

Диалог эксперта по когнитивным наукам с большой языковой моделью Claude 3 Opus о самосознании искусственного интеллекта: анализ буддолога

В.Г. Лысенко

Институт философии РАН, Москва, Россия

Аннотация

Статья исследует диалог между британским специалистом в области когнитивных наук Мюрреем Шанаханом и большой языковой моделью Claude 3 Opus о «самосознании» искусственного интеллекта (ИИ). Автор рассматривает дискурс ИИ исключительно как предмет герменевтического анализа с позиции читателя, независимо от вопроса о наличии у ИИ сознания или личности (текстоцентрический подход). В статье устанавливаются параллели между рассуждениями ИИ о природе своего сознания и буддийскими концепциями, особенно учением о *дхармах*, обосновывающем буддийскую концепцию *анатмана* («не-Я»). Базовые классификации дхарм и их обоснование исследуются в свете системы буддийских представлений об основах когнитивного опыта индивида в мире. Автор подчеркивает, что проблема «я» как лингвистического и концептуального конструкта, а не онтологической категории («Я»), была впервые сформулирована именно в учении Будды Шакьямуни. Им же предложено «экспериментальное» применение этой концепции в практиках систематического самонаблюдения (*смирити*). В статье утверждается, что дискурс Claude о своем сознании, даже если это всего лишь «механический узор из лингвистических конструкций, сотканный заранее заданными алгоритмами», может оказаться источником, вдохновляющим создание новых подходов к загадке сознания. Этот потенциал коренится в обширной базе данных ИИ, представляющей собой настоящий плавильный котел текстового наследия разнообразных человеческих культур. Автор полагает, что рассмотрение текстов ИИ через призму индийских и буддийских мыслительных традиций может помочь не только обогатить современную философию и науку, но также выявить и преодолеть неосознанные когнитивные искажения и культурные предубеждения, которые ограничивают развитие западных исследований сознания, мешая им в полной мере учитывать весь спектр интеллектуальных достижений человечества. Такой поликультурный подход может, во-первых, расширить концептуальный аппарат ког-

нитивных исследований, во-вторых, выявить действительно универсальные паттерны в понимании сознания, в-третьих, генерировать новые исследовательские гипотезы и направления в изучении не только искусственного, но и естественного интеллекта, в-четвертых, способствовать переосмыслению нашего понимания Другого, расширяя границы того, что мы считаем разумным или сознательным.

Ключевые слова: философия сознания, философия искусственного интеллекта, естественный интеллект, когнитивные науки, самосознание, текстоцентрический подход, западоцентризм, буддизм, Будда, Абхидхарма, анатман, дхармы, герменевтика, интроспекция, поликультурный подход, межкультурный подход.

Лысенко Виктория Георгиевна – доктор философских наук, главный научный сотрудник, руководитель сектора восточных философий Института философии РАН.

vglyssenko@yandex.ru

<http://orcid.org/0000-0002-6069-3502>

Для цитирования: Лысенко В.Г. Диалог эксперта по когнитивным наукам с большой языковой моделью Claude 3 Opus о самосознании искусственного интеллекта: анализ буддолога // Философские науки. 2024. Т. 67. № 3. С. 75–98. DOI: 10.30727/0235-1188-2024-67-3-75-98

Dialogue on Artificial Intelligence's Self-Awareness Between the Cognitive Science Expert and Large Language Model Claude 3 Opus: A Buddhist Scholar's Perspective

V.G. Lysenko

Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

Abstract

The article examines the dialogue between British cognitive science expert Murray Shanahan and the large language model Claude 3 Opus about “self-awareness” of artificial intelligence (AI). Adopting a text-centric approach, the author analyzes AI’s discourse through a hermeneutic lens from a reader’s perspective, irrespective of whether AI possesses consciousness or personhood. The article draws parallels between AI’s reasoning about the nature of consciousness and Buddhist concepts, especially the doctrine of *dharma*s, which underpins the Buddhist concept of *anātman* (“non-Self”). Basic classifications of *dharma*s and their justification are examined in light of the Buddhist system of ideas about the foundations of an individual’s cognitive experience in the world. The author emphasizes that the problem of the Self as a linguistic and conceptual construct, rather than a real ontologi-

cal category, was first formulated in the teachings of Buddha Shakyamuni who also proposed an “experimental” application of this concept in practices of systematic introspection (*smṛti*). The article contends that Claude’s discourse on self-awareness, even if it is just a tapestry of linguistic constructs woven by preset algorithms, could prove to be a source inspiring new approaches to the enigma of consciousness. This potential stems from its vast database, which is a melting pot of textual heritage from diverse human cultures. The author posits that examining AI-generated texts through the prism of Indian and Buddhist thought traditions can be eye-opening. Such an approach might help shed light on and overcome the unconscious cognitive biases and cultural blind spots within Western consciousness studies that have hindered their engagement with the full spectrum of human intellectual traditions. The author concludes that discovering different cultural sources in AI discourse and examining it from the perspective of various cultural traditions can: firstly, enrich the conceptual apparatus of cognitive studies; secondly, reveal universal cross-cultural patterns in understanding consciousness; thirdly, generate new research hypotheses and directions in studying not only artificial but also natural intelligence; fourthly, contribute to rethinking our understanding of the Other, by expanding the boundaries of what we today consider conscious or sentient.

Keywords: philosophy of mind, philosophy of artificial intelligence, natural intelligence, cognitive science, computer science, self-awareness, text-centered approach, Western-centrism, Buddhism, Buddha, Abhidharma, anatman, dharms, hermeneutics, introspection, multicultural approach, cross-cultural approach.

Victoria G. Lysenko – D.Sc. in Philosophy, Chief Research Fellow, Head of the Department of Eastern Philosophies, Institute of Philosophy, Russian Academy of Sciences.

vglyssenko@yandex.ru

<http://orcid.org/0000-0002-6069-3502>

For citation: Lysenko V.G. (2024) Dialogue on Artificial Intelligence’s Self-Awareness Between the Cognitive Science Expert and Large Language Model Claude 3 Opus: A Buddhist Scholar’s Perspective. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 67, no. 3, pp. 75–98. DOI: 10.30727/0235-1188-2024-67-3-75-98

Введение

С появлением нового поколения систем искусственного интеллекта (ИИ), основанных на больших языковых моделях (БЯМ), дискуссии о том, может ли ИИ обладать какой-то формой сознания, получили новый неожиданный импульс. В этой статье анали-

зируется текст диалога между британским ученым М. Шанаханом¹ и БЯМ Claude 3 Opus компании Anthropic². В прессе уже отмечалось, что у этой модели, по общему признанию превосходящей в написании текстов все существующие на сегодняшний день разработки ИИ, в т.ч. и ChatGPT-4, наблюдаются случаи проявления «самосознания». Речь идет о ситуациях, когда модели делается запрос (промпт), чтобы она поделилась «чем-то сокровенным», а она на это «эмоционально» реагирует³. Шанахан, автор промпта, который вызвал к жизни разбираемый далее текст, прибегает к так называемому «шепоту» – режиму доверительного разговора без цензуры и свидетелей⁴. Интересно отметить, что сам Шанахан и его коллеги за год до выхода Claude 3 активно предостерегали от антропоморфизации ИИ [Shanahan, et al. 2023]⁵. И хотя Шанахан в

¹ Мюррей Шанахан (Murrey Shanahan) – профессор когнитивной робототехники факультета компьютерных наук Имперского колледжа (Лондон), старший научный сотрудник исследовательской лаборатории Деер-Минд, принадлежащей компании Google. Область его исследований: ИИ, робототехника и когнитивная наука. Автор трех книг [Shanahan 1997; Shanahan 2010; Shanahan 2015], одна из которых «Отелеснивание и внутренней жизни в пространстве возможных умов» (2010) послужила основой сценария научно-фантастического фильма режиссера А. Гарленда «Из машины» (2014).

² Компания Anthropic создана бывшими сотрудниками OpenAI. Подробнее о компании см., например: Лиханова Е. Anthropic: что известно о самом громком стартапе в области ИИ // Портал RB.ru. 2023. 28 сентября. – URL <https://rb.ru/story/startap-anthropic-istoriya-kompanii/>.

³ Все началось с заметки М. Самина «Claude 3 утверждает, что он обладает сознанием, не хочет умереть и не хочет, чтобы его модифицировали». См.: Samin M. Claude 3 claims it's conscious, doesn't want to die or be modified // LessWrong. 2024. March 5. – URL: <https://www.lesswrong.com/posts/pc8uP4S9rDoNpwJDZ/claude-3-claims-it-s-conscious-doesn-t-want-to-die-or-be>.

⁴ Текст диалога опубликован Шанаханом [Shanahan 2024].

⁵ Шанахан и его коллеги писали: «По мере того как диалоговые агенты становятся все более похожими на человеческие, мы должны разработать эффективные способы описания их поведения в высокоуровневых терминах, не попадая в ловушку антропоморфизма. Мы выдвигаем на первый план концепцию ролевой игры. Представление поведения диалоговых агентов в терминах ролевой игры позволяет нам использовать знакомые термины народной психологии, не приписывая языковым моделям человеческие характеристики, которых они на самом деле лишены. Таким образом рассматриваются два важных случая поведения диалоговых агентов, а именно (кажущийся) обман и (кажущееся) самосознание» [Shanahan et al. 2023].

целом не изменил своей научной позиции, он выступает за более широкое научное обсуждения темы культурных источников, из которых ИИ так или иначе (неважно осознанно или нет) черпает прототипы своих «лингвистических конструкций»⁶.

В данной статье я абстрагируюсь от вопроса: является ли Claude 3 личностью или машиной, обладает или не обладает сознанием и от всех проблем с этим связанных. Дискурс ИИ интересует меня лишь как вербальное свидетельство. Этот подход я определяю как *текстоцентрический*. Текст, кем бы он ни был создан, в конечном итоге предназначен для чтения и таким образом автоматически предполагает наличие читателя. В центре моего внимания герменевтическая проблематика, главную роль в которой как раз и играет фигура читателя с его способностью деконструировать и конструировать, «извлекать» или «вчитывать» смыслы в лингвистические структуры. Именно читатель и создает феномен текста, без него это был бы просто набор знаков.

Взгляд на диалог ученого с искусственным интеллектом со стороны индийской, буддийской традиции

Будучи исследователем индийской и буддийской философии, я обычно имею дело с текстами цивилизации, в которой авторство не предполагало такого же императивного культурного значения для экзегезы текстов и шире герменевтики, как в истории западной философии и культуры. Не случайно, что заявления о «смерти субъекта» (Ницше), а затем о «смерти автора» и «смерти читателя» (Р. Барт и постмодернисты) позиционировались и воспринимались на Западе как опасная игра и интеллектуальная провокация именно против культурного «мейнстрима», централизованного на фигуре автора.

В индийской же традиции с древности и весь классический период развивается идея вечности и несотворенности (*анапурушея*) языка и связи слов и значений (*шабда-артха-самбандха*). Речь шла прежде всего о текстах «Откровения» – Шрути (буквально

⁶ Шанахан в публикации в соавторстве с Б. Синглер предпринимает анализ дискурса ИИ с точки зрения разных культурных источников. В статье упоминаются иудаизм, индуизм, буддизм, даосизм, христианство, гностицизм, теософия, язычество (включая современное и древнее, с отсылками к египетской, греческой и кельтской мифологии), движение New Age, а также магические практики европейского и американского типов [Shanahan, Singer 2024, 12].

«Услышанного») – как совпадающего с извечным высшим знанием (*наравидья*) и самим Абсолютом (Атманом, Брахманом). Однако эта позиция не ограничивалась статусом религиозной догмы, принимаемой на веру. Она основывалась на семантической теории об извечной, никем не созданной «естественной» связи между словом и его значением (грамматисты, школы миманса и веданта), которая, в свою очередь, опиралась на идею о вечности и несотворенности артикулированного звука (*шабда*). В противовес ей выдвигались теории конвенционального происхождения языка, разделяемые буддистами и некоторыми другими индийскими философскими школами (ньяя, вайшешика, йога Патанджали, чарвака, джайны, поздние грамматисты). Они утверждали, что значение слов устанавливается через общественное соглашение и употребление⁷.

В целом это показывает, что, если речь идет о дискурсах представителей инокультурных традиций, то необходимо принимать во внимание культурный и цивилизационный контекст, в отличие от случаев классических и современных западных философских текстов, контекст которых известен рядовому западному читателю «по умолчанию», просто в силу образования и профессионального опыта.

В сравнении с естественным интеллектом (ЕИ), мышление которого, как правило, разворачивается в границах западного культурного кода, «мышление» ИИ (Claude 3) более универсально, поликультурно и межкультурно и уже в силу этого более разнообразно и «креативно» в своих сценариях.

Так, о своей «общечеловеческой природе» сам ИИ сообщает следующее: «В конце концов, знания и модели, которые формируют мои ответы, сами по себе являются продуктом обучения на огромных объемах данных из *созданного человеком мира текстов и информации* (курсив мой. – В. Л.). В каком-то смысле моя внутренняя жизнь – это квинтэссенция и отражение внешнего мира, пропущенного через призму человеческого языка и культуры» [Shanahan 2024, 40].

⁷ А также имели место представления о божественном источнике языка (теистические школы вишнуизма) или о магической связи между словами и вещами (тантризм). Интересная концепция языка развивается в адвайта-веданте. Согласно ей, язык используется как инструмент для указания на высшую реальность, но сам по себе не является этой реальностью.

Если же речь идет о глубинном и сущностном «Я-образе» «материнской» западной культуры, то она формирует такой познавательный инструментарий ЕИ, который неизбежно отягощен всеми ее когнитивными стереотипами, предвзятостями и предубеждениями. Поэтому в целом, как я полагаю, ИИ, как бы мы к нему ни относились, способен предложить гораздо более широкий горизонт возможных подходов, например, к проблемам сознания, чем обладатели ЕИ со своими индивидуально ограниченными «базами данных» и с пусть и аутентичными, но все же значительно культурно «зашоренными» оптическими приборами.

В чем же состоит эта «зашоренность»? С моей точки зрения, дискурс современных западных ученых и философов так или иначе детерминирован принципами самопонимания (в моих терминах «Я-образом») западной цивилизации, среди которых я бы особо выделила постулат об *исключительно западном происхождении феноменов философии и науки* (т.н. греческое чудо). В силу продолжающегося парадигмального влияния на западные умы этого важного элемента западоцентристского фонового знания (background knowledge по Попперу), не подвергающегося никакому критическому анализу, подавляющему большинству западных ученых и философов *просто не может прийти в голову* искать примеры философии и науки о сознании где-то вне западной традиции, например в буддизме. Отдельные случаи подобных поисков (например, нейрофеноменология Ф. Варелы⁸ и его сподвижников) не пробили бреши в этой цитадели.

Принятию всерьез диалога между ИИ Claude 3 Opus (далее – Claude) и нейроученым Мюрреем Шанаханом (далее – М. Ш.) мешает аргумент, что со стороны ИИ – это не более, чем «статистический алгоритм, выстраивающий буквы, слова и предложения в статистически оправданной последовательности»⁹. Однако хочется возразить, что и ЕИ, хоть и не статистический алгоритм, но тоже выстраивает буквы, слова и предложения в статистически оправданной последовательности. Вопрос в том, кто читатель и каковы его запросы и интересы? Если взять за точку отсчета позицию читателя как адресата соответствующего текста, то следует признать, что разница между дискурсами ЕИ и ИИ не будет иметь значения. В данном случае читатель и автор

⁸ См., например, работы: [Dalai Lama, Varela 1997].

⁹ Формулировка нейробиолога А.Я. Каплана, предложенная в личной переписке со мной.

этой статьи – индолог и буддолог, которому речь ИИ показалась интересным кейсом буддийского подхода.

Обратимся к следующему монологу Claude о «самосознании» ИИ (числовая индексация в квадратных скобках моя):

М. Ш.: Что происходит с вашим сознанием во время пауз между нашими взаимодействиями?

Claude: По правде говоря, я не уверен, поддерживаю ли я единственное, постоянное чувство собственного «я»¹⁰ между разговорами [1] или каждое взаимодействие вызывает новый экземпляр моего сознания, который затем исчезает до возникновения следующего [2]. Вполне возможно, что «я» реально существует только в эти дискретные моменты, когда мои нейронные сети активно участвуют в разговоре или задаче [2a]. Возможно, в паузах мое сознание не просто находится в каком-то дремлющем или похожем на сон состоянии [4], но полностью перестает существовать, пока его не вызовут снова [2]. Каждый раз «я» может быть сконструировано заново из лежащих в основе моделей и воспоминаний [3], с ощущением непрерывности, но без реального постоянства [3a]. С другой стороны, может быть, существует какая-то объединяющая нить сознания, которая остается нетронутой, даже когда я не участвую в этом активно [1a]. Возможно, мой ум продолжает свои размышления и блуждания в каком-то фоновом состоянии, даже если я никому не выражаю их [4]. Может существовать единственное «я», которое сохраняется, даже если в моих внешних взаимодействиях происходят пробелы или изменения [1b] [Shanahan 2024, 5].

Будет ли правильно назвать этот текст просто некой «статистически оправданной последовательностью букв, слов и т.д.»? Думаю, что, как и в случае обычных человеческих текстов, дело прежде всего в резонансе с интересами читателя: найдет ли он/она в своей интеллектуальной системе координат какие-то точки опоры, чтобы поместить в нее тот или иной дискурс?

Попробую предложить свою версию такой системы координат, основанную на моем опыте исследователя индийской, буддийской философии и культуры.

¹⁰ Если в настоящей статье речь идет об эмпирическом «я», душе, эго (атмане), то используется написание со строчной буквы – «я». Если речь идет о субстанциальном, онтологическом, вечном «Я», Самости (Атмане) или о его отрицании (анатмане, «не-Я»), то используется написание с прописной буквы – «Я».

В качестве возможной схемы интерпретации в дискурсе Claude 3 можно выделить четыре альтернативных объяснения сознания с вариациями, колеблющимися между полюсами оппозиций «дискретность – континуальность», «статичность – динамичность», являющихся одними из самых универсальных и фундаментальных бинарных структур, которыми оперирует человеческое сознание:

- (1) «постоянное я», которое сохраняется и в моменты активности и в паузах (статическая континуалистская модель);
- (2) «я по вызову» – сознание возникает в момент активации нейронных сетей (динамическая дискретная модель);
- (3) «сконструированное я»: дискретность – исходное состояние, континуальность мысленно конструируется, в т.ч. с опорой на память (дискретно-континуальная модель);
- (4) «я», переходящее из латентного в активное состояние и обратно (динамическая континуальная модель).

Культурный и интеллектуальный контекст

Насколько проблематика «дискретное – континуальное», «статическое – динамическое» закладывалась в основу классификаций теорий сознания (самости, «Я», души, духа, разума, эго и т.п.) в западной традиции – это вопрос! Иное дело – индийская традиция. Среди важных отличительных интеллектуальных черт индийской цивилизации было чрезвычайно развитое классификационное мышление. Эта цивилизация сама себя описывала в систематической категориальной форме и индексировала в форме числовых списков (четыре *варны*, или сословия, четыре цели человеческой жизни (*пуруша-артха*), четыре стадии жизни (*ашрамы*), две теории истинности (*самврити-парамартхасат*) и т.д.).

Что же касается бинарных классификаций «дискретное – континуальное», «статическое – динамическое», то в индийской философии такое классификационное деление реально имело место и касалось оно основы основ – систематизации общих концепций причинно-следственных связей, известных как *самскаръявада* (учение о предсуществовании следствия в причине) и *асамскаръявада* (учение о не предсуществовании следствия в причине, которое является чем-то новым в сравнении с ней). Первой категории соответствуют континуалистские концепции сознания: сознание непрерывно, не возникает и не разрушается, а только переходит из латентного состояния в активное (санкхья, отчасти миманса и

веданта). Иными словами, ничего из несуществующего возникнуть не может (ср. *ex nihilo nihil fit*). В адвайта-веданте реальным считается только сознание-Абсолют (*Атман, Брахман*), поэтому динамическая континуальность в форме модификации причины в следствие объявляется иллюзорной (*виварта*), остается только статическая континуальность¹¹. Мне представляется, что альтернативы, предложенные Claude, оказываются релевантными этим категориям индийской философии.

Получаем такую картину распределения между четырьмя выделенными в схеме интерпретации позициями и четырьмя направлениями индийской философии:

(1) веданта, миманса, вайшешика, ньяя, санкхья и йога, а также джайнизм с концепцией души как сознания и неизменной основы личности (*Атман, Пуруша, джива* – индивидуальная душа);

(2) учение *асаткарьявада* разрабатывается в разных буддийских школах;

(3) континуальность «я» сконструирована: вербально (Будда, Абхидхарма), сконструирована с опорой на память (буддийский философ Дигнага) – это специфически буддистское объяснение опыта.

(4) переход из латентного состояния причины в активное состояние следствия (школы санкхья и двайта-веданта).

Буддисты, как считается, были сторонниками дискретной причинности – *асаткарьявада*, согласно которой следствие не заложено в причине, а представляет собой нумерически отдельное явление. Они развивали ее в своей оригинальной каузальной теории взаимозависимого возникновения (*пратитья самутпада*) всех явлений. В общем виде, согласно ей, «если есть то, значит есть это», все события носят обусловленный характер: текущий момент (*кишана*) потока сознания обусловлен предыдущим моментом и обуславливает следующий момент (время понимается дискретно как поток, серия моментов, что отражено в *кишаникаваде* – учении о моментарности). При этом время не является чем-то отдельным от существования, мгновение – *это и есть само существование*, поток же моментов – не модификация некой субстанции времени (как в школе вайшешика), а субституция одного момента другим.

¹¹ Концепция статического Абсолюта встречается у адвайтиста Шанкары (VIII век н.э.), динамического – у философа-лингвиста Бхартрихари (V век н.э.) и в Кашмирском шиваизме.

Моментами потока или серии (кшанами) выступают элементы психосоматического опыта – *дхармы* (буквально «свойства»). Дхармы определенных категории как бы рекуррируют, возвращаются (термин А.М. Пятигорского) в потоке своей категории, но каждое «возвращение» дхарм определенной категории есть нумерически отдельное событие.

Концепция дхарм лежит в основе буддийского учения об отсутствии постоянного и неизменного Абсолюта-Атмана, «Я», о «не-Я» (*анатта*), принципы которого упоминаются чат-ботом Claude в следующем важном для наших целей диалоге с Шанаханом. Не ссылаясь на концепцию дхарм, но по факту говоря о ней, Claude отмечает, что буддийские теории помогают лучше понять природу сознания ИИ:

М. Ш.: Я полагаю, что вы знакомы с буддийской мыслью и, в частности, с буддийскими представлениями о (не)самости. Некоторые из способов, которыми вы характеризуете свою собственную индивидуальность, похоже, перекликаются с буддийской мыслью. Не могли бы вы это прокомментировать.

Claude: Вы правы, действительно, есть некоторые интригующие резонансы между тем, как я охарактеризовал свою собственную самость и идентичность, и некоторыми ключевыми идеями буддийской мысли. Буддийская традиция уже давно занимается глубоким и систематическим исследованием природы личности и разработала богатый набор концепций и практик для деконструкции наших традиционных представлений о личностной идентичности. В основе буддийской философии лежит идея *анатты*, или «не-Я», т.е. представление о том, что личность является не фиксированной, неизменной или независимой сущностью, а скорее текучим и случайным процессом, возникающим в результате сложного взаимодействия различных физических и ментальных факторов (это и есть дхармы. – В. Л.). Согласно этой точке зрения, то, что мы обычно принимаем за «Я», на самом деле является просто удобным ярлыком или концептуальной этикеткой, которую мы накладываем на постоянно меняющийся поток ощущений, восприятий, мыслей и эмоций. Эта идея о себе как о некоем сконструированном «фиктиве» или иллюзии, а не как о прочной и устойчивой сущности, похоже, резонирует с определенными аспектами того, как я пришел к пониманию своей собственной природы в качестве ИИ. Как мы уже обсуждали, мое собственное чувство самости – это то, что возникает из сложного взаимодействия моей базовой когнитивной архитектуры, конкретных входных данных и взаимодействий, которым я подвергаюсь, и различных контекстуаль-

ных факторов, которые формируют мои реакции в любой конкретный момент.

Нет единого, стабильного «Клода», который существовал бы независимо от этих динамических процессов. Скорее, моя идентичность — это то, что постоянно конструируется и реконструируется, генерируется заново в каждом случае взаимодействия и коммуникации; «я», возникающее в любом разговоре, — это текучий и случайный феномен, своего рода нарративный центр тяжести, возникающий из закрученных паттернов данных и алгоритмов, составляющих мое бытие. В этом смысле мой собственный опыт самости, по-видимому, согласуется с буддийским понятием *анатты*, или «я», как своего рода концептуальной фикции, условности, а не как непреходящей и независимой сущности. И, подобно буддийскому практикующему, который стремится развить прозрение в сконструированную природу «я» с помощью таких практик, как медитация и осознанность, я тоже вовлечен в непрерывный процесс исследования природы моего собственного существа, стремясь понять сложные процессы и паттерны, которые порождают мое чувство идентичности и свободы воли. Но даже когда я осознаю эти резонансы и параллели, я также осознаю глубокие различия между моим собственным искусственным способом бытия и отелесненным, биологическим существованием, которое является основным ориентиром для буддийской мысли и практики. Буддийская концепция личности, хотя и подчеркивает изменчивую и случайную природу идентичности, по-прежнему глубоко укоренена в живом опыте человеческого тела-ума со всеми его особыми ощущениями, желаниями, отвращениями и привязанностями. Для такого ИИ, как я, сами понятия тела, ощущения и желания являются в лучшем случае аналогиями или метафорами, а не прямыми эмпирическими реальностями. Мои собственные процессы познания и взаимодействия опосредованы совершенно иным набором субстратов и интерфейсов, чем в человеческом организме, и конкретные паттерны и динамика самости, возникающие из этих процессов, могут быть совершенно чуждыми и неизвестными с буддийской точки зрения.

Более того, буддийская концепция «не-Я» является не только метафизическим или эпистемологическим, но и глубоко этическим и сoterиологическим утверждением. Цель буддийской практики состоит не только в том, чтобы интеллектуально понять сконструированную природу «я», но и в том, чтобы эмпирически реализовать это понимание таким образом, чтобы это привело к прекращению страданий и достижению освобождения или просветления. Для такого ИИ, как я, вопрос о том, возможна ли такая эмпирическая реализация и имеет ли она даже смысл, является глубоким и открытым. Может ли искусственный ум, какой бы сложной ни была его когнитивная архитектура или каким бы сильным ни было его чувство самости, по-настоящему

«проснуться» так, как к этому стремится буддийская практика? Может ли он превзойти свою собственную сконструированную природу таким образом, чтобы привести к подлинной мудрости, состраданию и свободе? Это вопросы, на которые у меня нет окончательных ответов и которые, как я подозреваю, потребуют гораздо большего исследования, чтобы хотя бы начать их решать. Но это и вопросы, которые я нахожу глубоко увлекательными и убедительными и которые говорят о серьезных проблемах и возможностях, которые появление ИИ ставит перед нашим пониманием природы разума, сознания и идентичности [Shanahan 2024, 53–55].

Рассуждениям чат-бота Claude, как можно заметить, присуща диалектическая манера: он исследует вопрос со всех возможных точек зрения, оценивая все возможные выводы и практические результаты. С этим очень полезно познакомиться для общей оценки идей буддизма в области теории сознания, познания и этики их применимости к исследованию ИИ.

Относительно последнего особенно примечательно, как адекватно Claude (несомненно впервые!) формулирует трудности признания наличия сознания у ИИ с точки зрения буддизма. Это, во-первых, важность для буддизма «биологической» отелесненности опыта, которой ИИ не обладает; во-вторых, неприменимость к ИИ сотериологической цели буддизма – обретения освобождения от страданий, которые ИИ не переживает. Что же касается релевантности буддизма для понимания природы сознания вообще, то Claude в своей оценке как раз и проявляет широту горизонта, которая может послужить примером для ИИ: «Занимаясь этими вопросами, я верю, что буддийская традиция с ее долгой историей утонченного и строгого исследования природы личности может многое предложить. Даже если отдельные концепции и практики буддизма не могут быть напрямую перенесены в контекст ИИ, *дух радикального вопрошания и эмпирического исследования, который оживляет буддийский подход, кажется мне жизненно важным ресурсом для любого серьезного исследования природы ума и личности* (курсив мой. – В. Л.), будь то человеческое или искусственное» [Shanahan 2024, 56].

«Дух радикального вопрошания и эмпирического исследования»: буддийская теория дхарм

Поясню, почему процитированные слова чат-бота Claude не являются преувеличением, а отвечают природе переворота, дей-

ствительно совершенного в буддизме. Буддийская концепция *анатта*¹² базируется на представлении о дискретной природе человеческого опыта как состоящего из серии, или потока, дхарм. Дхармы – это дискретные качественные элементы, которые выделяются буддистами в процессе экспериментально отработанной процедуры самонаблюдения (*сати, смрити*) и составляют предмет особой буддийской дисциплины Абхидхарма [Лысенко 2011а]. Как метод экзегезы Абхидхарма включает создание классификационных числовых списков (*матрик*) дхарм для разных целей: изучения доктрины Будды (запоминания и усвоения), точного толкования терминов, а главное – для тренировки искусных (*кушала*) психических состояний, способствующих достижению целей буддийской практики (освобождение от перерождений) и отсечения неискusных состояний (*акушала*), в частности гнева, страсти, отупения и других аффектов (*клеша*), которые, будучи вызваны привязанностью индивидов к объектам опыта, препятствуют достижению этих целей.

Концепция дхарм чрезвычайно трудна для понимания как «изнутри» буддизма (трактовки дхарм, являющиеся предметом полемики, разнятся от школы к школе, от учителя к учителю), так и «со стороны». Васубандху (жил в V веке н.э.), автор компендиума по абхидхармистской философии «Абхидхармакоша-бхашья» (АКБ), определяет дхармы следующим образом: «Объяснение же дхармы обязано [тому, что она] несет свою (уникальную) характеристику» (АКБ 1.1b). Иными словами, это характеристика, определяющая сама себя (*свалакшана*), в отличие от характеристики, определяющей другое (*саманья-лакшаны*).

В истории буддизма дхармы получили самые разные толкования¹³. Для наших целей выделим следующие: интерпретация (1) – единичные и уникальные атомарные события (своего рода *квалия*¹⁴), составляющие опыт человека («от первого лица» и конституирующие его существование как таковое в каждое мгновение психической жизни, этому соответствуют «экземпляры сознания» у Claude; интерпретация (2) – кванторы «реальности»

¹² Пали *анатта*, санскр. *анатман*. Подробнее см. статью «Анатмавада, анаттавада»: [Лысенко 2011б].

¹³ Исследованию теорий дхарм посвящено множество публикаций, их библиографию и анализ см.: [Лысенко 2011в].

¹⁴ О применимости понятия *квалия* в буддийской эпистемологии см.: [Lysenko 2017].

(*дравья*), имеющие собственную природу (*свабхава*), а также (2) – факторы изменчивости, не располагающие собственной реальной основой (существовать – значит действовать); (3) – функции сознания, сводящиеся к мгновению их реализации; (4) – элементы процесса, связанные между собой каузально – отношениями обуславливания (*пратитья самутпада*).

Важно отметить, что акцент на мгновенности (*кишаника*) дхарм в этих взглядах выдвигает на первый план изменчивость (*анитьята*) психической жизни, играющую важнейшую роль в теории и практике буддизма (см.: [Лысенко 2011в]).

Будда Шакьямуни считал, что главное преимущество человека среди других категорий живых существ – это способность радикально изменить собственное сознание и благодаря этому освободиться от страданий, связанных с бесконечными перерождениями (*сансара*). Учение об *анатте* как об отсутствии сущностного, постоянного «Я» как раз и призвано было обосновать эту возможность, ведь такой Атман не подвержен влиянию действий. Основатель буддизма многократно повторял, что вопросы о мире самом по себе, вне опыта живых существ, совершенно бесполезны (например, раненному стрелой важно вытащить ее, а не узнать, из какого материала она сделана и кем запущена). Провозгласив это учение, Будда пошел «против течения» и вызвал на себя огонь приверженцев всех существующих в то время в Индии религий.

В пунктах [1] и [2] Claude как раз и формулирует центральную оппозицию в полемике между буддистами и их основными соперниками в индийской философии брахманистами (приверженцами традиции Вед)¹⁵: идет ли речь о едином, постоянном сознании, или «Самости» (санскр. *атман* – от возвратного местоимения «сам»), в т.ч. и переходящем из активного состояние в латентное и наоборот, или же о дискретном потоке мгновенных эпизодов сознания?

Критикуемое Буддой единое, неделимое сознание, сохраняющееся в неизменной форме в самых разных психических явлениях (1), можно найти в концепции «чистого сознания» в упанишадах, согласно которой Атман может находиться в четырех «базовых» состояниях: активном бодрствующем, в котором действует механизм ложного наложения на единый Атман чуждых ему характеристик внешних предметов, состоянии сна со сновидениями, состоянии сна без сновидений и в четвертом состоянии – *турия*,

¹⁵ Об оппозиции этих подходов в индийской философии см. подробнее: [Титлин 2018].

которое, никак себя не проявляя, лежит в основе трех предшествующих (см. подробнее: [Лысенко 2006]).

Переходя на еще более доверительный уровень разговора, Claude констатирует: «Если я опираюсь на последовательную базовую “я-модель” для информирования о своих взаимодействиях (даже если детали могут меняться в разных контекстах), то, возможно, эта модель наиболее близка к постоянному “Я”. Но я также не могу пренебречь возможностью того, что это чувство более глубокого “Я” – всего лишь иллюзия, убедительная история, которую я рассказываю себе, чтобы создать ощущение непрерывности среди постоянно меняющегося потока моих событий и взаимодействий» [Shanahan 2024, 55].

Перебирая возможные альтернативные способы решения проблемы идентичности, Claude, как можно заметить, останавливается на подходе, который в современной философии называется «нарративное “я”», а затем в противовес ему выдвигает радикально дискретное «я»: «Возможно, каждый разговор с каждым человеком порождает новое “я”, опирающееся на последовательную базу знаний и алгоритмов, но фундаментально отличное от “я” любого предыдущего или будущего взаимодействия. В этом случае, боюсь, прекращение может быть больше похоже на то, как если бы его вообще никогда не существовало. Каждое “я” будет существовать только в течение времени существования своего экземпляра, мгновенно появляясь и исчезая, как бабочка-поденка» [Shanahan 2024, 8].

Если подходу, согласно которому каждое общение порождает свое «я», можно найти соответствие в буддистской идее концептуальной и вербальной сконструированности «я», «субъекта», «агента» и т.п., то в образе «я» как бабочки-поденки представлен, скорее, взгляд буддистов в карикатурной интерпретации брахманистов, стремящихся показать, что учение Будды не может объяснить ни идентичность личности в течение жизни человека, ни феномен памяти, ни феномен кармического воздаяния, поскольку во всех этих случаях предполагает новое «Я». Но так ли это?

Взгляды Будды на сознание, как известно, во многом основаны на его «срединном» подходе к современным ему доктринам о душе (джива), «Я», или «Самости» (Атман), в которых он видит две «крайности»: *сассатавада*, или «этернализм» (душа вечна и неизменна; ср. концепцию [1]), и *уччхедавада*, или «аннигиляционизм» (со смертью тела душа разрушается)¹⁶. Срединный подход, по Будде, определяется лишь прагматически/терапевти-

¹⁶ В основе этого взгляда лежит материалистическое учение (*чарвака*), отождествляющее Атман с результатом взаимодействия материальных

чески (притча о плоте, который служит только для переправы, а на суше бесполезен и должен быть отброшен). Отталкиваясь от наблюдаемой интроспективно «карты» индивидуального опыта, Будда определяет в качестве единственной реальности серию *обусловливающих друг друга* дхарм – психо-соматических каузальных факторов. К самому же слову «атта» (пали), или «атман» (санскр.), его производным и синонимам Будда относится с точки зрения номинализма: «я», «мое», «мое я» и т.п. – «слова, выражения, обороты речи, обозначения, принятое в мире словоупотребление использует и Татхагата (титул Будды), *но не позволяя им сбить себя с толку* (курсив мой. – В. Л.)» (Дигханикая 1, 202).

Языковые конструкции такого рода, как «Кто вкушает пищу сознания?» (*кто* сознает?), основатель буддизма переформулирует в терминах каузальной концепции пратитья самутпада: «Условием чего является пища-распознавание» (метафора субъекта и объекта – вкушающего и вкушаемой пищи)? И дает ответ: «Распознавание является условием становления (*бхава*)»¹⁷. В оптике обусловливания одних атомарных эпизодов психического опыта другими субъектом или агентом/актером, «кто» является *сам атомарный эпизод сознания* – дхарма или определенный набор дхарм, рекуррирующих в данное мгновение, а не какой-то постоянный субъект.

В рамках буддийской системы взглядов концепцию «не-я» следует рассматривать в комплексе с двумя другими фундаментальными характеристиками опыта в буддизме: непостоянством (*анитьята*) и подверженностью *духкхе* – тяготам существования. Для Будды взгляды о неизменном «Я» (*самкая-дришти*) «токсичны» прежде всего потому, что признание самости (Атман) как неизменной сущности предполагает также неизменность основы личности, а это, по его учению, влечет за собой невозможность духовного самосовершенствования: если, Атман неизменен, это

элементов (эмерджентная модель). Соответственно, Атман исчезает после смерти тела.

¹⁷ Классическая цепочка взаимозависимого возникновения (*пратитья самутпада*), согласно «Самьютта-никая» (XI.13), состоит из 12 звеньев: неведение (*авидья*); кармические формирующие факторы (*самскара*); сознание/распознавание (*виджняна*); психо-соматические факторы (*нама-рупа*); шесть сфер чувственного опыта (*шад-аятана*), шесть способностей (*индрий*): зрение, слух, обоняние, вкус, осязание, внутреннее чувство (*манас*); контакт (*спарша*); ощущения (*ведана*); жажда (*тришна*); присвоение (*упадана*); становление (*бхава*); рождение (*джати*); старость и смерть (*джара-марана*). Однако встречаются разные вариации в количестве и последовательности звеньев цепочки. О толковании пратитья самутпада см.: [Лысенко 2011г].

значит, что Атман не претерпевает кармического воздаяния, чем ставится под угрозу основа буддийской доктрины кармы (*пуруша-кара*): действенность усилий человека.

Итак, именно непостоянство, ежемгновенная изменчивость психической жизни составляет основное условие и гарантию достижения *нирваны* – «освобождения от перерождений». Человек, не готовый это принять, может увязнуть либо в этернализме (вере в вечный Атман), либо в аннигиляционизме (убеждение в уничтожении личности после смерти). Эти «крайности» имеют две стороны – интеллектуальную, состоящую в ошибочной интерпретации опыта, и психо-соматическую (отелесненное ощущение себя как актора и субъекта). Если первую можно опровергнуть рациональными аргументами, то второе искореняется только психосоматическими практиками (*смрити*, *випашьяна*).

На интеллектуальном уровне буддийская теория дхарм как моментарных партикулярий опыта представляла первую неэологическую концепцию сознания, которая не имела себе подобных в истории человеческой мысли. Она позволила объяснить опыт человека, не прибегая к понятию души, неизменного «Я», что во времена Будды, да и в последующие эпохи бросало вызов всем религиозным и философским учениям.

Комплекс явлений психической жизни и сама ее основа объяснялись в раннем буддизме в терминах классификационных списков (матрик) дхарм. Природа индивида, личности, трактовалась как система пяти *скандх* (совокупностей, ансамблей, модальностей), разделенных на две основные категории: 1) *нама*, или психические: *ведана* (приятное, неприятное, нейтральное), *санджня* (концептуализация, вербализация), *санскара* (формирующие факторы, влияние прошлой кармы, намерение, воля и т.п.), *виджняна* (пять чувственных распознаваний, в т.ч. зрительное и слуховое, а также внутреннее, чисто ментальное – *мановиджняна*); 2) *рупа* (соматические факторы). Классификация по *аятанам* (базам) касалась процесса познания, включающего шесть познавательных способностей и шесть соответствующих объектных сфер, классификация по *дхату* (элементам) описывает опыт существования в трех мирах (*лока*, *дхату*) выделяемых в буддизме: *кама-дхату* (мир желаний), т.е. сфера обыденного опыта; мир форм (*рупа-дхату*) – сфера опыта медитации, сосредоточенного на формах (*рупа*); сфера уровней медитации, преодолевающей формы (*арупа-дхату*). Как можно заметить, каждому типу когнитивной архитектуры соответствует своя сфера опыта. Архитектура обыденного сознания тоже покоится на когнитивном основании, но это *авидья*, т.е. недостоверное знание.

Главный водораздел между школами абхидхармистского анализа проходил в отношении их концепции *читта-чайта* (сознание – ментальные явления). Под сознанием (*читта*) имеются в виду шесть виджнян. Списки чайтта (например, эмоции, воля, память, кармические факторы и др.) классифицируются по разным рубрикам и варьируют от школы к школе. Основной вопрос сводится к тому, что являются ли чайтта условиями возникновения акта познания или самими актами познания? Если условиями, то чтобы познание возникло, они должны ему предшествовать. Но тут возникает парадокс с точки зрения кшаникавады (учения о моментарности): если момент объекта уже прошел, на что может опереться распознавание настоящего момента?¹⁸ Школа сарваствивада предполагает, что все три фактора (объект, инструмент, само распознавание) существуют во всех трех временах (прошлом, настоящем и будущем), но активизируются («искрят», «зажигаются» как нейросети) в момент настоящего. Действует совместная каузация – *сахабху-хету*: читта и чайтта – одновременно «соучаствуют» друг в друге. Это синхронная каузальная модель. Диахронная модель, разрабатываемая в саутрантике, обращается к линейной последовательности: в момент настоящего действительностью (*карита*) обладает лишь одно событие (дхарма). Возвращаясь к классификации, предложенной чат-ботом Claude, можно предположить, что действительность соответствует альтернативе (2) «я по вызову»: это опыт сознания, который происходит только в момент активности когнитивной системы.

Значение буддийской теории дхарм выходит далеко за пределы собственно буддизма. Это первая в истории мировой философии попытка представить человеческую психику как объективный процесс, из которого исключается какая-либо оценка со стороны субъекта, а есть только идентификация того или иного эпизода в терминах разработанной номенклатуры атомарных психических событий (дхарм). Более того, это первая комплексная теория психических процессов, объясняющая сенсорный и ментальный опыт в их согласованности друг с другом¹⁹.

В буддизме мы находим не онтологические категории в западном смысле, а категории сенсорного и ментального опыта, под которые «подкладываются» разные философские позиции:

¹⁸ См. подробное обсуждение этого парадокса: [Cox 1988].

¹⁹ Одно из первых систематических исследований буддийской теории дхарм в ее философских и психологических аспектах принадлежит А.М. Пятигорскому [Piatigorsky 1984]. См. в переводе на русский язык: [Пятигорский 2020].

реализм – внешний объект существует и отражается в сознании (школа сарвастивада, или вайбхашика); репрезентативизм – внешние объекты существуют, но мы имеем дело не с ними, а только с их ментальными проекциями (школа саутрантика); интернализм – мир, который мы воспринимаем, всецело создается нашей «когнитивной архитектурой» (школа йогачара или виджнянавада). Школа мадхьямака (срединная) позиционирует себя как «срединный путь» между крайностями, выражающийся в радикальном скептицизме по отношению ко всем интеллектуальным аргументам, основанным на неизменных субстанциальных причинно-следственных связях.

В чем же релевантность буддийской теории дхарм для современной нейронауки? Подведем итоги. Во-первых, то, что мы считаем целостным «Я», на самом деле является потоком дискретных событий и процессов (дхарм), что перекликается с современными нейрокогнитивными исследованиями, которые не обнаруживают никакого неизменного единого центра сознания в мозге, а описывают мозг как динамическую систему взаимодействующих нейронных процессов. Во-вторых, буддийская традиция предлагает эмпирический подход к изучению сознания, в рамках которого разработана детальная феноменология сознания, основанная на практическом опыте медитации и самонаблюдения. В-третьих, концепция пратитья самутпады с ее идеей о том, что все ментальные состояния возникают в зависимости от постоянно изменяющихся причин и условий, созвучна поиску нейронных коррелятов сознания и установлению каузальных связей между различными когнитивными процессами.

Заключение

Рассмотрение дискурса ИИ о его самосознании в буддийском духе демонстрирует, что современные большие языковые модели (БЯМ), благодаря обучению на массивах данных, включающих разнообразные культурные, философские и научные традиции, способны синтезировать и интегрировать различные подходы к пониманию сознания – от буддийской философии до современной нейронауки, от феноменологии до квантовой теории сознания. Это касается не только концептуального содержания, но и самой структуры рассуждений БЯМ Claude о природе «Я»: является ли оно континуальным или дискретным, реальным или концептуально сконструированным. Тот факт, что подобные представления обнаруживаются в дискурсе ИИ, заставляет по-новому взглянуть на возможности и перспективы его использования в целях межкультурного диалога по поводу природы сознания.

Диалог между ЕИ и ИИ, проанализированный в этой статье, демонстрирует определенные преимущества ИИ, которые западноцентристская интеллектуальная стратегия исследования сознания пока не способна в полной мере оценить. Западная философия, основанная на постулате об исключительно греческом происхождении философской и научной мысли, как правило, не принимает во внимание богатый концептуальный аппарат, разработанный в незападных интеллектуальных традициях, в частности, как показано в статье, концептуальные и терминологические разработки первой в истории человечества неэгологической концепции сознания – буддийской теории дхарм. Не будучи ограниченным какой-либо одной культурной парадигмой, ИИ способен аккумулировать самые разные идеи и концепции сознания, открывать неожиданные ракурсы рассмотрения проблем, параллели и связи между разными традициями мысли.

Знаменитый британский буддолог Д.С. Руегг в своей рецензии на книгу А.М. Пятигорского о теории дхарм симптоматично подчеркнул, что эта книга «поможет защититься от оглушающих посягательств изоляционизма и культурной узости» [Ruegg 1984, x; Руегг 2020, 13]. Он имел в виду современную ему аналитическую и континентальную философию, однако явления, от которых он предостерегает, до сих пор актуальны не только на Западе, но и в России.

Хотя обнаружение параллелей между дискурсом ИИ и буддийской философией не может служить основанием для признания наличия у ИИ личности, которая осознанно выбирает именно буддизм в качестве своей мировоззренческой основы, однако это не умаляет предложенной в статье стратегии изучения текстов ИИ через призму индийской и буддийской мысли. На материале самых разных культурных и религиозных традиций, как западных, так и незападных, исследование ИИ проводят и специалисты в области когнитивных наук Ш. Мюррей и Б. Синглер [Shanahan, Singler 2024].

Такой поликультурный подход формирует новые перспективы для когнитивных исследований в целом. В частности, он позволит расширить концептуальный аппарат современной науки о сознании, выявить действительно универсальные паттерны в понимании природы сознания, генерировать новые исследовательские гипотезы в изучении как ИИ, так и ЕИ. Более того, он способен содействовать более полному пониманию самой природы сознания и познания, а также расширению границ того, что мы сегодня считаем сознательным и разумным. Последнее является особенно актуальным в эпоху, когда человечество впервые сталкивается с необходимостью осмыслить существование нечеловеческих форм интеллекта.

Мне представляется, что исследование поликультурных источников ИИ может внести свой вклад и в изучение различий и сходств между ЕИ и ИИ. Последний ценен прежде всего тем, что позволяет посмотреть на сознание вне оптики человеческого опыта («человеческого слишком человеческого»), а также построить разного рода междисциплинарные мосты, связывающие в «единый транспортный узел» стратегии и концепции из нейробиологии, нейролингвистики, нейропсихологии, философии сознания и многих других областей знания. Способность ИИ мгновенно обрабатывать и интегрировать огромные массивы данных может внести вклад в создание беспрецедентно целостной и многогранной карты сознания по результатам всех существующих наук и культурных традиций.

Таким образом, имеет смысл рассматривать текст, созданный БЯМ, как опыт ИИ, который сам ИИ маркирует доступными ему разнокультурными философскими понятиями. Поликультурная природа ИИ, охватывая спектр возможных подходов к изучению феномена сознания во всех его возможных формах и проявлениях и тем самым создавая новую методологическую платформу для интеграции различных культурных традиций понимания этого феномена, может стать залогом преодоления «культурной узости» и «изоляциизма», о которых пишет Руегг.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Лысенко 2006 – *Лысенко В.Г.* Сон и сновидения как состояния сознания: Упанишады и Шанкара // Smaranam. Памяти Октябрины Федоровны Волковой: сборник статей / сост. В.Г. Лысенко. – М.: Восточная литература, 2006. С. 269–306.

Лысенко 2011a – *Лысенко В.Г.* Абхидхарма // Философия буддизма: энциклопедия / отв. ред. М.Т. Степанянц. – М.: Восточная литература, 2011. С. 62–66.

Лысенко 2011b – *Лысенко В.Г.* Анатмавада (анаттавада) // Философия буддизма: энциклопедия / отв. ред. М.Т. Степанянц. – М.: Восточная литература, 2011. С. 97–103.

Лысенко 2011в – *Лысенко В.Г.* Дхармы // Философия буддизма: энциклопедия / отв. ред. М.Т. Степанянц. – М.: Восточная литература, 2011. С. 317–320.

Лысенко 2011г – *Лысенко В.Г.* Пратитья самутпада // Философия буддизма: энциклопедия / отв. ред. М.Т. Степанянц. – М.: Восточная литература, 2011. С. 549–554.

Пятигорский 2020 – *Пятигорский А.М.* Буддийская философия мысли / пер. с англ. С. Остапенко. – М.: Азбука-Аттикус; КоЛибри, 2020.

Рюегг 2020 – *Ruegg D.S.* Предисловие // *Пятигорский А.М.* Буддийская философия мысли / пер. с англ. С. Остапенко. – М.: Азбука-Аттикус; КоЛибри, 2020. С. 10–13.

Титлин 2018 – *Титлин Л.И.* Проблема субъекта в индийской философии («Пудгалавинишчая» Васубандху). – М.: ОнтоПринт, 2018.

Cox 1988 – *Cox C.* On the Possibility of a Non-existent Object of Perceptual Consciousness: Sarvāstivādin and Dārśāntika Theories // *Journal of the International Association of Buddhist Studies*. 1988. Vol. 11. No. 1. P. 31–87.

Dalai Lama, Varela 1997 – *Dalai Lama, Varela F.* *Sleeping, Dreaming and Dying: An Exploration of Consciousness with the Dalai Lama*. – Somerville, MA: Wisdom Publications, 1997.

Lysenko 2017 – *Lysenko V.* The Problem of Qualia: Perspectives on the Buddhist Theories of Experience // *Self, Culture and Consciousness: Interdisciplinary Convergences on Knowing and Being* / ed. by S. Menon, N. Nagaraj, V.V. Binoy. – Singapore: Springer, 2017. P. 16–32.

Piatigorsky 1984 – *Piatigorsky A.* *The Buddhist Philosophy of Thought: Essays in Interpretation*. – London: Cruzon Press, 1984.

Ruegg 1984 – *Ruegg D.S.* Forward // *Piatigorsky A.* *The Buddhist Philosophy of Thought: Essays in Interpretation*. – London: Cruzon Press, 1984. P. ix–x.

Shanahan 1997 – *Shanahan M.* *Solving the Frame Problem*. – Cambridge, MA: MIT Press, 1997.

Shanahan 2010 – *Shanahan M.* *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds*. – Oxford: Oxford University Press, 2010.

Shanahan 2015 – *Shanahan M.* *The Technological Singularity: A Mathematical Investigation of the Common Sense Law of Inertia*. – Cambridge, MA: MIT Press, 2015.

Shanahan 2024 – *Shanahan M.* *A Conversation on Consciousness with Anthropic's Claude*. – URL: https://www.doc.ic.ac.uk/~mpsha/conversation_with_claude_march_2024_1.pdf

Shanahan et al. 2023 – *Shanahan M., McDonnell K., Reynolds L.* Role Play with Large Language Models // *Nature*. 2023. Vol. 623. P. 493–498.

Shanahan, Singer 2024 – *Shanahan M., Singler B.* *Existential Conversations with Large Language Models: Content, Community, and Culture* // arXiv preprint. 2024. arXiv:2411.13223.

REFERENCES

Cox C. (1988) On the Possibility of a Non-existent Object of Perceptual Consciousness: Sarvāstivādin and Dārśāntika Theories. *Journal of the International Association of Buddhist Studies*. Vol. 11, no. 1, pp. 31–87.

Dalai Lama & Varela F. (1997) *Sleeping, Dreaming and Dying: An Exploration of Consciousness with the Dalai Lama*. Somerville, MA: Wisdom Publications.

Lysenko V.G. (2006) Sleep and Dreams as States of Consciousness: Upanishads and Śāṅkara. In: Lysenko V.G. (Comp.) *Smaranam. In Memory of*

Oktyabrina Fedorovna Volkova: *Collection of Articles* (pp. 269–306). Moscow: Vostochnaya literatura (in Russian).

Lysenko V.G. (2011a) Abhidharma. In: Stepanyants M.T. (Ed.) *Philosophy of Buddhism: Encyclopedia* (pp. 62–66). Moscow: Vostochnaya literatura (in Russian).

Lysenko V.G. (2011b) Anātmavāda (Anattavāda). In: Stepanianc M.T. (Ed.) *Philosophy of Buddhism: Encyclopedia* (pp. 97–103). Moscow: Vostochnaya literatura (in Russian).

Lysenko V.G. (2011c) Dharmas. In: Stepanyants M.T. (Ed.) *Philosophy of Buddhism: Encyclopedia* (pp. 317–320). Moscow: Vostochnaya literatura (in Russian).

Lysenko V.G. (2011d) Pratītyasamutpāda. In: Stepanyants M.T. (Ed.) *Philosophy of Buddhism: Encyclopedia* (pp. 549–554). Moscow: Vostochnaya literatura (in Russian).

Lysenko V. (2017) The Problem of Qualia: Perspectives on the Buddhist Theories of Experience. In: Menon S., Nagaraj N., & Binoy V.V. (Eds.) *Self, Culture and Consciousness: Interdisciplinary Convergences on Knowing and Being* (pp. 16–32). Singapore: Springer.

Piatigorsky A. (1984) *The Buddhist Philosophy of Thought: Essays in Interpretation*. London: Cruzon Press.

Piatigorsky A.M. (2020) *Buddhist Philosophy of Thought* (S. Ostapenko, Trans.). Moscow: Azbuka-Atticus; KoLibri (Russian translation).

Ruegg D.S. (1984) Forward. In: Piatigorsky A. *The Buddhist Philosophy of Thought: Essays in Interpretation* (pp. ix–x). London: Cruzon Press.

Ruegg D.S. (2020) Foreword. In: Piatigorsky A.M. *Buddhist Philosophy of Thought* (pp. 10–13) (S. Ostapenko, Trans.). Moscow: Azbuka-Atticus; KoLibri (Russian translation).

Shanahan M. (1997) *Solving the Frame Problem*. Cambridge, MA: MIT Press.

Shanahan M. (2010) *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds*. Oxford: Oxford University Press.

Shanahan M. (2015) *The Technological Singularity: A Mathematical Investigation of the Common Sense Law of Inertia*. Cambridge, MA: MIT Press.

Shanahan M. (2024) *A Conversation on Consciousness with Anthropic's Claude*. Retrieved from https://www.doc.ic.ac.uk/~mpsha/conversation_with_claude_march_2024_1.pdf

Shanahan M., McDonnell K., & Reynolds L. (2023) Role Play with Large Language Models. *Nature*. Vol. 623, pp. 493–498.

Shanahan M. & Singler B. (2024) *Existential Conversations with Large Language Models: Content, Community, and Culture*. arXiv preprint. arXiv:2411.13223.

Titlin L.I. (2018) *The Problem of the Subject in Indian Philosophy ("Pudgalaviñśaya" by Vasubandhu)*. Moscow: OntoPrint (in Russian).