

ВСЁ ЛИ УЧИТЫВАЮТ КРИТИКИ ДЖ. СЁРЛА ПРИ ВЫХОДЕ ИЗ 'КИТАЙСКОЙ КОМНАТЫ'?

О. А. Казанцев 

Южный федеральный университет,
Россия, 344006, г. Ростов-на-Дону, ул. Большая Садовая, 105/42

 okazancev@sfedu.ru

Статья представляет критическую реплику на работу С. Б. Куликова *Всё ли учёл Дж. Сёрл при входе в 'Китайскую комнату'*? В исследовании предпринимается попытка проверить основательность не столько выдвинутого автором возражения против сёрловской критики сильного искусственного интеллекта, сколько предложенной С. Б. Куликовым стратегии защиты от возражений подобного рода.

В начале статьи даётся обоснование того, почему ответ от прогресса и появления качественно иных архитектур интеллектуальных систем (например, нейросетей) не получится отвергнуть посредством обращения к биологическому натурализму Дж. Сёрла и его же ответу от моделирования работы мозга. Затем предпринимается попытка соотнести выдвинутое С. Б. Куликовым возражение мысленному эксперименту американского философа с другими критическими ответами. В результате получается разработать собственный способ классификации множества опровержений сёрловской аргументации, а также разместить среди полученных типов и критику с помощью новых архитектур интеллектуальных систем.

На последнем исследовательском этапе предлагаются другие, отличные от используемых С. Б. Куликовым, способы защиты интуиции Дж. Сёрла против сильного искусственного интеллекта. Делается вывод, что без обновления концептуальных оснований аргумента 'Китайской комнаты' изначальная сёрловская гипотеза оказывается в уязвимом положении.

Ключевые слова: 'Китайская комната', мысленные эксперименты, сильный искусственный интеллект, нейросети, большие языковые модели, картотека, 'Китайская картотека'

Для цитирования: Казанцев О. А. Всё ли учитывают критики Дж. Сёрла при выходе из 'Китайской комнаты'? *Омский научный вестник. Серия «Общество. История. Современность»*. 2026. Т. 11, № 3. С. 110–120. <https://doi.org/10.25206/2542-0488-2026-11-3-110-120>. EDN: WIWCII.



© Казанцев О. А., 2026.

Контент доступен под лицензией Creative Commons Attribution 4.0 License.

DO THE CRITICS OF J. SEARLE TAKE EVERYTHING INTO ACCOUNT WHEN LEAVING THE CHINESE ROOM?

О. А. Kazantsev 

Southern Federal University,
Bolshaya Sadovaya St., 105/42, Rostov-on-Don, 344006, Russia

 okazancev@sfedu.ru

This article is a critical response to Sergey B. Kulikov's *Did John Searle Take Everything into Account When Entering the Chinese Room?* It seeks to examine not so much the soundness of the objection to Searle's critique of strong artificial intelligence put forward by Kulikov as the adequacy of Sergey B. Kulikov's proposed strategy of defending against objections of this kind.

First, it is shown why an appeal to progress and the emergence of qualitatively different architectures of intelligent systems, such as neural networks, cannot be rejected by invoking John Searle's biological naturalism and his brain simulation reply. Then an attempt is made to relate Kulikov's objection to the American philosopher's thought experiment to other critical responses. On this basis, a distinctive way

of classifying the variety of refutations of Searle's argument is developed, and critiques that rely on new architectures of intelligent systems are located within the resulting types.

At the final stage, alternative ways are proposed — different from those employed by Kulikov — of defending John Searle's intuitions against strong artificial intelligence. It is concluded that, without a renewal of the conceptual foundations of the Chinese Room argument, Searle's original hypothesis remains in a vulnerable position.

Keywords: Chinese Room argument, thought experiments, strong Artificial Intelligence, neural networks, large language models, filing cabinet, Chinese Filing Cabinet

For citation: Kazantsev O. A. Do the critics of J. Searle take everything into account when leaving the Chinese Room? *Omsk Scientific Bulletin. Series Society. History. Modernity.* 2026;11(3): 110–120. <https://doi.org/10.25206/2542-0488-2026-11-3-110-120>. EDN: WIWCII. (In Russ.).



© Kazantsev O. A., 2026.
The content is available under a Creative Commons Attribution 4.0 License.

Введение

Применительно к возникшей вокруг известного мысленного эксперимента Ф. Фут литературе используется термин 'вагонеткология' ('trolleyology') [1, с. 26]. Само наличие подобной '-логии' подчеркивает, что экспериментирование британского философа с вагонеткой привнесло значимую инновацию в этическую теорию. Появился новый способ проверки старых моральных интуиций [2, р. 89], который вышел за дисциплинарные пределы философской этики и породил множество как академических, так и пародийных вариаций [3]. И, по аналогии с вагонеткологией, допустимо указать на существование 'комнатоведения'¹ — области внутри философии сознания и сопредельных тематических регионов, сформированной вокруг обсуждения, критики и переложения на иные проблемы аргумента Дж. Сёрла 'Китайская комната'².

Законы жанра в 'комнатоведческой' литературе по большей части сводятся к своеобразному 'критическому императиву': *разработай такой ответ на аргументацию Сёрла, который покажет слабость, несерьёзность, ошибочность, порочность, непродуманность и т. д. его идей*. Во многом американский философ-аналитик собственноручно канонизировал именно такой способ рассуждений о 'Китайской комнате'³, сформулировав на материалах обсуждений с анонимными корреспондентами шесть классических ответов против своей гипотезы ('сильный' искусственный интеллект⁴, *понятый прежде всего как особенная компьютерная программа со способностью иметь собственные когнитивные состояния, невозможен*) — ответ от систем, ответ от робота, ответ от моделирования работы мозга, комбинированный ответ, ответ от других сознаний, ответ от нескольких обиталищ [4, с. 381]. Позднее к этой цепочке критики добавятся ответ от виртуального разума (*The Virtual Mind Reply*) [5], рассмотрение рассуждений Дж. Сёрла через призму модальной логики [6], сомнения относительно универсальности некоторых из интуиций мыслителя [7] и др. Что уж говорить о том экспериментальном разнообразии, которое возникает при переносе инсценировки из 'Китайской комнаты' в другие предметные рубрики или модификации в рамках всё тех же дебатов: 'моральная комната' [8], 'кабинет математики' [9], 'комната суммирования' [10], 'пазлы Лема' [11], 'комната случайных операций' [12], 'английская комната' [13], 'демон Сёрла' [14], 'шахматная комната' [15] 'пример Хелен Келлер' [16] и не только.

Приведённые выше возражения и вариации по мотивам аргументации Дж. Сёрла — малая доля от 'комнатоведческого' массива из критики, контр-критики и экстраполяций мысленного эксперимента. В недавней статье С. Б. Куликова была предложена ещё одна возможность для проверки надёжности сёрловского экспериментирования. А именно автор выдвинул предположение о том, что аргументация 'Китайской комнаты' применима лишь к одному, но не ко всем имеющимся проектам по созданию сильного ИИ: «Сёрл доказывает невозможность лишь того варианта [сильного] ИИ, который соответствует его собственным допущениям. Если же пересмотреть онтологические основания и учесть альтернативные архитектуры, в особенности нейроморфные, то могут открыться пути к созданию сильного ИИ, выходящие за границы его критики» [17, с. 79]. Но после рассуждений от противного (от того, что экспертные системы и нейросети, возможно, имеют потенциал для преодоления сёрловской критики) [17, с. 80] С. Б. Куликов возвращается к скепсису американского философа: «неизбежен вывод, что, несмотря на отсутствие развитых форм искусственного интеллекта во времена 'входа' Сёрла в его 'Китайскую комнату', весьма затруднительным было бы говорить о том, что он при построении мысленного эксперимента недоучёл какие-либо скрытые факторы. Очевидный прогресс компьютерных технологий скорее подтверждает критику возможности сильного ИИ» [17, с. 81–82].

Солидаризируясь с автором в том, что подобная апелляция к прогрессу в лице нейросетей не опровергает позиции Сёрла, согласиться с представленным рассуждением затруднительно. Ведь после введения предположения об ограниченности сёрловской дефиниции сильного ИИ, исследователь переходит через небольшое источниковедческое отступление [17, с. 81] к 'двойному обнулению' претензий генеративных моделей на интенциональность. А именно С. Б. Куликов опровергает потенциал нейросетей на возникновение сознания с помощью обращения к критике Сёрлом ответа от моделирования работы мозга, а затем довершает собственные рассуждения отсылкой к биологическому натурализму создателя 'Китайской комнаты' [17, с. 81]. Но в таком виде предложенная аргументация проблематична.

Во-первых, ответ от моделирования работы мозга и его сёрловская критика, к которым обращается С. Б. Куликов, не имеют непосредственного отношения к генеративным нейросетям в частности

и к коннекционизму как подходу в целом⁵. Косвенно это подтверждается самим Сёрлом при полемике с четой Чёрчлендов, авторов коннекционистского ответа (*The Connectionism Reply*) [18, р. 227]: «мои оппоненты Чёрчленды используют одну из разновидностей *ответа [от] систем*» [19]. Напрямую указанную неприменимость можно увидеть при сопоставлении ответа от моделирования работы мозга с основополагающими принципами в функционировании нейросетей: если сёрловские оппоненты из Беркли и MIT предлагали смоделировать в программе мозговую активность «носителя китайского языка, когда он понимает рассказы на китайском языке и даёт ответы на них» [4, с. 387], то проектировщики современного поколения ИИ используют некоторые принципы из работы естественных нейронных сетей для отладки функционирования нейросетей искусственного происхождения [20, с. 414]. Иными словами, программа, созданная для имитации деятельности мозга, и программа, вычислительная архитектура которой *создана на основе реальной деятельности мозга*, не одно и то же.

Во-вторых, С. Б. Куликов для разоблачения претензий генеративных моделей предлагает перейти с эпистемологического уровня (*проявлений познавательных способностей ИИ*) на уровень онтологический (*в чём реализованы познавательные способности ИИ*): «[человеческий] интеллект возникает на базе биологического субстрата, сформированного эволюцией: мозг как материальный носитель интеллектуальных функций обладает внутренней причинностью и способностью к самоорганизации. Компьютерные устройства, напротив, представляют собой артефакты технической деятельности. [...] в компьютерных системах отсутствует онтологическая основа для возникновения независимого интеллекта» [17, с. 81]. *Использование подобного 'онтологического аргумента' в пределе делает сам мысленный эксперимент 'Китайская комната' бессмыслицей.* Ведь если для опровержения потенциала какой-либо архитектуры интеллектуальной системы на роль сильного ИИ было бы достаточно сослаться на то, что вычисления или вычислительная машина не равны биологическому мозгу, такую тривиальную констатацию фактов с трудом можно назвать убедительной аргументацией [21, р. 186]. Определённые теоретические шаги и конкретное мысленное экспериментирование потому-то и необходимы Сёрлу в статье *Сознание, мозг и программы*, чтобы обосновать претензию биологического натурализма на объяснение природы интенциональности, сознания и когнитивных состояний: «[нет] никаких оснований, почему бы мы не могли передать машине способность понимать английский или китайский языки, ибо в некотором важном смысле наши тела и наши мозги суть в точности такие же самые машины. Но, с другой стороны, я вижу очень сильные основания для утверждения, что мы не смогли бы передать такие способности машине в том случае, когда функционирование этой машины определено только в терминах вычислительных процессов. [...] один аспект предлагаемого мной аргумента состоит в том, что лишь нечто такое, что обладает этими каузальными способностями и могло бы обладать интенциональностью. Быть может, другие физические и химические процессы могли бы продуцировать в точности те же эффекты» [4, с. 393].

Ниже предлагается рассмотреть идею С. Б. Куликова о том, что сёрловская аргументация не покрывает всего современного поля по созда-

нию сильного ИИ, без отсылок к критике ответа от моделирования работы мозга и биологическому натурализму, *в чистом виде*. Притом попробуем оценить предложение автора в контексте других 'комнатоведческих' дискуссий. Подобный шаг позволит лучше 'измерить' критический потенциал, заключённый в изначальной гипотезе С. Б. Куликова, а также определить типичность или нетипичность предложенного возражения Сёрлу. Для этого начнём с рассмотрения основных подходов к классификации замечаний, контраргументов и ответов в полемике о 'Китайской комнате'. И, в конечном счёте, постараемся понять: *всё ли учитывают критики Сёрла при выходе из 'Китайской комнаты'?*

Классификационные подходы в дискуссии о 'Китайской комнате'

Помимо основного способа разметки возражений сёрловским интенциям, предложенного автором 'Китайской комнаты', в обширном корпусе из критической и апологетической литературы встречаются и работы иного типа. Исследования, направленные на разработку и обоснование альтернативных способов классифицировать разнородные аргументы в пользу той или иной стороны внутри 'комнатоведческих' баталий. Среди таких альтернатив следует выделить классификационные подходы за авторством Д. Деннета и Д. Хофштадтера, Р. Дампера, а также Д. Коула. Существуют и другие модели распределения возражений против суждений Сёрла, но остановимся на заявленных трёх как наиболее гетерогенных относительно друг друга.

Д. Деннет и Д. Хофштадтер предлагают рассматривать 'Китайскую комнату' (как и всё семейство из аналогичных мысленных экспериментов, призванных «стимулировать искреннее, прочувственное интуитивное озарение») в качестве *насоса интуиции* [22, с. 20]. Подобный ракурс позволяет разложить воображённый Сёрлом опыт на составные блочные элементы (ручки или регуляторы), прокручивая каждый из которых можно получить самые разнообразные и довольно неожиданные экспериментальные вариации [14, с. 334]. Сюда будут попадать и предложенные самим изобретателем 'Китайской комнаты' реплики, вроде интериоризации комнатного содержания внутрь головы оператора-имитатора, и любые модификации мысленного эксперимента, к примеру демон Хогелэнда [14, с. 335–336]. При таком подходе допустимо раскладывать возражения против сёрловских интенций, имеющие модифицированную экспериментальную форму, через соотнесение с коррекцией комнатной процедуры в том или ином элементарном блоке (имитационный материал, степень аккуратности при имитировании, размер имитации, размер оператора, скорость работы оператора) [14, с. 334].

Р. Дампер выдвигает классификационный подход, основанный на *модальной реконструкции* 'Китайской комнаты' [23, р. 169–170]. Используя подход Р. Соренсена к логической структуре деструктивных мысленных экспериментов (направленных на опровержение, а не на подтверждение гипотетического утверждения) [23, р. 171], исследователь выделяет пять элементов в аргументе Сёрла: исходное модальное утверждение; логическое следование; обнаружение странного следствия; выявление абсурдности такого странного следствия; содержательная возможность [23, р. 171–172]. При этом из каждого следует определённый класс опро-

вержений, благодаря чему классификация выстраивается по точке приложения критики.

Д. Коул выводит классификацию иного формата: различие и распределение возражений против позиции, экспериментирования и аргументации Сёрла по степени уступчивости (т. е. согласия / частичного согласия / полного несогласия с ним) [5]. В конечном счёте так можно выделить три группы ответов аргументу 'Китайской комнаты': признание того, что оператор комнаты ничего не понимает, но понимание производится чем-то иным (это, к примеру, ответ от систем); допущение того, что комнатный индивид действительно не понимает, но всё дело в выбранном типе программы и / или 'железа', которые можно изменить, заменить или подвергнуть апгрейду для возникновения понимания (например, ответ от робота); и все иные критические подходы, утверждающие, что либо человек в комнате на самом деле понимает, либо сама конструкция эксперимента сомнительна для целей возникновения понимания, либо иные опровергающие Сёрла варианты [5].

У всех трёх описанных классификационных подходов есть как свои сильные, так и слабые стороны. Например, вариант Д. Деннета и Д. Хофштадтера позволяет не только классифицировать возражения к эксперименту Сёрла, но и придумывать новые. Тем не менее в потенциале представление 'Китайской комнаты' как суммы регуляторов предполагает наличие неудобно большого числа классов, что не позволяет использовать классификацию по её назначению: *составить карту, которая будет компактнее отображаемой территории*. Подход Р. Дампера помогает не только выдвинуть относительно небольшое (по крайней мере меньшее, чем наличная сумма ответов, критических проектов и проч. в самой дискуссии и по сравнению с предыдущим подходом) количество классов, но вместе с тем показать заполненность этих классификационных позиций [23, р. 174]. И всё же основная проблема введённой типологии состоит в том, что аргументация Сёрла изначально составлена на естественном языке. Поэтому её формализация будет обладать той или иной степенью волюнтаризма [23, р. 175]. В конце концов, предложение Д. Коула рассортировать возражения и контраргументы внутри дискуссии о 'Китайской комнате' по градации согласия / несогласия с сёрловскими интенциями помогает разметить стороны спора. Правда, сами аргументационные различия в подобном подходе стираются.

В рамках сказанного все три способа классифицировать аргументы против позиции Сёрла следует признать недостаточно эффективными для целей разметки 'комнатоведческого' ландшафта. По этой причине предложим собственную классификацию, обоснуем её полезность и обозначим место в такой схеме для стартового возражения С. Б. Куликова.

Структурный подход для разметки критических ответов против позиции Дж. Сёрла

Для того, чтобы нивелировать слабые стороны предыдущих подходов и сохранить сильные, предлагаем обратить внимание на различие теоретического аргумента и собственно мысленного эксперимента внутри 'Китайской комнаты'. Подобную демаркацию проводят и Д. Коул [5], и Дж. Морал [18, р. 215]. Под теоретическим аргументом будем прежде всего понимать предпосылки к сёрловскому экспериментированию, такие как: постулирова-

ние внутренней интенциональности, порождаемой мозгом; признание недостаточности синтаксиса для возникновения семантики; описание компьютерных программ как прежде всего синтаксических структур; ментальные или когнитивные состояния как состояния семантические присущи деятельности мозга [18, р. 219–220]. В таком случае под мысленным экспериментом 'Китайская комната' будет обозначаться основной сценарий Сёрла с не знающим китайского языка оператором внутри помещения, книгой правил, бумагой и чернилами, а также различные комнатные вариации [18, р. 216–217]. Такое вычленение двух элементов в рассуждениях американского философа отсутствует в классификационном подходе Д. Деннета и Д. Хофштадтера и смазано внутри классификации Д. Коула. Зато в предложении Р. Дампера оно заметно. Поэтому в собственном классификационном подходе предлагается *до некоторой степени* модификация дамперовской схемы.

Сконструированный подход предполагает выделение трёх укрупнённых типов ответов, возражений и критики в сторону сёрловской позиции касательно 'Китайской комнаты'. А именно предлагается выделить *аргументы входа, аргументы обработки и аргументы выхода*⁶. И соответствующие им три типа контраргументов, направленных на защиту предположений касательно сильного ИИ американского философа. По существу, в сконструированном подходе предлагается экстраполировать критикуемую Сёрлом сумму концепций (функционализм, бихевиоризм, теорию вычислительного разума, символический подход и т. д.) до масштабов всей дискуссии вокруг 'Китайской комнаты'. Имеется в виду, что все критические атаки на собственно сёрловскую позицию можно распределить на либо бьющие по предпосылкам мысленного эксперимента (того, что можно именовать тем самым теоретическим аргументом), либо направленные против интерпретации происходящего внутри 'Китайской комнаты' (практически все модификации комнаты и рассуждения об их внутренних процессах), либо обосновывающие иной исход диалога со скрытым в четырёх стенах оператором (более редкий тип контраргументации, о котором ещё будет сказано ниже). При таком раскладе, например, ответы от систем, от робота, от моделирования работы мозга и комбинированный ответ будут отнесены к аргументам обработки, а ответы от других сознаний и от нескольких обиталищ — к аргументам входа.

Подобную классификацию можно обозначить как *структурную*, ведь она предполагает сортировку по выбранной мишени из структуры аргументации Дж. Сёрла: американский философ был не прав либо в *посылках мысленного эксперимента* (аргументы входа), либо в *самом экспериментальном опыте* (аргументы обработки), либо в *выводах из посылок и проведённого мысленного эксперимента* (аргументы выхода). Разумеется, оппонент сёрловской позиции может совершать критику и предпосылок, и собственно мысленного эксперимента, и выводов одновременно. Но всё же, зачастую, он вынужден делать акцент на чём-то одном хотя бы из определённой *критической экономики*.

Например, иницируя анализ аргумента 'Китайской комнаты' с неточности, ошибочности или спорности той или иной предпосылки (аргумент входа), критик может предложить модификацию или иной взгляд на изначальный вариант экспериментирования (аргумент обработки), а после перейти

к ошибочности или спорности экспериментальных и теоретических предположительных результатов (аргумент выхода). В такой схеме очевидно, что при попытке уже 'адвоката' Сёрла отстоять взгляды своего 'подзащитного' против достаточно проработать *всего лишь* аргумент входа. Показать, что обнаруженная погрешность в сёрловских посылах несущественна, или её попросту нет, а ошибается только оппонент, или при прочих равных остальные посылки вкупе с мысленным экспериментом продолжают давать те же выводы. При удачном исходе такой защиты и логически, и риторически продолжать разбор аргументов обработки или выхода даже не требуется.

Аналогичную реконструкцию критики можно произвести, если противник Сёрла согласен с посылками (избегает аргумента входа), но критикует происходящее внутри мысленных стен 'Китайской комнаты' (использует аргумент обработки) и заодно пробует развенчать экспериментальные итоги (выводит из аргумента обработки аргумент выхода). В такой критической стратегии так же следует остановиться на ключевом для оппонента аргументе обработки, чтобы затем переходить к демонтажу аргумента выхода не потребовалось.

Как видно по двум предложенным стратегиям критики, возможным исходя из структурного подхода, наиболее действенной должна стать ещё не обозначенная критика с применением аргумента выхода. О ней скажем позднее, а пока вернёмся к оппонированию гипотезы Сёрла со стороны С. Б. Куликова. Его предложение скептически рассмотреть посылку американского философа о том, что такое сильный ИИ, подпадает внутри структурного подхода под разряд аргументов входа. Притом ещё раз заметим, что автор предлагает *исключительно и только* аргумент входа, ведь не рассматривает затем изменённую вариацию самого мысленного эксперимента (аргумент обработки) и скорректированные выводы из него (аргумент выхода): «[п]редположим, что аргументация Сёрла логически безупречна [т. е. безупречны аргументы обработки и выхода]. Но и тогда вполне возможно, что она анализирует сильный ИИ в рамках ограниченной онтологической модели [что означает применение аргумента входа]» [17, с. 79]. Классифицировав выдвинутое С. Б. Куликовым возражение, становится понятным, почему предлагаемая контраргументация всё-таки не работает: он пытается критиковать аргумент входа при помощи *контраргумента обработки*. Подобное логически и риторически выглядело бы более основательным, если контраргумент обработки дополнить контраргументом входа, а также модифицировать мысленный эксперимент под критикуемый аргумент входа. Иными словами, следует без отсылки к тривиальным суждениям или к другим входным, предвещающим экспериментирование, тезисам, попробовать ослабить аргумент входа соответствующим контраргументом; *и если того окажется недостаточно, требуется прибавить модифицированную версию эксперимента* 'Китайской комнаты'. Ниже попробуем осуществить предложенный сценарий.

'Китайская комната' между весной и зимой искусственного интеллекта

Как было сказано ранее, С. Б. Куликов предлагает аргумент входа против той дефиниции, которую даёт Сёрл применительно к сильному ИИ. Последняя признаётся исторически и концептуаль-

но ограниченной⁷ [17, с. 79]. Среди подкрепляющих подобный взгляд положений автор ссылается на следующий: «[у]поминаний работы 1943 года [У. Мак-Каллока и У. Питтса], в которой предлагалась исторически первая версия нейросети, у Сёрла не прослеживается. Зато есть ссылка на более позднюю работу Джозефа Вейценбаума, посвящённую описанию программного комплекса для имитации диалога человека и компьютера и тем самым способного генерировать новый контент» [17, с. 81]. Указанное отсутствие ссылочного аппарата по коннекционизму и нейросетевым архитектурам предлагается воспринимать как подкрепление того, что мысленный эксперимент 'Китайская комната' изначально не был рассчитан на искусственные нейросети. Представленное обоснование предлагаемого нивелировать при помощи двух контраргументов. Один из них обозначим как *стилистический*, а другой — как *исторический*.

Несмотря на то, что знание или незнание, учёт или неучёт Сёрлом в статье *Сознание, мозг и программы* не только символического подхода как объекта критики, но и подхода коннекционистского требует отдельного исследования, притом скорее биографического характера, всё же возможно снизить степень у основательности тезиса С. Б. Куликова. Например, с помощью стилистического контраргумента отсутствие ссылочной базы в сёрловском тексте можно отнести к манере письма у почившего философа. Последнюю отмечает в одной из своих публикаций российский исследователь творчества Сёрла М. Н. Белянин⁸: «[п]римечательно, что в *Речевых актах* отсутствует список литературы. Тем самым автор подчеркивает новизну предмета (что небесспорно), и новизну подхода (что несомненно). В одной из поздних книг Сёрл мимоходом отметит, что вся философская классика так и написана — с минимумом или вовсе без библиографических ссылок. Всегда стремясь следовать этому правилу, Сёрл отступит от него единственный раз — в полемически построенной работе *Открывая сознание заново*» [24]. Сотворённый собственными усилиями мыслителя образ *философа-простеца* не должен вводить в заблуждение при непосредственном анализе сёрловских текстов и аргументации.

Другой контрдовод, а именно исторический, переводит взгляд с особенностей стилистики Сёрла на приметы его времени. В последние десятилетия прошлого столетия конкурентное соотношение между символическим и коннекционистским подходами, изначально пребывавшее в состоянии относительного равновесия (по финансированию, признанию в академическом сообществе и сообществе разработчиков, и т. д.), изменилось. *Символически* прекращение равновесного состояния в пользу преобладания символического подхода и наступления «коннекционистской зимы» связывается в литературе по истории ИИ с выходом книги М. Минского и С. Пейперта *Перцептроны* [20, с. 483–484]. По каким бы причинам ни пришла коннекционистская зима, она пришлась как раз на период с 1973 по 1980 годы [25]. Поэтому ретроспективно отсутствие ссылок в статье Сёрла можно объяснить ещё и тем, что аргументация философа направлена против символического подхода как наиболее сильного противника⁹. Противника, который успел к моменту публикации *Сознание, мозг и программы* вывести из игры другого претендента на создание сильного ИИ, т. е. коннекционизм. Поэтому если сёрловский

аргумент справляется с символизмом, то и проигравший коннекционистский подход подобным способом можно одолеть. По крайней мере, такой могла быть *риторическая логика* Сёрла.

Конечно, и стилистический, и исторический контраргументы полностью не снимают поставленную проблему (учитывает ли аргумент 'Китайской комнаты' коннекционизм или нет). Тем не менее, как представляется, выдвинутые доводы всё же позволяют устранить предложенное С. Б. Куликовым обоснование. Но что же делать с ответами на позицию Сёрла, представляющими коннекционизм вообще и нейросетевые модели в частности как реальное техническое преодоление критики сильного ИИ? Возможно ли более основательным образом отринуть взятое в чистом виде допущение С. Б. Куликова об альтернативных архитектурах? Ведь помимо упомянутых ранее Чёрчлендов и Л. Д. Робертса подобного рода аргументы или их обсуждение представляют обширный пласт в 'комнатоведческой' литературе [26–30]. Попробуем в масштабах настоящей статьи предложить модифицированную версию сёрловского мысленного эксперимента, скорректированную под обновлённую предпосылку (коннекционистский подход в лице генеративных нейросетей вместо прежнего символического). Если собственная модификация 'Китайской комнаты' окажется по крайней мере столь же убедительной, как аналогичные вариации от самого Сёрла¹⁰ (интерпретированная комната или искусственный мозг из клапанов и шлангов)¹¹, будем полагать поставленную задачу выполненной. В противном случае, при котором изменённый экспериментальный сценарий не будет справляться с тестируемым потенциалом генеративных нейросетей на статус сильного ИИ, зафиксируем ограниченность аргумента 'Китайской комнаты'. Тогда предположение С. Б. Куликова подтвердится: сёрловская критика ограничена онтологически определёнными архитектурами интеллектуальных систем.

Но что следует взять за основу при переносе генеративных ИИ в контекст мысленного эксперимента Сёрла? Имеем в виду, как наиболее достоверно представить 'Китайскую комнату', работающую наподобие выбранной архитектуры? Если отринуть предложенный самим американским философом путь простого умножения количества операторов (сёрловский 'китайский спортзал') и не соглашаться с буквальным имитированием взаимодействия нейронов (к чему обращается С. Б. Куликов), то что именно необходимо вычленив из функционирования генеративных нейросетей? Предлагаем за исковое основание взять описание того, как работает одна из многих нейросетевых архитектур. Например, каким способом действуют рекуррентные модели при обработке исходных и при *генерировании* новых данных [31]: «[рекуррентные нейросети] создают и обучают для многократного последовательного применения. При каждом запуске [такая сеть] использует своё внутреннее состояние с предыдущего шага как входные данные. Благодаря этому она может помнить свои прошлые действия. При генерации текста рекуррентной нейронной сетью обычно сначала подаётся короткий фрагмент текста («промпт»), затем сеть предсказывает следующую букву (или слово) и добавляет её к промпту. Далее она создаёт следующую букву или слово, опираясь на текущее состояние сети (учитывающее весь обработанный текст) и только что сгенерированную букву или слово. Фактически сеть

использует созданный ею текст как часть входных данных. Эти сети получили название рекуррентных из-за повторного обращения к выходным данным» [32, с. 111].

Тогда успешное встраивание генеративной нейросети на основе рекуррентной архитектуры можно представить внутри экспериментального пространства 'Китайской комнаты' в том случае, если: оператор в преобразённых комнатных условиях сможет имитировать относительно длинный диалог с пользователем снаружи; *пользователь* или собеседник снаружи сможет отсылать оператора к предыдущим вопросам и ответам в диалоговом контексте; оператор сможет успешно имитировать подобного рода общение без возникновения понимания; при всём перечисленном при ведении диалога сможет порождаться новый контент; пользователь будет удовлетворён такой коммуникацией, поэтому у него не появится сомнений в языковой компетентности оператора. Как представляется, при реализации задуманного получится нивелировать некоторые значимые нападки на изначальный эксперимент Сёрла. К примеру, имеющиеся в работе П. Кугеля: 'Китайская комната' не способна запоминать свои предыдущие *когнитивные состояния* (вопросы пользователя и ответы на них) и *осмысленно* сменять их (в соответствии с логикой ведения беседы при её восприятии собеседником) [33, р. 154]. Идентичную проблематику поднимает и Д. Б. Волков¹²: «['Китайскую комнату' можно поставить в тупик, если] задать ей вопрос о фактах, доступных обеим сторонам, ответ на который нельзя было бы подготовить заранее. [...] Например, сколько вопросов было уже задано с начала беседы?» [34, с. 8].

'Китайская картотека', экспериментальный реализм и аргумент выхода

Собственный вариант мысленного эксперимента назовём 'Китайской картотекой'. Вообразим общество (нацию, империю) амбициозное, способное и компетентное для проведения мега-проектов. Наподобие представленных в литературном примере Х. Л. Борхеса [35] или в мысленном эксперименте Н. Блока [36]. Такой общественной системе не составит труда направить ресурсы на практическое воплощение философской провокации Сёрла. Но, имея сведения о дискуссиях, критике и возражениях касательно 'Китайской комнаты', инициаторы проверки практикой сёрловских идей решают поступить следующим образом: помимо чернил, бумаги и прописанных на понятном оператору языке книги правил ввода-вывода китайских символов, человеку-внутри-комнаты дают также *доски, гвозди и молоток*. Зачем? Покажем через прокручивание возможной беседы с комнатным оператором.

1. Пользователь (далее — П.) вне комнаты задаёт вопрос.

2. П.: Привет. Что ты можешь рассказать о Великой Китайской стене? (далее — вопрос-1)

3. Оператор (далее — О.) получает бумажку с вопросом, записанным китайскими иероглифами, после чего ищет в сводах с правилами по манипулированию подобными символами, что следует записать в ответном бумажном послании. И спустя некоторое время находит программную строку для ответа на вопрос-1, после чего отвечает.

4. О.: Привет. Великая Китайская стена — крупнейшее рукотворное оборонительное сооруже-

ние в истории, возведённое ещё в Древнем Китае (далее — ответ-1).

5. П. получает письмо от О., но не удовлетворяется таким ответом и почти сразу пишет новый, уточняющий, вопрос.

6. П.: Можешь рассказать больше по моему предыдущему вопросу? (далее — вопрос-2).

Именно на этом этапе беседы человеку из комнаты потребуются доски, гвозди и молоток. Лишь заранее отметим, что наш О. всегда делает по два экземпляра своих ответных посланий для П., один из которых оставляет себе.

7. О. ищет в сводах правил строчку, в которой бы присутствовали китайские символы «можешь», «рассказать», «больше», «по моему», «предыдущему вопросу». Найдя совпадающую символическую комбинацию, человек обнаруживает, что книги на доступном ему языке не сразу призывают к записи нового ответа. Наоборот, к более кропотливому действию: «если видишь перечисленные в такой или приблизительно такой последовательности символы, то сколоть одну подвесную полку. Положи в неё вопрос-1. Затем сколоть ещё три подвесные полки одна за другой, пронумеровав их для своего удобства. Во вторую сколоченную полку положи ответ-1, в третью — вопрос-2, а четвёртую пока оставь пустой. Далее найди в других сводах правил максимально соответствующий комбинации символов из вопроса-1 заголовок страницы. Затем найди, что из символической последовательности на этой странице совпадает с символами из ответа-1. В итоге то, что не совпадает с символической комбинацией из ответа-1, запиши в новой ответной последовательности, после чего дубликат положи в оставшуюся четвёртую полку, а оригинал передай во внешний мир».

8. О.: Да, конечно, могу рассказать больше. Великая Китайская стена — это гигантская сеть укреплений, проходящая через весь Северный Китай. Длина основного её участка составляет около 8850 км. Она возводилась на протяжении более 2000 лет (до эпохи династии Мин, 1644 год). В среднем высота этого сооружения составляет от 6 до 8 метров, а толщина достигает 5–7 метров.

Даже не имея феноменальной сверхпамяти, оператор в усовершенствованной комнате всё же может запоминать и ориентироваться в беседе с человеком за комнатными границами. Притом возникновение понимания касательно коммуникативного содержания для продолжения коммуникации не требуется. Достаточно, чтобы своды правил (т. е. программа на понятном оператору языке, созданная для имитации адекватного вопроса-ответного лингвистического поведения китайца) содержали не только вводные и выводные последовательности из китайских иероглифов. Следует добавить к языковым правилам расширенные инструкции, которые позволят 'Китайской комнате' стать 'Китайской картотекой'. Быть может, представить подобную картотечную вселенную, позволяющую О. 'запомнить' абсолютно все диалоги со всеми пользователями, затруднительно, но вполне допустимо представить множественные картотеки по каждой отдельной беседе.

Для придания большей реалистичности такой модификации мысленного эксперимента Сёрла можно внести дополнительные уточнения. Например, пользователи-собеседники также могут и должны будут иметь свои собственные руководства по общению с 'Китайской картотекой'. Ко-

нечно, не такие объёмные, как своды правил для оператора, но как минимум содержащие указания следующего характера: «обязательно используйте в целях продолжения общения с картотекой фразы 'предыдущий вопрос', 'предыдущий ответ', 'расскажи об этом подробнее' и др.». Также возможно домыслить происхождение тех самых сводов правил, которыми пользуется человек внутри комнаты-картотеки. Их могут составлять и даже постоянно обновлять, заменять и дополнять огромная группа из социологов, статистиков и психологов. Последние проводят регулярные опросы и интервью со всем населением воображаемого общества, внося в своды правил для оператора наиболее частотные ответы условного среднестатистического носителя китайского языка.

Трудно, но с некоторыми усилиями можно представить, что это же экстравагантное сообщество экспериментаторов, стараясь приложить сёрловские идеи к имитации генерирования заявленного выше нового контента без возникновения понимания, будет в силах составить дополнительные правила для оператора именно с этой целью. Иначе говоря, научить человека внутри комнаты-картотеки помимо 'запоминания' диалога ещё и 'улучшать' тексты, к примеру. Возможно, с помощью заранее оговорённых с пользователями ограничений по объёму. Если есть логическая и физическая возможности — а они есть (большие языковые модели на рекуррентной или трансформерной платформах тому наглядный пример) — вывести и зафиксировать в программных строках правила по улучшению вводного текста и его преобразованию в улучшенный выводной текст (по требованиям улучшения со стороны пользователя). Опять же представления о том, что такое *улучшенный текст*, можно хотя бы попытаться социологически, психологически и статистически измерить, а после перевести в своды правил для оператора; и регулярно обновлять последние вслед за культурными, языковыми и иными трансформациями общества. Но подобный творческий функционал комнаты-картотеки, надо признать, находится на границах представимости, не говоря о шестивии по лезвию реализуемости вне мысленного эксперимента.

Казалось бы, при всех прочих равных, оговорах и условностях, изначальный экспериментальный сценарий Сёрла допустимо преобразовать в предложенный здесь. Подобное расширение позволяет имитировать гораздо более сложное языковое и, на первый взгляд, *сознательное* (понимающее, запоминающее, меняющееся) поведение. Притом не приводящее к возникновению понимания и верной интерпретации синтаксических правил, даже не требующее семантики для собственного осуществления. Следовательно, выдвинутое стартовое возражение С. Б. Куликова о том, что генеративные нейросети преодолевают сёрловскую критику, можно считать опровергнутым. Но представленная экспериментальная сложность, к счастью или к сожалению, до сих пор разрушила крайне простыми вопросами. Для демонстрации последних вновь стоит обратиться к Д. Б. Волкову, а точнее к тому, что он называет *аргументом внутреннего времени*: «[п]рограмму, которая бы правильно отвечала на вопросы о самой беседе, всё равно можно помыслить. Тогда члены жюри должны задавать вопросы о внутреннем чувстве. Например, *вопрос, сколько времени прошло с момента первого вопроса?* Разговор с узником может длиться сколь угодно

долго, интервалы между вопросами могут быть любыми, поэтому программа не сможет заранее предусмотреть ответ. [...] Чтобы поставить программу в тупик, можно попросить узника сравнить между собой два интервала времени или что-нибудь подобное. Насколько я понимаю, нельзя написать инструкции так, чтобы предусмотреть правильный разумный ответ» [34, с. 8].

К аналогичным языковым входным данным можно добавить и такие вопросы: про банку 'кока-колы' (о том, что находится перед оператором, внутри комнаты?) [37]; про сверку времени (о том, который час?) [38]; а также целый ряд вопросов *субкогнитивного* характера. К ним можно отнести вопросы с применением неологизмов («какое слово кажется вам более красивым — *blutch* или *farfaletta*?»), с использованием категорий («оцените ручки как оружие»), с опорой на человеческие ощущения («ощущение, когда держишь во рту глоток 'кока-колы', больше похоже на покалывание в онемевшей ноге или на то, как холодную воду выливают на голову?») и любые иные конструкции, направленные на раскрытие физиологического, культурного, жизненного опыта говорящего [13]. Перечисленные типы вопросных позиций указывают на то, что существуют такие языковые входные данные, из которых принципиально — в рамках аргументации Сёрла, даже с её усовершенствованием — не могут быть получены адекватные сцепки символов *на выходе*. Собственно, это и есть заявленный выше аргумент выхода: при полном согласии со всеми теоретическими предпосылками американского философа (*вход*) вместе с признанием продуманности самого сценария мысленного эксперимента (*обработка*) 'Китайская комната' всё равно не будет работать (*выход*). Как минимум комнатный оператор не сможет исчерпывающим образом имитировать поведение, которое позволит признать его полноценным и понимающим носителем китайского языка.

Возможные попытки дополнительно модифицировать комнату-картотеку так, чтобы преодолеть представленные ограничения, приведут либо к ослаблению критической позиции Сёрла, либо сами по себе будут выходить далеко за пределы как физической, так и логической возможности. К примеру, можно попробовать ограничить задавание указанных типов вопросов со стороны пользователей. Но это потребует расширения пользовательского руководства до масштабов сводов правил у оператора или попросту приведёт к тому, что оператора уже не смогут признать полноценным носителем китайского языка. А дальнейшее умножение сводов правил и прочего комнатного инвентаря приведёт к сведению к минимуму *экспериментального реализма*. Например, даже если представить себе оператора и со сверхпамятью, и с невероятной ловкостью, и с небывалой быстротой¹³, попытки улучшить иные комнатные элементы попросту сделают из 'Китайской комнаты' версию *Вавилонской библиотеки*. Сущности, осмысленный поиск чего-то конкретного в которой и логически, и физически невозможен [39].

Таким образом, следует констатировать, что и в изначальной, и в модифицированной версии мысленный эксперимент Сёрла обладает фундаментальными ограничениями. А именно такими, которые — даже при устранении критики со стороны аргументов входа и обработки — приводят к невозможности нивелировать аргументы выхода.

По крайней мере, подобрать некий контраргумент такого рода представляется весьма затруднительным предприятием. Попробовать реализовать последнее следует в рамках самостоятельного исследования.

Заключение

Исследование предложенного С. Б. Куликовым возражения против позиции Сёрла (развитие исключительно вычислительных моделей и программной базы не позволит создать сильный ИИ), а также выдвинутой самим автором критики подобного возражения привело к нескольким выводам.

Первый: критиковать ответ от прогресса (т. е. появление новых архитектур, например, генеративных нейросетей на трансформерной или рекуррентной платформе) через апелляцию к биологическому натурализму и аргументацию против ответа от моделирования работы мозга нерелевантно.

Второй: для более подробного изучения силы и ограничений выдвинутого С. Б. Куликовым возражения был выбран путь контекстуализации, т. е. включения ответа от прогресса в лице генеративных нейросетей внутрь 'комнатоведческих' дискуссий. Последнее потребовало поиска и сравнения различных подходов к классификации критических ответов против позиции Сёрла.

Третий: при проведении сравнительного анализа найденных классификационных подходов сделан вывод касательно их недостаточности в контексте рассмотрения возражения С. Б. Куликова и внутри дебатов по 'Китайской комнате' в принципе. В связи с этим разработан собственный способ классифицировать ответы за и против позиции Сёрла — структурный подход. Данный подход предполагает выделение аргументов входа, обработки и выхода, а также различение соответствующих контраргументов.

Четвёртый: возражение С. Б. Куликова как аргумент входа предлагается снять при помощи двух концептуальных шагов. На первом предполагаемое неучитывание Сёрлом при проектировании мысленного эксперимента и общей аргументации против сильного ИИ ослабляется через обращение к стилистике философа и историческому контексту публикации *Сознание, мозг и программы*. На втором, заключающем шаге, ответ от прогресса в лице генеративных нейросетей нивелируется посредством тестового 'прогона' модифицированной версии 'Китайской комнаты' под названием 'Китайская картотека'. Последняя, как предполагается, способна имитировать работу генеративной рекуррентной нейросети.

Пятый: несмотря на то, что изначальное возражение С. Б. Куликова возможно снять при помощи приведённой аргументации, усложнённая версия 'Китайской комнаты' тем не менее сохраняет прежние и редко подсвечиваемые концептуально слабые места. Все они так или иначе сводятся к тому, что комнатный оператор не сможет ответить на все возможные вопросы. Ведь последнее потребует либо возникновения у обитателя комнаты-картотеки понимания, либо нереалистичного усложнения мысленного эксперимента.

Финальный, пятый, вывод обосновывает необходимость либо отказаться от позиций Сёрла — притом не только по вопросу сильного ИИ, но и природы сознания, — либо существенным образом пересобрать сам его аргумент. Притом на бо-

лее глубинном уровне, чем это было сделано в лице 'Китайской картотеки'. Подобная реконфигурация мысленного эксперимента для реабилитации взглядов философа станет целью следующей статьи.

Примечания

¹ Разумеется, из подобной дисциплины следует исключить рассуждения и вариации по мотивам другой известной комнаты философов — 'комнаты Мэри'.

² Заранее обозначим, что в настоящей статье под 'Китайской комнатой' будет пониматься не просто одноимённый мысленный эксперимент, но аргумент. Притом в его структуру экспериментальная ситуация входит как составная, но не ключевая составляющая. Подобная интерпретация критики Сёрлом 'сильного' искусственного интеллекта укоренена в самом наименовании 'Китайской комнаты' — *The Chinese Room Argument*.

³ Дж. Морал допускает, что смешение теоретического аргумента и мысленного эксперимента — и даже смешение внимания с первого на второе — внутри 'Китайской комнаты' могло быть намеренной *макиавеллистской стратегией* со стороны Сёрла для запутывания и разобщения оппонентов [18, p. 254].

⁴ Далее будем писать «сильный ИИ» вместо «'сильный' искусственный интеллект».

⁵ Далее по тексту подход по созданию сильного ИИ, связанный с использованием данных нейрофизиологии для проектирования архитектур с параллельным вычислением и распределённой памятью, будем обозначать при помощи термина 'коннекционистский подход'. Среди продуктов последнего как раз можно выделить генеративные модели на базе трансформерной, рекуррентной и иных архитектур [40, с. 274].

⁶ Возможно, более подходящими обозначениями являются — вместо *входа* и *выхода* — ввод и вывод, но в контексте критического комментирования статьи С. Б. Куликова были выбраны именно эти наименования.

⁷ Тем не менее впоследствии автор пишет, что сёрловское мысленное экспериментирование стоит выше частных технических воплощений (сильного) ИИ, что делает его позицию в некотором смысле трансцендентальной: «утверждения Сёрла нацелены на основания [сильного] ИИ. Они затрагивают его природу в принципе, а не только касаются сугубо технических решений. И в этом смысле Сёрл в не меньшей степени 'трансценденталист', нежели многие его противники из стана антинатурализма» [17, с. 80]. Но насколько это соответствует рассуждениям самого Сёрла (по крайней мере, периода публикации *Сознание, мозг и программы*)? Представляется, что соответствие это как минимум неполное, ведь сам американский философ, по сути, предлагает *практическую (техническую)* критику, преодолевающую ограничения созданной им же 'Китайской комнаты'. Иными словами, он призывает сменить подход к созданию ИИ с вычислительного на иной, учитывающий каузальные силы, порождающие интенциональность и, следовательно, сознание внутри человеческого мозга [4, с. 399]. Пусть предлагаемый Сёрлом подход слишком радикально отличается от мейнстрима, это всё же вполне эмпирический путь к понимаю (и к созданию) сильного ИИ, а не выстраиваемая трансцендентальная преграда для последнего: «[может ли мыслить артефакт, машина, сделанная человеком?] В предположении, что возможно искусственно произвести машину с нервной системой, нейронами, обладающими аксонами и дендритами, и со всем прочим, в достаточной степени похожими на нашу нервную систему, наши нейроны и т. д., ответ на этот вопрос представляется опять же тривиально положительным. Если вы в состоянии точно продублировать причины, то вы в состоянии продублировать и следствия. И на самом деле, возможно, быть может, продублировать сознание, интенциональность и всё такое прочее, используя какие-то другие химические принципы, чем те, что

реализованы в людях. Это, как я сказал, *вопрос эмпирический*» [4, с. 393–394].

⁸ См., например, работу М. Н. Белянина по вопросам формирования идейного контекста в творчестве Сёрла для создания 'Китайской комнаты' [42].

⁹ Поэтому обратный приведённому историческому контраргументу аргумент Р. Френча (о том, что публикация *Сознание, мозг и программы* пришлась на «доконнекционистский год», следовательно, изначальная сёрловская аргументация могла учитывать только символический подход [13]) не верен полностью или частично: коннекционизм уже существовал и как теория, и как набор из практических проектов.

¹⁰ Притом ту вариацию 'Китайской комнаты', которую Сёрл предложил против коннекционизма в полемике с Чёрчлендами ('*китайский спортзал*'), вслед за устоявшейся линией в литературе [18, p. 228], не будем признавать достаточной и приемлемой. Именно поэтому необходима иная модификация мысленного экспериментирования против коннекционистского подхода как претендента на воплощение сильного ИИ.

¹¹ Разумеется, последняя инстанция признания или непризнания убедительности модифицированного мысленного эксперимента — это *читатель*. Здесь же под убедительностью будем подразумевать, справляется ли экспериментальная вариация — хотя бы при первом приближении — с предлагаемым критическим ответом 'Китайской комнаты'. Да, не все комментаторы полагают, что и перечисленные сёрловские варианты в принципе работают так, как заявлено мыслителем. Но будем стараться при собственном модифицировании предоставить такую версию, которая была бы — *при прочих равных* — сопоставима с экспериментальными модификациями из-под пера самого Сёрла.

¹² Притом автор так же, как и мы (см. в следующем разделе статьи), считает возможным преодолеть критические ответы 'Китайской комнате' *от контекстуализации общения с ней* (т. е. с помощью отправки вопросов о контексте заданного в предыдущих вопросах): «[И. В. Солнцев] предлагает спросить Комнату о содержании предыдущего вопроса. Этот вариант, по моему мнению, можно парировать. Комната теоретически может иметь все комбинации вопросов и ответов. Тогда и на этот у неё будет правильный ответ» [34, с. 8]. Правда, далее предлагается вообразить не то, что внутри комнатных сводов правил ввода и вывода есть все возможные ответы и вопросы, их бесконечно большое и постоянно растущее количество. Предложение состоит в модификации '*архитектуры*' 'Китайской комнаты' через большее её уподобление устройству рекуррентных нейросетевых моделей.

¹³ Возможно, доведённую до максимальных физических границ; или сам оператор находится в помещении с ускоренным ходом времени относительно пользователей. И то, и другое, конечно, уводит всё дальше и дальше из плоскости философских спекуляций в область *научной фантастики*.

Список источников / References

1. Эдмондс Д. Убили бы вы толстяка? Задача о вагонетке: что такое хорошо и что такое плохо? / пер. с англ. Д. Кралечкина. 3-е изд., испр. Москва: Изд-во Института Гайдара, 2019. 264 с. ISBN 978-5-93255-552-1.

Edmonds D. Ubili by vy tolstyaka? Zadacha o vagonetke: chto takoye khorosho i chto takoye plokho? / trans. from Engl. D. Kralechkin. 3rd ed., rev. Moscow, 2019. 264 p. ISBN 978-5-93255-552-1. (In Russ.).

2. Appiah K. A. Experiments in Ethics. Cambridge, MA: Harvard University Press, 2008. 274 p. ('Гарвардское университетское издательство (Harvard University Press) внесено в реестр нежелательных организаций на территории Российской Федерации).

3. Пляскина А. С., Подьяков А. Н. Визуальные репрезентации «проблемы вагонетки» в интернет-мемах: политексту-

альный тематический анализ // Вопросы психологии. 2022. Т. 68, № 6. С. 64–79. EDN: AWINAW.

Playskina A. S., Podd'yakov A. N. Visual Representations of the "Trolley Problem" in Internet Memes: A Polytextual Thematic Analysis. *Voprosy Psikhologii*. 2022;68(6):64–79. EDN: AWINAW. (In Russ.).

4. Сёрл Дж. Р. Мозг, сознание и программы // Аналитическая философия: становление и развитие: антология / пер. с англ. А. Блинова. Москва: Дом интеллектуальной книги; Прогресс-Традиция, 1998. С. 376–400.

Searle J. R. Minds, Brains, and Programs. *Analytic Philosophy: Formation and Development: Anthology*. Moscow, 1998. P. 376–400. (In Russ.).

5. Cole D. The Chinese Room Argument. *The Stanford Encyclopedia of Philosophy* / ed. by E. N. Zalta, U. Nodelman. 2024.

6. Shaffer M. J. A Logical Hole in the Chinese Room. *Minds and Machines*. 2009;19:229–235. <https://doi.org/10.1007/s11023-009-9151-9>.

7. Tanaka K. Minds, Programs, and Chinese Philosophers: A Chinese Perspective on the Chinese Room. *Sophia*. 2004; 43:61–72. <https://doi.org/10.1007/bf02782437>.

8. Fisher M., List C., Slavkovik M., Weiss A. Ethics and Trust: Principles, Verification and Validation (Dagstuhl Seminar 19171). *Dagstuhl Reports*. 2019;9(4):59–86. <https://doi.org/10.4230/DagRep.9.4.59>.

9. Gavalas D. From Searle's Chinese Room to the Mathematics Classroom: Technical and Cognitive Mathematics. *Studies in Philosophy and Education*. 2007;26(2):127–146. <https://doi.org/10.1007/s11217-006-9018-y>.

10. Levesque H. J. Is It Enough to Get the Behavior Right? *IJCAI 2009: Proceedings of the 21st International Joint Conference on Artificial Intelligence*. 2009. P. 1439–1444.

11. Лем С. Тайна китайской комнаты // Молох / пер. с польск. В. Язневича. Москва: АСТ, АСТ Москва, Хранитель; Минск: Харвест, 2006. С. 246–255.

Lem S. The Secret of the Chinese Room. *Molokh* / trans. from Polish V. Yaznevich. Moscow; Minsk, 2006. P. 246–255. (In Russ.).

12. Teşeleanu G. Books, Rooms and Searle: Short Paper. *Cyber Security, Cryptology, and Machine Learning 2025*. Cham: Springer, 2025. P. 326–336. https://doi.org/10.1007/978-3-032-10759-6_22.

13. French R. M. The Chinese Room: Just Say "No"! *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000. P. 657–662.

14. Хофштадтер Д. Размышления // Глаз разума: Фантазии и размышления о самосознании и о душе / сост. Д. Хофштадтер, Д. Деннет; пер. с англ. М. Эскиной. Самара: Бахрах-М, 2003. С. 331–340.

Hofstadter D. Reflections. *The Mind's. Fantasies and Reflections on Self and Soul* / comp. by D. Hofstadter, D. Dennett; trans. from Engl. M. Eskina. Samara, 2003. P. 331–340. (In Russ.).

15. Puccetti R. The Chess Room: Further Demythologizing of Strong AI. *Behavioral and Brain Sciences*. 1980;3(3):441–442. <https://doi.org/10.1017/S0140525X00005938>.

16. Rapaport W. J. How Helen Keller Used Syntactic Semantics to Escape from a Chinese Room. *Minds and Machines*. 2006;16:381–436. <https://doi.org/10.1007/s11023-007-9054-6>.

17. Куликов С. Б. Всё ли учёл Дж. Сёрл при входе в 'Китайскую комнату'? // Омский научный вестник. Сер. Общество. История. Современность. 2026. Т. 11, № 1. С. 78–83. <https://doi.org/10.25206/2542-0488-2026-11-1-78-83>. EDN: AFHRKU.

Kulikov S. B. Did J. Searle Take Everything into Account When Entering the Chinese Room? *Omsk Scientific Bulletin. Series Society. History. Modernity*. 2026;11(1):78–83. <https://doi.org/10.25206/2542-0488-2026-11-1-78-83>. EDN: AFHRKU. (In Russ.).

18. Moulal J. The Chinese Room Argument. *John Searle* / ed. by B. Smith. Cambridge: Cambridge University Press, 2003. P. 214–260.

19. Сёрл Дж. Разум мозга — компьютерная программа? // PsychoSearch. URL: <https://psychosearch.ru/teoriya/psikhika/338-searle-john-razum-mozga-kompyuternaya-programma> (дата обращения: 08.06.2026).

Searle J. Is the Brain's Mind a Computer Program? *PsychoSearch*. URL: <https://psychosearch.ru/teoriya/psikhika/338-searle-john-razum-mozga-kompyuternaya-programma> (accessed: 08.06.2026). (In Russ.).

20. Марков С. Охота на электроовец. Большая книга искусственного интеллекта: в 2 т. Т. 1. Москва, 2024. 568 с. ISBN 978-5-600-04012-0.

Markov S. Okhota na elektroovets. *Bol'shaya kniga iskusstvennogo intellekta: in 2 vols. Vol. 1*. Moscow, 2024. 568 p. ISBN 978-5-600-04012-0. (In Russ.).

21. Roberts L. D. Searle's Extension of the Chinese Room to Connectionist Machines. *Journal of Experimental & Theoretical Artificial Intelligence*. 1990;2(2):185–187. <https://doi.org/10.1080/09528139008953722>.

22. Деннет Д. Насосы интуиции и другие инструменты мышления / пер. с англ. З. Мамедьярова, Е. Фоменко. Москва: АСТ: CORPUS, 2019. 576 с.

Dennett D. Intuition Pumps and Other Tools for Thinking / trans. from Engl. Z. Mamed'yarova, E. Fomenko. Moscow, 2019. 576 p. (In Russ.).

23. Damper R. I. The Logic of Searle's Chinese Room Argument. *Minds and Machines*. 2006;16(2):163–183. <https://doi.org/10.1007/s11023-006-9031-5>.

24. Белянин М. Н. От Остина к Сёрлу // Проза.ру. 2016. URL: <https://proza.ru/2016/11/19/805> (дата обращения: 01.07.2026).

Belyanin M. N. From Austin to Searle. *Proza.ru*. 2016. URL: <https://proza.ru/2016/11/19/805> (accessed: 01.07.2026). (In Russ.).

25. Алексеева Е. А. Противостояние символизма и коннекционизма в истории развития искусственного интеллекта // Электронный научно-образовательный журнал «История». 2020. Т. 11, вып. 11 (97). <https://doi.org/10.18254/S207987840013021-2>. EDN: CPPFID.

Alekseyeva E. A. The Opposition of Symbolism and Connectionism in the History of Artificial Intelligence Development. *ISTORIYA*. 2020;11(97). <https://doi.org/10.18254/S207987840013021-2>. EDN: CPPFID. (In Russ.).

26. García-Valdecasas M. Are Large Language Models Intentional? The Limits of Referential Grounding. *Philosophy & Technology*. 2026;39(62). <https://doi.org/10.1007/s13347-026-01079-4>.

27. Borg E. LLMs, Turing Tests and Chinese Rooms: The Prospects for Meaning in Large Language Models. *Inquiry*. 2026;69(6):2807–2837. <https://doi.org/10.1080/0020174X.2024.2446241>.

28. Šekrst K. Chinese Chat Room: AI Hallucinations, Epistemology and Cognition. *Studies in Logic, Grammar and Rhetoric*. 2024;69(1):365–381. <https://doi.org/10.2478/slgr-2024-0029>.

29. Rakover S. S. The Computer That Simulated John Searle in the Chinese Room. *New Ideas in Psychology*. 1999;17(1):55–66. [https://doi.org/10.1016/S0732-118X\(98\)00016-6](https://doi.org/10.1016/S0732-118X(98)00016-6).

30. Copeland B. J. The Curious Case of the Chinese Gym. *Synthese*. 1993;95(2):173–186. <https://doi.org/10.1007/BF01064587>.

31. Graves A. Generating Sequences with Recurrent Neural Networks. *arXiv*. 2013. <https://doi.org/10.48550/arXiv.1308.0850>.

32. Тогелиус Ю. Общий искусственный интеллект / пер. с англ.; под науч. ред. А. Смирнова. Москва: Изд-во Института Гайдара, 2025. 208 с. ISBN 978-5-93255-691-7.

Togelius J. Obschchiy iskusstvennyy intellekt / trans. from Engl.; sci. ed. A. Smirnov. Moscow, 2025. 208 p. ISBN 978-5-93255-691-7. (In Russ.).

33. Kugel P. The Chinese Room Is a Trick. *Behavioral and Brain Sciences*. 2004;27(1):153–154. <https://doi.org/10.1017/S0140525X04210044>. EDN: HKVFRT.

34. Волков Д. Б. Аргумент внутреннего времени // Финиковый компот. 2015. № 8. С. 6–8.
Volkov D. B. The Argument of Internal Time. *Date Palm Compote*. 2015;8:6–8. (In Russ.).
35. Борхес Х. Л., Биой Касарес А. Нынешний натурализм // Хроники Бустоса Домека / пер. с исп. Е. Лысенко. Москва: АСТ, Астрель, Neoclassic, 2011. С. 48–58.
Borges J. L., Bioy Casares A. Naturalism Today. *Chronicles of Bustos Domecq* / trans. from Spanish E. Lysenko. Moscow, 2011. P. 48–58. (In Russ.).
36. Block N. Troubles with Functionalism. Perception and Cognition / ed. by C. W. Savage. *Minnesota Studies in the Philosophy of Science*. Vol. 9. Minneapolis: University of Minnesota Press, 1978. P. 261–325.
37. Васильев В. В. Кока-кола и секрет 'Китайской комнаты' // Философия сознания: классика и современность. Москва: Издатель Савин С. А., 2007. С. 86–94. EDN: VTHRZF.
Vasil'yev V. V. Koka-kola i sekret 'Kitayskoy komnaty'. *Filosofiya Soznaniya: Klassika i Sovremennost'*. Moscow, 2007. P. 86–94. EDN: VTHRZF. (In Russ.).
38. Ben-Yami H. A Note on the Chinese Room. *Synthese*. 1993;95(2):169–172. <https://doi.org/10.1007/BF01064586>.
39. Реализм против Теории Пыли, или Как изгнать Демона Второго Рода из Вавилонской библиотеки // Хабр. 2025. URL: <https://habr.com/ru/articles/912242/> (дата обращения: 01.07.2026).
Realizm protiv Teorii Pyli, ili Kak izgnat' Demona Vtorogo Roda iz Vavilonskoy biblioteki. *Habr*. 2025. URL: <https://habr.com/ru/articles/912242/> (accessed: 01.07.2026). (In Russ.).
40. Марков С. Охота на электроовец. Большая книга искусственного интеллекта: в 2 т. Т. 2. Москва, 2024. 784 с. ISBN 978-5-600-04012-0.
Markov S. Okhota na elektroovets. Bol'shaya kniga iskusstvennogo intellekta: in 2 vols. Vol. 2. Moscow, 2024. 784 p. ISBN 978-5-600-04012-0. (In Russ.).
41. Damper R. I. The Chinese Room Argument: Dead but Not Yet Buried. *Journal of Consciousness Studies*. 2004;5–6(11): 159–169.
42. Беянин М. Н. Джон Серл и «трудная проблема сознания» // Вестник Московского университета. Сер. 7: Философия. 2011. № 4. С. 38–50. EDN: OKMLMZ.
Belyanin M. N. John Searle and the Hard Problem of Consciousness. *Lomonosov Philosophy Journal*. 2011;4:38–50. EDN: OKMLMZ. (In Russ.).

Информация об авторе:

КАЗАНЦЕВ Осип Андреевич, преподаватель кафедры «Конфликтология и национальная безопас-

ность» Института социологии и регионоведения Южного федерального университета, г. Ростов-на-Дону.

SPIN-код: 8274-0814

AuthorID (РИНЦ): 1136782

ORCID: <https://orcid.org/0009-0002-0312-989X>

e-mail: koznew@bk.ru

Прозрачность финансовой деятельности: автор не имеет финансовой заинтересованности в представленных материалах и методах. Конфликт интересов отсутствует.

Автор прочитал и одобрил окончательный вариант рукописи.

При подготовке данной рукописи автор не использовал генеративный искусственный интеллект или аналогичные автоматизированные инструменты.

Статья поступила в редакцию 10.07.2026; одобрена после рецензирования 28.07.2026; принята к публикации 08.09.2026.

Information about the author:

KAZANTSEV Osip Andreevich, Lecturer of the Conflict Science and National Security Department, Institute of Sociology and Regional Studies, Southern Federal University, Rostov-on-Don.

SPIN-code: 8274-0814

AuthorID (RSCI): 1136782

ORCID: <https://orcid.org/0009-0002-0312-989X>

e-mail: koznew@bk.ru

Financial transparency: the author has no financial interest in the presented materials or methods. There is no conflict of interest.

The author has read and approved the final manuscript. The author did not use the Artificial Intelligence or similar automated tools in the article.

The article was submitted 10.07.2026; approved after reviewing 28.07.2026; accepted for publication 08.09.2026.