищем информации, к которому люди при необходимости обращаются. Сенсомоторные энактивисты разделяют позицию приверженцев аутопоэтического энактивизма, понимая восприятие как деятельность, сочетающую прием информации, и влияние на нее, но остаются нейтральными в своих оценках телеологической природы отношений человека и среды [10].

Главная цель радикального энактивизма – улучшение и укрепление позиций антирепрезентационизма. Критикуя сенсомоторный энактивизм, сторонники радикального подхода полагают, что работа по конструированию окружающей действительности возможна и в отсутствие предшествующего опыта и понимания среды [11]. Основное отличие от аутопоэтического энактивизма – разделение когнитивных функций, возникающих в процессе взаимодействия с внешней средой, на базовые (которые основаны на адаптивной сенсомоторике и присутствуют даже у простейших форм жизни) и ментальные способности высокого порядка (которые имеют социокультурный фундамент и проявляются только у человека). Иными словами, радикальный энактивизм выходит за рамки сенсомоторного и аутопоэтического направлений, допуская существование ментальных состояний только «в качестве продуктов языка и культуры в целом» [12] в сочетании с базовыми когнитивными возможностями.

Эмпирическое обоснование «воплощенного разума»

Представляется, что изучение сознания и искусственного интеллекта необходимо осуществлять и подкреплять как с позиции философии, так и со стороны естественных наук. Энактивизм исходно является междисциплинарным направлением, а его основатель Ф. Варела – нейробиологом, поэтому для подтверждения идеи «воплощенного разума» стоит обратиться к исследованиям человеческого мозга.

Особенного внимания заслуживает одна из недавно опубликованных работ, авторы которой предлагают установленный опытным путем «ответ на многовековые открытые дебаты в когнитивной науке и философии о связи между познанием, движениями человеческого тела и окружающей средой» [13, p. 14 769]. Эксперимент проводился с использованием электроэнцефалографии и виртуальной реальности (virtual reality – VR). В VR-пространстве были созданы две комнаты, между которыми появлялась дверь разной ширины: узкая, средняя и широкая. Участникам в VR-шлемах требовалось дожидаться изменения цвета двери (зеленый означал команду Go (идти), красный – NoGo (не идти) и либо пройти через дверь (Go), даже если она была недостаточно широкой для прохода (20 см), либо остаться на месте (NoGo). В случае перехода во вторую комнату участник должен был прикоснуться к специальному кругу, который сообщал ему о зарплате 0,1 евро в дополнение к базовой оплате 10 евро/час. Попытка завершилась заполнением опросника, характеризующего эмоциональное состояние, после чего виртуальная реальность перезагружалась и начинался новый раунд эксперимента (всего 240 попыток на каждого участника).

В результате эксперимента были получены данные (электроэнцефалограммы, поведенческие характеристики и субъективная оценка участников), анализ которых показал, что на восприятие окружающей среды, выраженное в виде изменения потенциалов головного мозга, оказывают сильное влияние как возможности, предоставляемые этой средой, так и действия, совершаемые человеком в рамках данных возможностей. Как отмечают исследователи, перемещаясь в пространстве, люди постоянно прогнозируют, предсказывают состояние мира вокруг себя, опираясь на перцепционные механизмы, зависящие, в свою очередь, от изменения положения тела и пространства вокруг него. Результаты эксперимента, по мнению авторов, являются убедительным доказательством необходимости применения «общей идеи энактивизма как холистического подхода для исследования сознания» [ibid., p. 14 777].

Другим интересным примером является реальный клинический случай, описанный нейропсихологом О. Саксом в книге «Человек, который принял жену за шляпу». Знаменитый певец П., преподающий музыку в консерватории, перестал узнавать людей в лицо, при этом начал видеть лица в предметах, а предметы – в лицах, в остальном оставаясь абсолютно здоровым человеком с превосходным музыкальным слухом. Профессор видел и узнавал детали предмета или изображения, но не

мог распознать картину в целостности, а также принимал желаемое за действительное. В частности, после осмотра у О. Сакса он искал шляпу и увидел ее в своей жене: «Он протянул руку, схватил свою жену за голову и... попытался приподнять ее, чтобы надеть на себя. Этот человек у меня на глазах принял жену за шляпу!» [14, с. 24] В ходе дальнейшего обследования выяснилось, что П. безошибочно определял геометрические формы и справлялся с распознаванием карикатур, но лица людей, даже родных и друзей, он узнавал только при наличии у последних ярких черт (больших зубов или необычной прически). Более того, профессор не мог идентифицировать предметы, при этом точно описывал их по частям. К примеру, находящуюся в руках розу он описал так: «Изогнутая красная форма с зеленым линейным придатком. <...> Тут нет простых симметрий, как у правильных многогранников, хотя, возможно, симметрия этого объекта – более высокого уровня... Это может быть растением или цветком» [там же, с. 27]. Как отметил О. Сакс, в результате зрительной агнозии у П. исчезла способность видеть мир в его эмоциональных, личных, конкретных воплощениях, и он действовал подобно вычислительной машине, выделяя основные детали и схемы, но не воспринимая их как цельные реальные объекты. Если рассмотреть данный пример с позиции энактивизма, то окажется, что вследствие нарушения восприятия П. частично потерял навык конструирования внешнего мира, у него нарушились познавательные способности, а сознание приняло искаженную форму.

В 1995 г. Д. Чалмерс сформулировал «трудную проблему сознания», согласно которой в нейронауках и аналитической философии отсутствует объяснение того, как и почему на основании работы физической системы возникает субъективный опыт [15]. Хотя фактически «трудная проблема сознания» представляет собой современную модификацию давно сложившейся психофизической проблемы, после нахождения удачной формулировки австралийским философом произошел всплеск интереса к вопросам взаимосвязи сознания и тела, в том числе по причине наращивания объемов информации и появления больших данных, вызвавших скачок в развитии технологий искусственного интеллекта, особенно машинного обучения в общем и искусственных нейронных сетей (ИНС) в частности. Повсеместное применение ИНС, характерное для настоящего времени, вынуждает ученых, разработчиков программного обеспечения и философов активно участвовать в дискуссиях, связанных с этическими и аксиологическими аспектами ИИ, а также задаваться вопросом о возможности появления «сильного» искусственного интеллекта.

Представляется, что в рамках классического подхода ответить на вопрос о возникновении сверхразума получится только при решении «трудной проблемы сознания», однако оно вряд ли будет найдено в ближайшее время. Если же обратиться к энактивизму, то «трудную проблему сознания» можно обойти, потому что в таком случае снимается бинарная оппозиция «сознание – тело». Сознание предстает как процесс, возникающий в цепи «центр познания – тело – окружающая среда».

У человека и животных в качестве центра познания выступает головной мозг, у искусственного интеллекта – процессор. Тело с органами чувств в случае ИИ заменяется сенсорными датчиками и различными манипуляторами, а окружающая среда остается неизменной. Итак, можно ли при подобных исходных положениях определить, кто обладает сознанием: человек, животное или машина?

Стоит предположить, что у современного искусственного интеллекта сознание отсутствует, так как его центр познания подчиняется формальным правилам, описанным с помощью двоичного кода, а значит, не способен «понять» непрерывную, постоянно меняющуюся в зависимости от обстоятельств реальность. Именно поэтому искусственные нейронные сети, обученные на специальных выборках, могут, к примеру, писать тексты или создавать стихотворения, однако они оказываются бессвязны и, что самое главное, бессмысленны.

Что касается живых существ, то с точки зрения энактивизма все они обладают когнитивными способностями. Опираясь на постулаты радикального энактивизма, представляется необходимым уточнить, что всё живое владеет базовыми свойствами сознания, но о его высших формах стоит говорить только тогда, когда цепь «центр познания – тело – окружающая среда» превращается в связку «центр познания – тело – социокультурная среда». Действительно, самыми умными и близкими к людям животными считаются те, чья жизнь так или иначе напоминает социокультурное устройство человечества, будь то пчелиный рой, стадо слонов или общающаяся с помощью особенного языка группа дельфинов. Человек же, исходя из вышеизложенного, является единственным, у кого в полной мере можно найти сознание, причем уровень этого сознания тем выше, чем выше социокультурное развитие личности, связанное с «необходимостью бесперебойной трансляции культуры и невозможно гарантировать автоматизм этой процедуры» [16, с. 151].

Таким образом, выработанные энактивистами положения позволяют предположить, что «сильный» искусственный интеллект возникнет только тогда, когда его «центр познания» будет работать на принципах динамики, а не дискретности, усовершенствуются сенсомоторные навыки, а также появятся социокультурные характеристики (в частности, ИИ овладеет языком). А до тех пор сознание в своем высшем проявлении остается неотъемлемой частью исключительно человеческой сущности.

Литература

1. Searle J. Minds, brains, and programs / J. Searle // Behavioral and brain sciences. – 1980. – Vol. 3. – № 3. – P. 417–457.
2. Бостром Н. Искусственный интеллект : этапы, угрозы, стратегии / Н. Бостром ; пер. с англ. С. Филина. – М. : Манн, Иванов и Фербер, 2016. – 496 с.
3. Дрейфус Х. Чего не могут вычислительные машины. Критика искусственного разума / Х. Дрейфус ; пер. с англ. Н. Родмана ; под ред. Б. В. Бирюков. – М. : Прогресс, 1978. – 334 с.

4. *Maturana H. R.* Autopoiesis and cognition : the realization of the living / H. R. Maturana, F. J. Varela. – D. Reidel Publishing Company, 1980. – 171 p.

5. *Whitehead A. N.* Process and reality / A. N. Whitehead. – N. Y. : Macmillan company, 1967. – 546 p.

6. *Князева Е. Н.* Энактивизм : концептуальный поворот в эпистемиологии / Е. Н. Князева // Вопросы философии. – 2013. – № 10. – С. 91—104.

7. *Мерло-Понти М.* Феноменология восприятия / М. Мерло-Понти ; пер. с фр. под ред. И. С. Вдовиной, С. Л. Фокина. – СПб. : Ювента ; Наука, 1999. – 603 с.

8. *Varela F.* The Embodied Mind : Cognitive Science and Human Experience / F. Varela, J. E. Thompson, E. Rosch. – Cambridge : MIT Press, 1991. – 328 p.

9. *Di Paolo E. A.* The enactive approach / E. A. Di Paolo, E. Thompson // The routledge handbook of embodied cognition. – N. Y., 2014. – P. 68–78.

10. *Ward D.* Introduction : the varieties of enactivism / D. Ward, D. Silverman, M. Villalobos // Topoi. – 2017. – Vol. 36. – № 3. – P. 365–375.

11. *Hutto D. D.* Radicalizing enactivism : basic minds without content / D. D. Hutto, E. Myin. – Cambridge : MIT Press, 2012. – 232 p.

12. *Иванов Д. В.* Радикальный энактивизм и проблема субъективности / Д. В. Иванов // Вопросы философии. – 2016. – № 11. – С. 60–69.

13. Sensorimotor brain dynamics reflect architectural affordances / Z. Djebbara, L. B. Fich, L. Petrini, K. Gramann // Proceedings of the National Academy of Sciences. – 2019. – Vol. 16. – № 29. – P. 14 769–14 778.

14. *Сакс О.* Человек, который принял жену за шляпу, и другие истории из врачебной практики / О. Сакс ; пер. с англ. Г. Хасина, Ю. Численко. – М. : АСТ, 2015. – 352 с.

15. *Chalmers D. J.* Facing up to the problem of consciousness / D. J. Chalmers // Journal of consciousness studies. – 1995. – № 2 (3). – P. 200–219.

16. *Силантьева М. В.* Человек : мифы и реальность / М. В. Силантьева // Вестник МГИМО – Университета. – 2013. – № 3 (30). – С. 150–151.

Московский государственный институт международных отношений (Университет) МИД России (МГИМО)

Уланова А. Е., аспирант кафедры философии

E-mail: aleksa.ulanova@yandex.ru

Moscow State Institute of International Relations (MGIMO University)

Ulanova A. E., Post-graduate of the Philosophy Department

E-mail: aleksa.ulanova@yandex.ru