

# Философия общения и искусственный интеллект: опыт сравнительного анализа дискуссий в отечественной и зарубежной литературе

**Резаев Андрей**

**Владимирович** – доктор философских наук, профессор, руководитель международной исследовательской лаборатории ТАНДЕМ. Санкт-Петербургский государственный университет. Российская Федерация, 191124, г. Санкт-Петербург, ул. Смольного, д. 1/3, 9-й подъезд;  
e-mail: rezaev@hotmail.com

**Трегубова Наталья**

**Дамировна** – кандидат социологических наук, доцент. Санкт-Петербургский государственный университет. Российская Федерация, 191124, г. Санкт-Петербург, ул. Смольного, д. 1/3, 9-й подъезд;  
e-mail: n.tregubova@spbu.ru

Статья посвящена сравнительному анализу трех дискуссий о проекте искусственного интеллекта, проходивших по обе стороны «железного занавеса», начиная с 1960-х гг.: 1) Э.В. Ильенков – Д.И. Дубровский (СССР), 2) Х. Дрейфус – представители компьютерных наук (США), 3) Х. Дрейфус – Г. Коллинз (США – Великобритания). В качестве основания для сравнительного анализа дискуссий выступает типология парадигм общения, предполагающая существование трех принципиально разных взглядов на человека как на социальное существо. Эти взгляды могут быть перенесены на представления о человеке в рамках различных направлений теории и критики искусственного интеллекта. Информационно-инструментальная парадигма ставит проблему общения в терминах взаимосвязи – взаимодействия и обмена образованиями материального и духовного порядка (Д.И. Дубровский, представители компьютерных наук). Социально-институциональная парадигма рассматривает общение как персонификацию, актуализацию и реализацию структурной характеристики общества – общественных отношений (Э.В. Ильенков). Экзистенциально-феноменологическая парадигма характеризует общение как реализацию внутренней потенции самосознающего индивида (Х. Дрейфус). Выделяются «два с половиной этажа» в исследовании проблем, связанных с вхождением технологий искусственного интеллекта в жизнь общества, для каждого из которых характерна особая парадигмальная структура и особый способ постановки исследовательских проблем: 1) философия, компьютерные и когнитивные науки; 2) социальные науки; 3) междисциплинарные исследования. В завершение выделены «новые старые вопросы», которыми сегодня структурируется поле социальной аналитики искусственного интеллекта и искусственной социальности.

**Ключевые слова:** искусственный интеллект, искусственная социальность, Э.В. Ильенков, Д.И. Дубровский, Х. Дрейфус, Г. Коллинз, парадигмы общения





# PHILOSOPHY OF SOCIAL INTERCOURSE AND ARTIFICIAL INTELLIGENCE: A COMPARATIVE ANALYSIS

**Andrey V. Rezaev** –  
DSc in Philosophy, Head  
of International Research  
Laboratory TANDEM.  
Saint Petersburg State  
University.  
1/3-9 Smolnogo St.,  
Saint Petersburg 191124,  
Russian Federation;  
e-mail: rezaev@hotmail.com

**Natalia D. Tregubova** –  
PhD in Sociology, Associate  
Professor of Comparative  
Sociology Chair.  
Saint Petersburg State  
University.  
1/3-9 Smolnogo St.,  
Saint Petersburg 191124,  
Russian Federation;  
e-mail: n.tregubova@spbu.ru

The paper aims to analyze three discussions pertaining to the artificial intelligence project that took place on both sides of the “Iron Curtain” since the 1960s: 1) E.V. Ilyenkov – D.I. Dubrovsky (USSR), 2) H. Dreyfus – computer scientists (USA), 3) H. Dreyfus – H. Collins (USA – UK). The authors observe the originality of the arguments of Soviet philosophers in comparison with the discussions in the US and UK. The basis for a comparative analysis of these discussions is the typology of social intercourse paradigms, which assumes the existence of three fundamentally different views of humans as social beings: the information-instrumental, the existential-phenomenological, and the social-institutional paradigms. These views correlate with the ideas about humans within different theoretical structures examining artificial intelligence. The information-instrumental paradigm poses the problem of social intercourse regarding communication, material, and informational exchange (D.I. Dubrovsky, computer scientists). The social-institutional paradigm considers social intercourse as personification, actualization, and implementation of the structural characteristics of social relations (E.V. Ilyenkov). The existential-phenomenological paradigm characterizes social intercourse as the realization of the internal potency of a self-conscious individual (H. Dreyfus). The paper exposes that the theoretical and methodological foundations of the discussions under consideration are similar to the paradigmatic foundations of the analysis of social intercourse in social philosophy and social sciences. There are “two and a half floors” in the research of problems associated with the entry of artificial intelligence technologies into the everyday life of society: 1) philosophy, computer and cognitive sciences; 2) social sciences; 3) interdisciplinary research. A particular paradigmatic structure and a unique research problem characterize each of them. In conclusion, “new old questions” are highlighted that structure the field of contemporary social analytics for studying artificial intelligence and artificial sociality.

**Keywords:** artificial intelligence, artificial sociality, E.V. Ilyenkov, D.I. Dubrovsky, H. Dreyfus, H. Collins, social intercourse paradigms

## Постановка проблемы

Едва ли можно сомневаться в том, что время, которое мы переживаем сегодня, властно и решительно требует новых подходов к осмыслению взаимодействий человека с «умными» машинами и алгоритмами. Однако даже беглый взгляд на содержание современных дискуссий об искусственном интеллекте (ИИ) позволяет заметить, что вопросы, в них обсуждаемые, формулировались еще пятьдесят-семь-



десять лет назад. Причем особенностью современных выступлений и дискуссий об инструментах и технологиях ИИ является тот факт, что их участники занимают не просто противоположные, но несоизмеримые позиции. Нам представляется логичным и необходимым проанализировать «старые» дебаты об ИИ, характерные для второй половины прошлого столетия, обращая внимание на их значимость для развития сегодняшнего и завтрашнего взглядов на искусственный интеллект и искусственную социальность (ИС)<sup>1</sup>.

Нашей целью в данной работе является анализ дискуссии об «умных машинах», имевшей место в СССР, в сравнении с дискуссиями, проходившими параллельно в США. На основании предпринятого анализа будет определена и обоснована их актуальность для современной социальной аналитики. В рамках настоящей статьи мы обратим особое внимание на три дискуссии: 1) Э.В. Ильенков – Д.И. Дубровский, 2) Х. Дрейфус – С. Пайперт, Дж. Маккарти и другие представители компьютерных наук, 3) Х. Дрейфус – Г. Коллинз.

Каковы причины, определившие выбор именно этих трех дискуссий? Во-первых, выбранные нами дискуссии являются *междисциплинарными* – они проходят между представителями разных дисциплин и исследовательских традиций. Их участники основываются на разных допущениях об искусственном и о естественном интеллекте, которые в процессе взаимной критики становятся очевидны. Во-вторых, в отличие от некоторых других дебатов<sup>2</sup>, в трех дискуссиях, к которым мы обращаемся, участники не могут договориться о самих основаниях спора, поскольку занимают принципиально разные позиции по поводу того, что есть человек – не сознание, не интеллект, не разум, а именно человек. Сам проект ИИ изначально был адисциплинарным: разработчики стремились воспроизвести человека в целом, и здесь им не могла помочь никакая конкретная научная дисциплина<sup>3</sup>. Нас же интересуют *адисциплинарные* дискуссии о проблеме ИИ, поскольку именно они позволяют увидеть в наибо-

<sup>1</sup> Под искусственной социальностью здесь и далее понимается система взаимодействий с участием человека и искусственного интеллекта, в которых ИИ может выступать как посредником во взаимодействии между людьми, так и самостоятельным участником [Резаев, Трегубова, 2019].

<sup>2</sup> Например, вокруг мысленного эксперимента с «китайской комнатой» Дж. Сёрля и дискуссии вокруг работ Р. Пенроуза.

<sup>3</sup> Сегодня «искусственный интеллект» – это субдисциплина внутри компьютерных наук, которая занимается решением технических задач. Вместе с тем изначальный проект ИИ – проект воспроизведения человека – остается адисциплинарным предприятием. Представляется, что и проблемы вхождения технологий ИИ в жизнь общества (проблемы искусственной социальности) в конечном счете могут быть решены только при выходе за существующие дисциплинарные границы. Вместе с тем сегодня эти проблемы, как мы увидим, разрабатываются в рамках конкретных дисциплин и междисциплинарных направлений.



лее чистом виде структуру разногласий о данной проблеме, о самой ее постановке.

Теоретико-методологическим основанием для сопоставления и анализа различных позиций и положений о проблемах ИИ и ИС в рамках настоящего рассуждения выступает концепция парадигм общения, сформулированная одним из авторов настоящей статьи [Резаев, 1993]<sup>4</sup>. Функционируя на различных уровнях социальной реальности, феномен общения предполагает принципиально разную постановку исследовательских проблем. Парадигмы общения относятся, в широком смысле, к способам исследования этих проблем в социальной философии:

1) *Информационно-инструментальная парадигма* ставит проблеме общения в терминах взаимосвязи – взаимодействия и обмена образованиями материального и духовного порядка – способностями, деятельностью, информацией, мыслями, чувствами.

2) *Социально-институциональная, или марксистская, парадигма* (ее теоретико-методологические основания оформились в рамках историко-материалистической традиции) рассматривает общение как персонификацию, актуализацию и реализацию структурной характеристики общества – общественных отношений.

3) *Экзистенциально-феноменологическая парадигма* характеризует общение как реализацию внутренней потенции самосознающего индивида.

Три парадигмы общения в социальной философии – это три принципиально разных ответа на вопрос, что такое человек как социальное, общающееся существо. Причем эти ответы вовсе не обязательно предполагают некоторое эмпирическое измерение, проверку (так, экзистенциально-феноменологическая парадигма отвергает саму возможность такой «проверки»).

В отличие от философского, социологический анализ общения структурируется не парадигмами, а дилеммами (вопросами), которые характеризуют разные решения при построении теоретических моделей [Резаев, Трегубова, 2017]. Различные социологические теории включают элементы разных парадигм общения; тем не менее не возникает универсальной социологической теории общения. Разные теоретики предлагают разные термины и концептуальные модели для характеристики различных сторон общения. Эти модели харак-

<sup>4</sup> Как отмечал Ю.В. Перов, один из признанных лидеров ленинградской-петербургской школы в отечественной философии, концепция парадигм общения «стала наиболее основательным из отечественных исследований социально-философских теорий общения... и поныне сохранила это свое значение, несмотря на то, что за прошедшее время появилось немало отечественных публикаций, существенно расширяющих и углубляющих представления о разработке проблематики общения в зарубежной философии» [Перов, 2004, с. 10].



теризуются выбором разных сторон в дилеммах – взаимоисключающих ответах на вопросы об общении. В результате возникают эклектичные (с точки зрения социальной философии), но гораздо более привязанные к эмпирическому анализу теории и концепции.

Данная структура знания – с одной стороны, парадигмы философского анализа, с другой – вопросы, характеризующие концепции в эмпирических науках, – будет далее наложена на дискуссии об ИИ и ИС.

Приведем два принципиальных аргумента, обуславливающих соотношение парадигмальной структуры феноменов общения с логикой развития ИИ.

Во-первых, различные парадигмы общения характеризуют разные взгляды на человека как общающееся существо. От понимания человека, как мы надеемся показать, зависит и взгляд на ИИ как на проект воспроизведения человеческих способностей.

Во-вторых, концепция парадигм общения помогает в социальной аналитике взаимозависимости «человек-алгоритм», которая становится все более актуальной и насущной задачей [Резаев, Трегубова, 2023]. Здесь возникают принципиальные затруднения: не ясно, что принимать за базовые элементы взаимозависимости, по каким основаниям выделять ее факторы и механизмы. Соответственно, логично для начала сформулировать более простую задачу – анализ взаимодействия между человеком и ИИ. Следующий шаг, который часто предлагают, – рассмотреть эти взаимодействия как последовательность коммуникаций (информационных и материальных обменов). Однако здесь и кроется проблема: взаимодействия между человеком и ИИ по их социальному, психологическому, этическому последствиям принципиально несводимы к коммуникациям. Именно здесь возникает эвристическая ценность типологии парадигм общения, которая позволяет зафиксировать разные перспективы на процесс общения людей и по аналогии на процесс взаимодействия человека с ИИ.

## **Проблема ИИ в советской философии**

Каким образом проблематика «искусственного интеллекта» актуализировалась в советской философской и научной мысли? В СССР то, что на английском языке получило название *artificial intelligence* и стало отдельным исследовательским направлением, развивалось под именем «кибернетики» [Kirtchik, 2023]. Проект ИИ изначально вдохновлялся стремлением воссоздать и превзойти человеческий разум, в то же время в Советском Союзе был альтернативный проект – построения коммунизма и воспитания нового советского человека. Поэтому для советских философов проблема «умных



машин»<sup>5</sup> возникала в связи с вопросом о том, что такое человек и какую роль ИИ может играть в жизни людей при капитализме и коммунизме.

В рамках настоящего рассуждения мы обратимся к анализу одной, наиболее яркой дискуссии (точнее, серии дискуссий) – между Э.В. Ильенковым и Д.И. Дубровским. Предметом дебатов между этими авторами становились разные вопросы: что определяет возникновение и развитие психики человека, что есть норма, патология и гениальность, что есть идеальное как философская категория. Отдельно следует выделить споры по поводу результатов Загорского эксперимента по обучению слепоглухонемых детей. Здесь нас будет интересовать не анализ самих дискуссий (тем более что эта задача уже предпринималась, и неоднократно), а то, какое место в аргументации каждого из авторов занимает проблематика «умных машин», что они имеют сказать о проблемах ИИ.

Проблематика ИИ в аргументах Э.В. Ильенкова и Д.И. Дубровского возникает как побочная: сравнение человека с «умной машиной» позволяет лучше высветить специфику самого человека и его место в этом мире. При этом Д.И. Дубровского, как мы увидим, следует отнести к информационно-инструментальной парадигме общения, Э.В. Ильенкова – к марксистской (социально-институциональной): оба автора мало пишут собственно об общении, однако они разделяют базовые предпосылки соответствующих парадигм о человеке. Отметим также: собственно об ИИ Ильенков и Дубровский не дискутируют, хотя они занимают разные позиции по поводу его принципиальной возможности. И различие это связано именно с разным представлением о человеке – а об этом авторы как раз спорят.

Как в аргументации каждого из авторов актуализируется проблематика ИИ?

Э.В. Ильенков – одна из наиболее ярких фигур в советской философии, оригинальный марксистский философ, полемизировавший с официальным марксизмом, человек, оказавший значительное влияние на своих современников. Проблематика ИИ была для него побочной, однако и в ней он сумел сформулировать тезисы, которые представляют интерес и сегодня. В своих работах, затрагивающих проблематику ИИ [Арсеньев и др., 1966; Ильенков, 1968]<sup>6</sup>,

<sup>5</sup> Следует отметить, что терминология советскими авторами используется разная – искусственный интеллект, машина, компьютер и др.

<sup>6</sup> В 1966 г. Э.В. Ильенков в соавторстве со своими коллегами-психологами А.С. Арсеньевым и В.В. Давыдовым публикует статью, основные положения которой затем раскрываются в его монографии 1968 г. «Об идолах и идеалах». Соавторы статьи разделяли взгляд на человека в рамках теории деятельности, который определяет их интерес к отличиям между человеком и машиной. Ильенков, как представляется, формулирует собственно философские аргументы, на которых



Э.В. Ильенков полемизирует с тем взглядом на проект ИИ, который был характерен для его основателей в США<sup>7</sup>. Взгляд этот (распространенный и сегодня) состоит в том, что люди могут/должны стремиться создать машину, которая будет умнее, эффективнее, работоспособнее их, которая воспроизведет и превзойдет человеческие способности. Ильенков критикует данное положение по двум основаниям: а) это невозможно, человек и машина принципиально отличны; б) современное распространение машин, восхищение и страх перед ними связаны с эксплуатацией человека человеком при капитализме.

Прежде всего Ильенков с соавторами проводят различие между мозгом и мышлением: «Мечтая о мыслящей машине, столь же, а может быть, и еще более совершенной, чем человек, многие кибернетики исходят из представлений, будто мыслит мозг. Поэтому им кажется, что достаточно построить модель мозга, чтобы получить и искусственное мышление. Увы, нет. Ибо мыслит не мозг, а человек с помощью мозга» [Арсеньев и др., 1966, с. 265]<sup>8</sup>. При этом «человек» понимается именно как социальное существо: «...для того, чтобы мозг отдельного индивида обрел способность мыслить, его обладатель должен быть с детства включен в систему общественно-человеческих отношений и развит в согласии с ее требованиями и нормами» [Ильенков, 1968, с. 304]. Поэтому, чтобы промоделировать мышление человека в машине, нужно, гипотетически, создать общество (цивилизацию) машин. А если машину включить в человеческое общество, придется сделать ее аналогичной человеку, иначе она будет ущербной, не будет способна использовать все достижения человеческой цивилизации.

Но чем именно компьютер отличается от человека? На данный вопрос можно предлагать разные ответы; ответ Ильенкова – логикой: у компьютера она формальная, у человека – диалектическая. Компьютер не переносит противоречий, поскольку основан на формальной логике, а для человека они – двигатель мышления: «...диалектическая логика... есть логика живого человеческого мышления, то есть способности отражать противоречия объективной реальности, выдерживать “напряжение противоречия”, находить им реальное, конкретно-содержательное разрешение... Машина же, построенная

---

мы и остановимся ниже, в то время как его коллеги-психологи обеспечивали для этих аргументов эмпирическую базу.

<sup>7</sup> Именно с таким пониманием ИИ полемизировал Х. Дрейфус в США, как мы увидим далее. В СССР подобное понимание ИИ не было доминирующим, но в каком-то виде существовало. Критика Ильенкова – это и критика западных реалий, и критика возможного их переноса на советскую почву.

<sup>8</sup> Следует заметить, что это положение разделялось многими советскими философами и психологами – однако далеко не всеми, как показывает дискуссия с Д.И. Дубровским.



по схеме математической логики, противоречий не любит, не выносит. Они разрушают схему ее работы» [Ильенков, 1968, с. 288–289]. Для Ильенкова способность осмыслять и разрешать противоречия и включенность человека в общественные отношения – это две стороны человеческой природы, понимаемой в категориях материалистической философии К. Маркса и Ф. Энгельса. Человек, мыслящий диалектически, вместе с себе подобными создает человеческое общество; как следствие, человек универсален. Машина же выполняет частные, специализированные функции.

Итак, машина и человек принципиально различны. Почему в современном обществе возникает стремление создать машины «умнее человека» и страх перед такими машинами? Аргумент Ильенкова состоит в том, что при капитализме человек *уже* превратился в раба машины – машины капитализма. При капитализме разделение труда приводит к такой специализации, что машины действительно в состоянии заменить многих людей. Поэтому страхи и энтузиазм вокруг проекта ИИ – не что иное, как кривое зеркало, в котором отражается капиталистическая действительность, маскирующее ее проблемы. Решение их, согласно Ильенкову, состоит в том, что сегодня именуют «человеко-ориентированным искусственным интеллектом» (human-centered AI): «...острейшая современная проблема, необходимость решения которой непосредственно чувствует каждый живой человек на земле, заключается вовсе не в том, чтобы поскорее сотворить еще одну машину “умнее и сильнее человека”, а в том, чтобы живого человека поскорее сделать умнее, сильнее и предусмотрительнее, нежели вышедшая из-под его контроля машинерия, современная громада производительных сил капиталистической индустрии» [Там же, с. 312]. Причем благо человека здесь понимается как полное развитие его (отдельного человека и всех людей) способностей при коммунизме.

Перед тем, как пойти далее, хотелось бы высказать несколько комментариев по поводу данной концепции.

Во-первых, она оригинальна. Многие философы и ученые, критикуя проект ИИ, выделяли различия между человеком и компьютером. Однако, насколько нам известно, именно Ильенков сформулировал его как противопоставление диалектической и формальной логики. Данное противопоставление имеет некоторое сходство с суждением о том, что человек, в отличие от компьютера, действует ситуативно, которое формулировалось в философии [Дрейфус, 1978] и в социальных науках [Wolfe, 1993; Сачмен, 2019]. Однако там данное суждение было сформулировано на основании других – экзистенциальных и феноменологических концепций человека. Для Ильенкова же, основывавшегося на марксистской философии, свойства мышления человека определяются тем, что он существо социальное, причем не столько в смысле использования языка, сколько в смысле совместной трудовой





деятельности. Что касается рассуждений Ильенкова о машинах при капитализме, сходные идеи были высказаны Н. Винером [1983], однако Ильенков развивает их более систематично.

Во-вторых, Ильенков в своей критике капитализма ставит проблему взаимозависимости «человек-машина». Он показывает: место, которое компьютеры и роботы занимают в нашей жизни, зависит от того, как организовано общество, – а организация общества, в свою очередь, зависит от той или иной реализации возможностей, возникающих вместе с научно-техническим развитием. Вместе с тем Ильенков не видит, что инструменты ИИ могут внести нечто принципиально новое в «машину капитализма», – он не рассматривает их как агентов, меняющих общественные отношения, превращающих их во что-то иное.

В-третьих, можно говорить о намеренном антропоцентризме Ильенкова: он отмечает, что кибернетика видит в человеке только сходное с машиной и потому не может правильно поставить вопрос о взаимоотношениях машины и человека. Вместе с тем антропоцентризм не мешает автору видеть цепочки взаимосвязей между людьми и вещами: мышление для него есть функция от системы людей и вещей, связанных общественными отношениями.

Наконец, человек для Ильенкова, как и для Маркса, – это родовой человек: каждый получает свое существование от совокупности отношений с другими людьми. Такое представление предполагает, что при коммунизме возможна полная реализация человеческих способностей, и религия/идеология как «кривое зеркало» реальности будет просто не нужна. Однако человек зачастую склонен одушевлять, а затем обожествлять/демонизировать, превращать в идола нечто просто потому, что он человек. Если это так, то проблема одушевления ИИ [Turkle, 2005] и склонность видеть в нем сакральный объект [Alexander, 1990] не являются только следствием развития «ложного сознания» при капитализме.

Переходя к анализу позиции Д.И. Дубровского, следует отметить, что он был одним из первых в Советском Союзе, кто начал заниматься анализом психофизиологической проблемы – проблемы взаимосвязи между мозгом/телом и сознанием/психикой человека. Д.И. Дубровский – автор информационной теории сознания, которая формулируется им в многочисленных статьях и монографиях, начиная с 1960-х гг. Его дискуссии с Э.В. Ильенковым касались фундаментального вопроса: каковы условия и механизмы мышления человека? Мышление – индивидуально или социально? Биологически или культурно детерминировано? И если Ильенков отвечает на этот вопрос, исходя из Марксова представления о человеке как ансамбле общественных отношений, то для Дубровского ответ заключатся в понимании работы индивидуального мозга. И здесь автор использует категорию информации, чтобы объяснить, как физические процессы,



происходящие в мозге, соотносятся с существованием психики, субъективной реальности (СР), как возможны сознание и свободная воля.

Для Дубровского проблематика ИИ возникает прежде всего в связи с исследованием психофизиологической проблемы. Автор замечает: «У компьютера нет субъективной реальности. Информационный процесс, лишенный качества СР, отличается по своей организации, по своим структурным, оперативным и целевым характеристикам от того информационного процесса, который специфичен для эго-системы головного мозга» [Дубровский, 2007]. Проблема ИИ предстает для него как проблема воспроизведения информационных процессов, характерных именно для человека. Возможно ли это? Согласно Дубровскому, теоретически да, но вовсе не обязательно теми методами развития вычислительной техники, которые используются сейчас.

Одним из направлений критики технологий ИИ с этих позиций является проблема важности физического воплощения для формирования мышления. Эта проблема – в связи с человеком – обсуждалась в полемике Ильенкова и Дубровского, в частности в критике вторым интерпретации результатов Загорского эксперимента. Вопрос состоит в том, возможно ли сформировать человека «с чистого листа», включив его в общественные отношения (позиция Ильенкова), или в формировании психики в целом и отдельных способностей человека в частности существенную роль играют биологические характеристики (позиция Дубровского). В отношении ИИ подобный вопрос связан с тем, возможно ли создать ИИ, включив его в социальные связи (обучив языку, например), или необходимо телесное воплощение, сходное с человеческим<sup>9</sup>. И здесь позиция Дубровского и его последователей состоит в том, что телесное воплощение имеет решающее значение: чтобы создать ИИ, который может то же, что и человек, нужно дать такому ИИ возможность действовать в физическом пространстве [Ефимов и др., 2023].

Позиция Д.И. Дубровского в основных ее положениях соотносится с тем, что пишут об ИИ его коллеги-когнитивисты (см., например, [Boden, 2016]): проблема рассматривается в категориях информационных процессов, принципиальная возможность ИИ не отрицается, для решения проблемы ИИ предлагается проводить больше когнитивных исследований.

Мы представили читателю суждения двух выдающихся советских философов относительно проблемы ИИ. Их позиции – противоречат или дополняют друг друга? На наш взгляд, они несоизмеримы, так как основаны на принципиально разных положениях о природе человека. Вместе с тем оба философа критикуют возможность

<sup>9</sup> Дискуссию в очень похожих терминах, как мы увидим, вели Х. Дрейфус и Г. Коллинз.



создания ИИ современными методами, только за счет моделирования мозга человека: без тела (Дубровский) и без «тела культуры» (Ильенков).

Если Э.В. Ильенков и Д.И. Дубровский представляют две из трех парадигм общения, возникает вопрос: были ли в советской философии представители экзистенциально-феноменологической парадигмы, которые писали об «умных машинах»? Насколько известно авторам настоящей статьи, в СССР не было философа-экзистенциалиста, который бы специально интересовался проблемами ИИ – в отличие от США, где такой фигурой является Х. Дрейфус. Вместе с тем можно обратиться к идеям Г.С. Батищева, ученика Э.В. Ильенкова. Г.С. Батищев начинал анализ проблем вхождения машин в жизнь людей с тех же позиций, что и его учитель: основное различие между человеком и компьютером он видел в том, что первый мыслит, осмысляя и разрешая противоречия, второй же противоречия не переносит [Батищев, 1963]. В более поздних работах философ анализирует экзистенциальное измерение жизни человека, и его анализ проясняет некоторые моменты вхождения технологий ИИ в жизнь общества.

Батищев дополняет аргумент Ильенкова о том, что при капитализме человек становится «винтиком» общественной машины, следующими рассуждениями: «Из процесса опредмечивания выходят наружу в качестве массовой конечной продукции вещи мертвые, бесплодные и холодные, такие, в которых *никто* не стремился дать себе самому живое продолжение и в которые *никто* не вложил себя самого, свою адресованную всем другим обшительную сущность, свою щедрую душу, свою неугасимую смысловую энергию. Так труд-средство порождает колоссальные нагромождения бездушно и бездуховно выполненных *вещей-результатов*, каждая из которых хотя и вполне *социальна*, ибо произведена по правилам общественного производства, но тем не менее в глубоком смысле *бессубъектна*» [Батищев, 1987, с. 127]. Эти рассуждения помогают ответить на вопрос: почему ИИ сегодня во многих случаях может заменить человека-работника? Дело в том, что сегодня человек стал подобен машине не только из-за узкой специализации, но и из-за того, что ценностные ориентации людей предполагают особое отношение к труду. Машина становится творцом (или со-творцом) произведения, потому что сам смысл творчества в конкретном разделении труда таков, что это допускает.

Еще одно положение Батищева помогает осмыслить характер взаимозависимости, возникающей между человеком и ИИ. Он выделяет три уровня проблем, стоящих перед человеком: проблемы с достаточной логикой (нам известно, как их решить), с недостаточной логикой (нужны творческие усилия, чтобы понять или придумать решение), с недостаточным субъектом (для того, чтобы их решить,



наших усилий – как мы есть сейчас – недостаточно) [Батищев, 2015]. Применяя данную типологию к работе ИИ, можно сказать, что первые проблемы ИИ решить может, в случае вторых ИИ будет инструментом в их решении человеком. Третьи же – в принципе не в компетенции ИИ. Проблема того, как использовать ИИ во благо человека и человечества, – это не проблема первого уровня: для ее решения нужны как минимум творческие усилия. Получается, что ИИ не может решить проблемы, которые сам порождает, – хотя, добавим, сами разработчики будут стремиться именно к такому решению.

По результатам рассмотренных положений можно сделать предварительные обобщения о том, как в рамках каждой из парадигм общения рассматривается проблема ИИ.

*В рамках информационно-инструментальной парадигмы* мозг человека – это устройство по обработке информации. Вопрос о том, рядоположен ли естественный и искусственный интеллект, решается исходя из того, какого именно представления о мозге как информационной машине придерживаются представители парадигмы. Принципиально, что здесь нет заранее предполагаемых (постулируемых) различий между мозгом и ИИ.

*В рамках социально-институциональной парадигмы* инструменты ИИ понимаются как новый вид машин, а машины (техника) – как продукт развития человеческого общества. Предполагается, что человек изначально является социальным существом, способным к производству и воспроизводству. Место машин в обществе – это их место в общественном разделении труда. И здесь возникают две проблемы: а) как «вписать» машины в человеческое общество (проблема согласования ценностей – value alignment problem), б) до какой степени можно имитировать человеческую социальность с помощью технологий ИИ.

*В рамках экзистенциально-феноменологической парадигмы* отличие ИИ от человека представляется самоочевидным: человек живой; человек смертен и знает об этом; человеку всегда чего-то не хватает; существование человека нельзя полностью зафиксировать и измерить методами эмпирических наук. То, что мы можем сказать об ИИ с позиций данной парадигмы, характеризуется двойной ущербностью: 1) ИИ ущербен в сравнении с человеком как неизмеримо более сложным существом; 2) в обществе, где ИИ может заменить человека во многих сферах, сам человек ведет ущербное существование.



## Дискуссии о проекте ИИ по другую сторону «железного занавеса»

Обратимся теперь к двум англоязычным дискуссиям о проекте ИИ. Одна из них касается полемики Хьюберта Дрейфуса с представителями компьютерных наук, которую он вел, начиная с 1960-х гг. [Дрейфус, 1978; Dreyfus, 1965; Papert, 1968; McCarthy, 1996; Dreyfus, 1992; Dreyfus, 1996; Dreyfus, 2012; см. также Астахов, 2020]. Другая – это дебаты между Хьюбертом Дрейфусом и Гарри Коллинзом – одним из немногих представителей STS, кто всерьез рассматривает проблему ИИ [Dreyfus, 1996; Collins, 1996; Selinger et al., 2007; Collins, 2018].

В обеих дискуссиях одной из сторон выступает Хьюберт Дрейфус – значимая фигура в философской критике проекта ИИ. Будучи академическим философом, Дрейфус столкнулся с проектом ИИ в самом начале его реализации и обнаружил, что создатели ИИ основывались, часто сами того не сознавая, на вполне определенном наборе идей о человеке и человеческом разуме (идеи Гоббса, Декарта, Канта, Фреге, Рассела) [Дрейфус, 1978; Dreyfus, 2012]. В своем анализе технологий ИИ Дрейфус основывается на философии Хайдеггера, Мерло-Понти и Витгенштейна. С философских позиций, которые развивал и защищал сам Дрейфус, идеи разработчиков ИИ о том, как мыслит и существует человек, выглядели нереалистичными упрощениями.

Содержание первой дискуссии – между Дрейфусом и его оппонентами из сферы компьютерных наук – сводится к критике философом невозможности воспроизвести целый ряд человеческих способностей с помощью компьютерных технологий. Его оппоненты обычно указывали, что искомые способности (выиграть в шахматы у гроссмейстера, перевести с одного языка на другой и т.д.) воспроизвести удалось или вскоре удастся. Однако для Дрейфуса важна не принципиальная неспособность достичь с помощью ИИ некоторых задач – важна *неспособность их решить так, как это делает человек*. Важно различие между универсальными алгоритмами, основанными на математике, и свойствами мышления и переживания человека, существующего в конкретном месте в конкретное время. Мозг не сводится к вычислительной машине, а использование языка – к манипуляции символами, и телесное воплощение является необходимым условием разума, подобного человеческому.

Сравнивая дискуссии, проходившие по разные стороны «железного занавеса», можно видеть, что критика Дрейфусом проекта ИИ в США находит явную параллель в полемике Ильенкова против кибернетиков в СССР. Оба философа критикуют не развитие компьютерных технологий как таковое, а необоснованные – с их точки



зрения – претензии на воспроизведение человеческого разума. Однако необоснованность претензий обосновывается ими по-разному: Дрейфус за точку отсчета принимает положение в мире отдельного человека, Ильенков – существование человеческого общества. Оба философа осуществляют критику представления о человеке как машине по переработке информации, основываясь один – на экзистенциально-феноменологической, другой – на социально-институциональной парадигме. При этом у каждой позиции есть свои ограничения. Ильенков, в пределе, обезличивает человека, не видит экзистенциального измерения индивидуального (уникального) существования. Дрейфус, в свою очередь, не видит общественных отношений, в которые включаются машины, – для него важна лишь разница между человеком и вычислительной машиной.

Вторая дискуссия связана именно со «слепотой» Дрейфуса в отношении социальных взаимодействий, в которые встраивается ИИ. Его оппонент Гарри Коллинз формулирует следующий тезис: чтобы создать ИИ, аналогичный человеческому, нужно не его телесное воплощение, а встраивание ИИ в систему социальных (речевых) взаимодействий: *embedded, not embodied AI*. Позиция Коллинза предполагает, что если ИИ, аналогичный человеческому, в принципе возможен, то он может быть создан на основе включения ИИ в разговоры. В языке запечатлен опыт человеческого существования, поэтому обучение использованию языка в разных ситуациях, согласно Коллинзу, равносильно обучению тому, как жить в мире людей. Ответ Дрейфуса на критику Коллинза состоит в том, что телесное воплощение как важнейшая характеристика существования человека является условием формирования человеческого разума, включая способность к речевому взаимодействию.

Особый интерес в рамках данной дискуссии для нас представляет обсуждение случая Мадлены – слепой женщины, страдавшей церебральным параличом. Данный случай был описан психологом Оливером Саксом [2010]. Мадлена уже в пожилом возрасте обрела способность действовать руками. Когда ей давали предметы, которых она не могла видеть и ранее не имела возможности ощупать, она была способна догадаться по форме предмета, как этот предмет называется. Данный случай стал предметом дебатов между Г. Коллинзом, с одной стороны, и Х. Дрейфусом и Э. Сэлинджером, с другой [Selinger et al., 2007]. Коллинз рассматривает случай Мадлены как подтверждение тезиса о том, что использование языка может стать заменой телесному опыту: Мадлена не могла видеть или осязать предметы, но участвовала в беседах с членами семьи и слушала книги. Значит, и ИИ может получать знания об окружающем мире, включаясь в разговоры людей. Контрагентом философов стало указание на то, что телесный опыт Мадлены, хотя и весьма ограниченный, – это опыт человека: у нее



несоизмеримо больше общего с любым другим человеком, чем с любым агентом ИИ.

Дискуссию о случае Мадлены можно сопоставить с обсуждением результатов Загорского эксперимента. В обоих случаях речь идет об анализе условий существования людей с ограниченными способностями и проведении аналогий с функционированием ИИ. Для Коллинза, как и для Ильенкова, возможность таких людей встроиться в общество демонстрирует социальную обусловленность человеческого мышления. Однако социальная обусловленность понимается Коллинзом не с позиций марксистской философии, а как включенность в конкретные ситуации взаимодействия. Интересно, что Коллинз, как и Ильенков, рассуждает о возможном обществе/цивилизации машин. Однако для Ильенкова человек как социальное существо радикальным образом отличается от машины, поэтому советский философ не предлагает встраивать машины в человеческое общество. Коллинз, напротив, не находит принципиальных аргументов против того, чтобы машина могла стать социальной. Вместе с тем он настаивает, что нынешние технологии ИИ имитируют человеческую социальность весьма несовершенным образом.

На другой стороне обеих дискуссий, Дубровский и Дрейфус утверждают, что телесное воплощение людей играет существенную роль в том, как они мыслят. Однако, опять-таки, на разных основаниях: Дубровский, опираясь на когнитивные исследования, подчеркивает разницу в человеческом восприятии; Дрейфус, приводя философские аргументы, указывает на предельное сходство в опыте всех людей. Как следствие, для Дубровского рассмотрение Загорского эксперимента демонстрирует лишь ограниченность современных подходов к созданию ИИ. Для Дрейфуса же случай Мадлены иллюстрирует невозможность создания ИИ, подобного человеку.

Разница в позициях и аргументах Дрейфуса и Дубровского – это различия между представителями экзистенциально-феноменологической и информационно-инструментальной парадигм. Различия между Ильенковым и Коллинзом, однако, определяются более сложным образом. Гарри Коллинза оказывается трудно отнести к одной парадигме. С социально-институциональной парадигмой его сближает интерес к социальным взаимодействиям как способу реализации общественных отношений, с экзистенциально-феноменологической – интерес к возможности человека уникальным образом действовать в конкретных ситуациях, с информационно-инструментальной – допущение о том, что проект ИИ в принципе может быть реализован. В целом же в работах Коллинза нет «большой» теории: большее внимание автор уделяет тому, как работают конкретные технологии ИИ в конкретных ситуациях и в чем эти технологии пока «не дотягивают» до человека. Таким образом, отличие Коллинза от Ильенкова – это отличие социального ученого,



сочетающего элементы разных парадигм, от философа, рассуждающего о проблеме ИИ на основании принятых им парадигмальных оснований.

## **Два с половиной этажа в здании социальной аналитики ИИ и искусственной социальности**

Итак, в первом приближении, для компьютерных и когнитивных наук характерна информационно-инструментальная парадигма как основание для понимания человека. Именно она стала основой для проекта ИИ. Некоторые философы также разделяют ее положения. Вместе с тем философские критики проекта ИИ выступают с позиций других парадигм: для них человек принципиально отличен от вычислительной машины. Для социальных ученых характерна иная структура теоретизирования: с одной стороны, интерес к более конкретным (частным) проблемам вхождения ИИ в жизнь общества, с другой – сочетание идей подходов из разных парадигм.

Рассмотрим положения отдельных теорий, которые позволят дополнить данную картину.

В рамках *информационно-инструментальной парадигмы* возможны разные суждения о проекте ИИ. Если мышление – информационный процесс, это еще не значит, что мышление – процесс вычислительный [Дубровский, 2017]. Если мозг и является машиной для вычислений, то вовсе не обязательно, что мы сумеем ее воспроизвести [Boden, 2016]. Вместе с тем экспоненциальное развитие современных технологий позволяет на это надеяться [Kurzweil, 2005]. В социальных науках в рамках данной парадигмы особый интерес представляет концепция максимизации эмоциональной энергии Р. Коллинза: предлагается научить ИИ максимизировать эмоциональную энергию [Коллинз, 2004] и включаться в социальные взаимодействия по принципам «эмоционального утилитаризма» [Резаев, Трегубова, 2022]. Вместе с тем эта социологическая концепция синтезирует отдельные положения трех парадигм общения (детальный анализ см. в [Резаев, Трегубова, 2017]).

*Социально-институциональную парадигму* в чистом виде представляет Э.В. Ильенков. В социологии в рамках концепции искусственной коммуникации Е. Эспозито [Esposito, 2017; Esposito, 2022], опираясь на теорию коммуникации Н. Лумана, анализирует, как агенты ИИ становятся социальными агентами, приобретая способность создавать осмысленную (с точки зрения людей) информацию. Концепция Эспозито, как и теория Лумана, сочетает элементы всех трех парадигм: интерес к информации, опору на феноменологический анализ проблемы смысла, взгляд на общество как на систему,





состоящую из коммуникаций. Отчасти сходную позицию занимает Г. Коллинз, для которого, как мы отмечали, телесное воплощение не является обязательным – достаточно быть способным включиться в разговоры. Также следует упомянуть концепцию надзорного капитализма Ш. Зубофф [2022], в которой предлагается анализ новых отношений эксплуатации с помощью компьютерных алгоритмов.

Наконец, в развитии положений *экзистенциально-феноменологической* парадигмы также могут быть выделены несколько направлений. В рамках постмодернистской философии предпринимаются попытки переосмыслить сами границы человеческого существования, упразднить различия между человеком и машиной [Haraway, 2006]. В рамках собственно социальных наук вхождение технологий ИИ в жизнь общества, задействуя элементы экзистенциальной или феноменологической философии, осмысляют Ш. Теркл [Turkle, 2005], Л. Сачмен [Сачмен, 2019], Дж. Александер [Alexander, 1990].

Рассматривая картину исследований в целом, можно выделить два с половиной «этажа» (уровня) в исследованиях проблем ИИ:

- На первом этаже находится так называемая «большая тройка»: философия, компьютерные науки, когнитивные науки. Эти дисциплины с самого начала проекта ИИ – с середины XX в. – интересовались данной проблематикой. Здесь, соотнося естественный и искусственный интеллект, исследователи основываются на представлениях о человеке, соответствующих одной из парадигм общения.
- На втором этаже располагаются социальные науки, где в рамках отдельных теорий исследователи сочетают положения разных парадигм. Если на первом уровне происходит преимущественно анализ проекта ИИ, то на втором уровне – стремятся ответить на вопрос: что происходит, когда технологии ИИ входят в повседневную жизнь общества?
- Наконец, в «мансарде» располагаются исследования в рамках STS и communication studies. В них авторы не рассматривают искусственный интеллект как нечто, принципиально отличное от технологий, существовавших до него. Здесь проследить связь с парадигмами общения наиболее проблематично.

Таким образом, представители философии, компьютерных и когнитивных наук осмысляют проект ИИ как таковой. Социальные науки рассматривают реализацию данного проекта и ее следствие – формирование искусственной социальности. Междисциплинарные направления STS и communication studies исследуют те же проблемы, но за отдельными примерами технологий ИИ они не видят целого – единого проекта воспроизводства и превосходства человеческих способностей с помощью вычислительных машин.



Проведенный анализ позволяет сформулировать *гипотезу*: хорошая теория искусственной социальности является: 1) эклектичной (сочетает элементы разных парадигм), 2) конкретной 3) и при этом принимает во внимание специфику технологий ИИ в сравнении с другими технологиями.

## **Вместо заключения: новые старые вопросы об искусственном интеллекте**

В завершение настоящего рассуждения сформулируем вопросы, которыми структурируется дискуссия о проблемах ИИ. Эти вопросы в явном или неявном виде были сформулированы еще на первом этапе обсуждения проекта ИИ. Однако сегодня они вновь и вновь встают в научных и популярных дискуссиях о том, какую роль технологии ИИ будут/могут/должны играть в жизни людей.

«Новые старые вопросы» подразделяются на вопросы об ИИ и вопросы об искусственной социальности. Вопросы об ИИ также можно назвать «парадигмальными», поскольку они определяют общий взгляд на возможность и характер реализации проекта ИИ. Ниже мы сформулируем их и приведем некоторые примеры ответов из числа тех теорий и концепций, которые обсуждались в настоящей статье.

Вопросы об искусственном интеллекте (парадигмальные вопросы):

1. *Кто мыслит: мозг или человек или общество?* Для Д.И. Дубровского и М. Боден – мозг. Для Х. Дрейфуса и Ш. Теркл – человек. Для Э.В. Ильенкова и Е. Эспозито – в некотором, причем разном, смысле – общество. Некоторые авторы в принципе не ставят этот вопрос.
2. *Рядоположен ли интеллект человека компьютерным вычислениям?* Данный вопрос делит авторов на противников и сторонников гипотезы о мозге как вычислительной машине (computational mind hypothesis). Наиболее яркий пример – полемика Х. Дрейфуса и Э.В. Ильенкова против представителей компьютерных наук.
3. *Рассматривается ли искусственный интеллект как отдельный феномен, представляющий особый интерес для исследователей?* Для почти всех упомянутых в данной статье авторов – да, для многих исследователей из STS – нет.
4. *Принципиально ли телесное воплощение людей для понимания искусственного интеллекта?* Для Е. Эспозито



и Г. Коллинза – а также, добавим, для А. Тьюринга (вспомним «тест Тьюринга») – нет, для Х. Дрейфуса и Д.И. Дубровского – да.

Вопросы об искусственной социальности:

5. *Существует ли проблема вхождения новых агентов ИИ в жизнь общества?* Для Е. Эспозито – да, для Э.В. Ильенкова – нет.
6. *Общество или индивид в фокусе внимания теории?* Для Е. Эспозито и Ш. Зубофф – общество, для Ш. Теркл и Р. Коллинза – индивид.
7. *ИИ – субъект, или псевдосубъект, или орудие человека?* Для Д. Харрауэй – субъект, для Ш. Теркл – псевдосубъект, для Э.В. Ильенкова – орудие.
8. *Почему мы склонны обожествлять/демонизировать машины?* Они действительно превосходят нас (Р. Курцвейл). Мы и так их рабы при капитализме (Э.В. Ильенков). Человек просто творит себе кумира из очередного объекта (Дж. Александер).

## Список литературы

Арсеньев и др., 1966 – *Арсеньев А.С., Ильенков Э.В., Давыдов В.В.* Машина и человек, кибернетика и философия // Ленинская теория отражения и современная наука / Под ред. К.Ф. Константинова. М.: Наука, 1966. С. 265–284.

Астахов 2020 – *Астахов С.* Феноменология против символического искусственного интеллекта: философия научения Хьюберта Дрейфуса // Логос. 2020. Т. 30. № 2. С. 157–193.

Батищев 1963 – *Батищев Г.С.* Противоречие как категория диалектической логики. М.: Высшая школа, 1963. 120 с.

Батищев, 1987 – *Батищев Г.С.* Социальные связи человека в культуре // Культура, человек и картина мира / Под ред. А.И. Арнольдова, В.А. Кругликова. М.: Наука, 1987. С. 90–135.

Батищев, 2015 – *Батищев Г.С.* Самопознание человека как культурно-созидательного существа: три уровня сложности задач // *Батищев Г.С.* Избранные произведения / Под общ. ред. З.К. Шаукеновой. Алматы: Институт философии, политологии и религиоведения КНМОН РК, 2015. С. 453–474.

Винер, 1983 – *Винер Н.* Кибернетика, или Управление и связь в животном и машине. 2-е изд. М.: Наука, 1983. 344 с.

Дрейфус, 1978 – *Дрейфус Х.* Чего не могут вычислительные машины: критика искусственного разума. М.: Прогресс, 1978. 334 с.

Дубровский, 2007 – *Дубровский Д.И.* Сознание, мозг, искусственный интеллект. М.: Стратегия-центр, 2007. 249 с.

Ефимов, Дубровский, Матвеев, 2023 – *Ефимов А.Р., Дубровский Д.И., Матвеев Ф.М.* Что мешает нам создать Общий искусственный интеллект?



Одна старая стена и один старый спор // Вопросы философии. 2023. № 5. С. 39–49.

Зубофф, 2022 – *Зубофф Ш.* Эпоха надзорного капитализма. Битва за человеческое будущее на новых рубежах власти. М.: Издательство Института Гайдара, 2022. 784 с.

Ильенков, 1968 – *Ильенков Э.В.* Об идолах и идеалах. М.: Политиздат, 1968. 319 с.

Коллинз, 2004 – *Коллинз Р.* Может ли социология создать искусственный разум? // *Бергер П., Бергер Б., Коллинз Р.* Личностно-ориентированная социология. М.: Академический проект, 2004. С. 566–598.

Перов, 2004 – *Перов Ю.В.* Стратегии философского осмысления социального общения // *Коммуникация и образование / Под ред. С.И. Дудника.* СПб.: Санкт-Петербургское философское общество, 2004. С. 9–32.

Резаев, 1993 – *Резаев А.В.* Парадигмы общения. Взгляд с позиций социальной философии. СПб.: Изд-во СПбГУ; Иваново: Полиформ, 1993. 212 с.

Резаев, Трегубова, 2017 – *Резаев А.В., Трегубова Н.Д.* Мир общения в социологическом измерении. М.: Университетская книга, 2017. 152 с.

Резаев, Трегубова, 2019 – *Резаев А.В., Трегубова Н.Д.* «Искусственный интеллект», «онлайн-культура», «искусственная социальность»: определение понятий // *Мониторинг общественного мнения: Экономические и социальные перемены.* 2019. № 6. С. 35–47.

Резаев, Трегубова, 2022 – *Резаев А.В., Трегубова Н.Д.* «Эмоциональный утилитаризм» и пределы развития искусственного интеллекта // *Мониторинг общественного мнения: экономические и социальные перемены.* 2022. № 2. С. 4–23.

Резаев, Трегубова, 2023 – *Резаев А.В., Трегубова Н.Д.* Взаимозависимость «человек-машина» за пределами искусственного интеллекта: случай биткойна // *Социологическое обозрение.* 2023. Т. 22. № 3. С. 263–286.

Сакс, 2010 – *Сакс О.* Человек, который принял жену за шляпу, и другие истории из врачебной практики. М.: АСТ, 2010. 318 с.

Сачмен, 2019 – *Сачмен Л.* Реконфигурации отношений человек-машина. М.: Элементарные формы, 2019. 488 с.

## References

Alexander, 1990 – Alexander, J. “The Sacred and Profane Information Machine: Discourse about the Computer as Ideology”, *Archives de sciences sociales des religions*, 1990, no. 69, pp. 161–171.

Arsen'ev, A., Il'enkov, E. & Davydov, V. “Mashina i chelovek, kibernetika i filosofiya” [Machine and Human: Cybernetics and Philosophy], in: F. Konstantinov (ed.) *Leninskaya teoriya otrazheniya i sovremennaya nauka* [Lenin's Theory of Reflection and Contemporary Science], Moscow: Nauka, 1966, pp. 265–283. (In Russian)

Astakhov, S. “Fenomenologiya protiv simvolicheskogo iskusstvennogo intellekta: filosofiya naucheniya Kh'yuberta Dreifusa” [Phenomenology versus Symbolic



Artificial Intelligence: The Philosophy of Learning of Hubert Dreyfus], *Logos*, 2020, vol. 30, no. 2, pp. 157–193. (In Russian)

Batishchev, G.S. “Sotsial’nye svyazi cheloveka v kul’ture” [Human Social Connections in Culture], in: A.I. Arnol’dov, V.A. Kruglikov (ed.) *Kul’tura, chelovek i kartina mira* [Culture, People and the Picture of the World], Moscow: Nauka, 1987, pp. 90–135. (In Russian)

Batishchev, G.S. “Samopoznanie cheloveka kak kul’turo-sozdatel’nogo sushchestva: tri urovnya slozhnosti zadach” [Self-knowledge of Human as a Culture-Creative Being: Three Levels of Task Complexity], in: Batishchev, G.S. *Selected works*. Almaty: Institut filosofii, politologii i religiovedeniya KNMON RK, 2015, pp. 453–474. (In Russian)

Batishchev, G.S. *Protivorechie kak kategoriya dialekticheskoi logiki* [Contradiction as a Category of Dialectical Logic]. Moscow: Vysshaya shkola, 1963. (In Russian)

Boden, 2016 – Boden, M. *AI: Its Nature and Future*. Oxford: Oxford University Press, 2016.

Collins, 1996 – Collins, H. “Embedded or Embodied? A Review of Hubert Dreyfus’ What Computers Still Can’t Do”, *Artificial Intelligence*, 1996, vol. 80, no. 1, pp. 99–117.

Collins, 2018 – Collins, H. *Artificial Intelligence: Against Humanity’s Surrender to Computers*. Madford, MA: Polity Press.

Collins, R. “Mozhet li sotsiologiya sozdat’ iskusstvennyi razum?” [Can Sociology Create an Artificial Intelligence?], in: Berger, P., Berger, B., Collins, R. *Lichnostno-orientirovannaya sotsiologiya* [Person-Oriented Sociology], trans. by V.F. Anurin. Moscow: Akademicheskii proekt, 2004, pp. 566–598. (In Russian)

Dreyfus, 1965 – Dreyfus, H.L. *Alchemy and Artificial Intelligence*. Rand Corp., Report No. P-3244, 1965.

Dreyfus, 1992 – Dreyfus, H. *What Computers Still Can’t Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press, 1992.

Dreyfus, 1996 – Dreyfus, H. “Response to My Critics”, *Artificial Intelligence*, 1996, vol. 80, no. 1, pp. 171–191.

Dreyfus, 2012 – Dreyfus, H. “A History of First Step Fallacies”, *Minds & Machines*, 2012, no. 22, pp. 87–99.

Dreyfus, H. *Chego ne mogut vychislitel’nye mashiny: kritika iskusstvennogo razuma* [What Computers Can’t Do: A Critique of Artificial Reason], trans. by N. Rodma. Moscow: Progress, 1978. (Trans. into Russian)

Dubrovskii, D.I. *Soznanie, mozg, iskusstvennyi intellekt* [Consciousness, Brain, Artificial Intelligence]. Moscow: Strategia-Tsent, 2007. (In Russian)

Efimov, A.R., Dubrovskii, D.I., & Matveev, F. “Chto meshaet nam sozdat’ obshchii iskusstvennyi intellekt? Odnaya staraya stena i odin staryi spor” [What Prevents Us from Creating Artificial General Intelligence? One Old Wall and One Old Dispute], *Voprosy Filosofii*, 2023, no. 5, pp. 39–49. (In Russian)

Esposito, 2017 – Esposito, E. “Artificial Communication? The Production of Contingency by Algorithms”, *Zeitschrift für Soziologie*, 2017, vol. 46, no. 4, pp. 249–265.

Esposito, 2022 – Esposito, E. *Artificial Communication: How Algorithms Produce Social Intelligence*. The MIT Press, 2022.



Haraway, 2006 – Haraway, D. “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20<sup>th</sup> Century”, in: Weiss, J., Nolan, J., Hunsinger, J., Trifonas, P. (eds.) *The International Handbook of Virtual Learning Environments*. Dordrecht: Springer, pp. 117–158.

Il'enkov, E.V. *Ob idolakh i idealakh* [On the Idols and the Ideals]. Moscow: Politizdat, 1968. (In Russian)

Kirtchik, 2023 – Kirtchik, O. “The Soviet Scientific Programme on AI: If a Machine Cannot ‘Think’, Can It ‘Control’?”, *BJHS Themes*, 2023, no. 8, pp. 111–125.

Kurzweil, 2005 – Kurzweil, R. *The Singularity is Near: When Human Transcends Biology*. Viking, 2005.

McCarthy, 1996 – McCarthy, J. “Hubert Dreyfus, What Computers Still Can't Do”, *Artificial Intelligence*, 1996, vol. 80, no. 1, pp. 143–150.

Papert, 1968 – Papert, S. *The Artificial Intelligence of Hubert L. Dreyfus: A Budget of Fallacies*. MIT AI Memo, No. 154, 1968.

Perov, Yu.V. “Strategii filosofskogo osmysleniya sotsial'nogo obshcheniya” [Strategies for philosophical comprehension of social communication], in: S.I. Dudnik (ed.) *Kommunikatsiya i obrazovanie* [Communication and Education]. St. Petersburg: Sankt-Peterburgskoe filosofskoe obshchestvo, 2004, pp. 9–32. (In Russian)

Rezaev, A.V., Tregubova, N.D. “Vzaimozavisimost' “chelovek-mashina” za predelami iskusstvennogo intellekta: sluchai bitkoina” [Human-machine Interdependence Beyond AI Development: The Case of Bitcoin], *Russian Sociological Review*, 2023, vol. 22, no. 3, pp. 263–286. (In Russian)

Rezaev, A.V., Tregubova, N.D. “Emotsional'nyi utilitarizm' i predely razvitiya iskusstvennogo intellekta” [“Emotional Utilitarianism” and the Frontiers of Artificial Intelligence Evolvement], *Monitoring of Public Opinion: Economic and Social Changes*, 2022, no. 2, pp. 4–23. (In Russian)

Rezaev, A.V., Tregubova, N.D. ““Iskusstvennyi intellekt”, “onlain-kul'tura”, “iskusstvennaya sotsial'nost'”: opredelenie ponyatii” [Artificial Intelligence, On-line Culture, Artificial Sociality: Definition of the Terms], *Monitoring of Public Opinion: Economic and Social Changes*, 2019, no. 6, pp. 35–47. (In Russian)

Rezaev, A.V., Tregubova, N.D. *Mir obshcheniya v sotsiologicheskoy izmerenii* [The World of Social Intercourse Through Sociological Lenses]. Moscow: Universitetskaya kniga, 2017, 152 pp. (In Russian)

Rezaev, A.V. *Paradigmy obshcheniya. Vzglyad s pozitsii sotsial'noi filosofii* [Social Intercourse Paradigms. Looking from Social Philosophy's Positions]. St. Petersburg: Izd-vo SPbGU; Ivanovo: Polinform, 1993, 212 pp. (In Russian)

Sacks, O. *Chelovek, kotoryi prinyal zhenu za shlyapu, i drugie istorii iz vrachebnoi praktiki* [The Man Who Mistook His Wife for a Hat, and Other Clinical Tales], trans. by G. Khasin, Yu. Chislenko. Moscow: AST, 2010. (Trans. into Russian)

Selinger, Dreyfus & Collins, 2007 – Selinger, E., Dreyfus, H. & Collins, H. “Interactional expertise and embodiment”, *Studies in History and Philosophy of Science*, 2007, no. 38, pp. 722–740.

Suchman, L. *Rekonfiguratsii otnoshenii chelovek-mashina* [Human – Machine Reconfigurations], trans. by A.S. Maksimova. Moscow: Elementarnye formy, 2019, 488 s. (Trans. into Russian)

Turkle, 2005 – Turkle, S. *The Second Self: Computers and the Human Spirit* (20<sup>th</sup> anniversary ed.). Cambridge, MA: MIT Press.



Wiener, N. *Kibernetika, ili upravlenie i svyaz' v zhivotnom i mashine* [Cybernetics: Or Control and Communication in the Animal and the Machine], trans. by I.V. Solov'ev G.N. Povarov. Moscow: Nauka; Glavnaya redaktsiya izdaniy dlya zarubezhnykh stran, 1983. (Trans. into Russian)

Wolfe, 1993 – Wolfe, A. *The Human Difference: Animals, Computers, and the Necessity of Social Science*. Berkley: University of California Press, 1993.

Zuboff, Sh. *Epokha nadzornogo kapitalizma. Bitva za chelovecheskoe budushchee na novykh rubezhakh vlasti* [The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power], trans. by A.F. Vasil'ev. Moscow: Izdatel'stvo Instituta Gaidara, 2022, 784 pp. (Trans. into Russian)