ФИЛОСОФСКИЕ ПРОБЛЕМЫ НАУЧНО-ТЕХНИЧЕСКОГО РАЗВИТИЯ

УДК 1:004.8 / 165.1
DOI 10.31249/scis/2026.03.03

ФИЛОСОФ И БОЛЬШАЯ ЯЗЫКОВАЯ МОДЕЛЬ В АРХИТЕКТУРЕ ПОСТНЕКЛАССИЧЕСКОЙ НАУКИ

Сабанина Наталия Рафаэлевна – соискатель, Институт философии РАН, Москва, Россия; myzeinatali@mail.ru; ORCID 0000-0002-3024-0668

Аннотация. Статья подготовлена в продолжение и развитие темы Круглого стола ИНИОН РАН «Наука в цифровую эпоху», посвященного памяти В.И. Аршинова (26 февраля 2026 г.). Обсуждаются проблемы и методологические основания практики, в которой большая языковая модель (БЯМ) выступает рабочим инструментом философского исследования и узлом гибридной социо-техно-антропосферы. Опираясь на доклады В.Г. Буданова, М.Ф. Януковича и В.В. Тарасенко, обзор современных исследований, связанных с влиянием ИИ на научное исследование, а также отечественную программу *машинной эпистемологии*, автор формулирует четыре методологические функции БЯМ, применимые в работе философа: топография концептуальных пространств, выявление «слепых пятен» аргументации, генерация контрфактических аксиоматик и распределенное чтение корпусов. Эти функции оказываются продуктивны благодаря структурным ограничениям БЯМ, прежде всего невозможности феноменологического эпохе и асимметрии интенциональной структуры: смысловой, нарративный (⊙-доминантный) режим человеческого мышления и вычислительный, геометрический (∇-доминантный) режим работы трансформера образуют двухполюсную конфигурацию граничного сопряжения. Рассматривается понятие гибридных ценностей, возникающих во взаимодействии человека и генеративной нейросети. Ценности в этом контексте образуют новую интенциональную структуру и определяются как инварианты интерпретации события на границе разнородных систем. Рассматривается концепт «гибридная научная литература», а также практики «корпусного чтения». Обсуждается технология «БЯМ-зрения», позволяющая формировать топографическую проекцию корпуса текста и адаптирующая его сложность к когнитивным возможностям читателя. В заключении предложен принцип «эпизодического доверия» для гибридных систем, основанный на ограничении саморепрезентации.

Ключевые слова: философия искусственного интеллекта; машинная эпистемология; цифровое эпохе; трансляционная интенциональность; интенциональная асимметрия; гибридные ценности; гибридная научная литература; корпусное чтение; философия сложностности В.И. Аршинова.

Для цитирования: Сабанина Н.Р. Философ и большая языковая модель в архитектуре постнеклассической науки // Науковедческие исследования. – 2026. – № 3. – С. 29–49. – DOI: http://www.doi.org/10.31249/scis/2026.03.03

Статья поступила в редакцию: 25.05.2026
Принята к публикации: 15.06.2026

PHILOSOPHER AND LARGE LANGUAGE MODEL IN THE ARCHITECTURE OF POSTNONCLASSICAL SCIENCE

Nataly R. Sabanina – applicant, Institute of Philosophy of the Russian Academy of Sciences, Moscow, Russia; myzeinatali@mail.ru; ORCID 0000-0002-3024-0668

Abstract. This article continues and develops the theme of the INION RAS Round Table "Science in the Digital Age," dedicated to the memory of V.I. Arshinov (February 26, 2026). The paper discusses problems and methodological foundations of a practice in which a large language model (LLM) serves as a working instrument of philosophical research and a node of the hybrid socio-techno-anthroposphere. Drawing on presentations by V.G. Budanov, M.F. Yanukovitch, and V.V. Tarasenko, a review of contemporary research on AI's impact on scientific inquiry, and the domestic program of machine epistemology, the author formulates four methodological functions of LLM applicable to philosophical work: topology of conceptual spaces, identification of "blind spots" in argumentation, generation of counterfactual axiomatics, and distributed corpus reading. These functions prove productive precisely because of LLM's structural constraints, primarily the impossibility of phenomenological epoche and the asymmetry of intentional structure: the semantic, arrative (⊙-dominant) mode of human thought and the computational, geometric (∇-dominant) mode of transformer operation form a bipolar configuration of boundary coupling. The concept of hybrid values emerging from human-generative neural network interaction is examined. Values in this context form a new intentional structure and are defined as invariants of event interpretation at the boundary of heterogeneous systems. The concept of "hybrid scientific literature" is considered, along with practices of "corpus reading." The technology of "LLM-vision" is discussed, enabling the formation of a topological projection of a text corpus and adapting its complexity to the reader's cognitive capacities. In conclusion the principle of "episodic trust" for hybrid systems is proposed, based on limiting self-representation.

Keywords: philosophy of artificial intelligence; machine epistemology; digital epoch; translational intentionality; intentional asymmetry; hybrid values; hybrid scientific literature; corpus reading; V.I. Arshinov's philosophy of complexity.

For citation: Sabanina, N.R. (2026) Philosopher and Large Language Model in the architecture of postnonclassical science. // *Naukovedcheskie issledovaniya = Science Studies*. – 2026. – N 3. – P. 29–49. – DOI: http://www.doi.org/10.31249/scis/2026.03.03

The paper was submitted: 25.05.2026
Accepted for publication: 15.06.2026

Введение: смена жанра философского труда в цифровую эпоху

Цифровая эпоха ИИ радикально изменила исследовательские методы: суперкомпьютеры, большие языковые модели и цифровые платформы создали гибридные формы научной жизни, а предмет исследования сместился от междисциплинарных взаимодействий привычных дисциплин к целостной цифровой социо-техно-антропосфере, в которой ИИ образует новый гибридный социум [Буданов, 2025, с. 228]. Что происходит с демаркацией

научного знания и типом рациональности [Буданов, 2023], как меняется роль наблюдателя и характер субъектности, чем определяется достоверность знания в архитектуре практик его производства в связи с массовым распространением технологий ИИ – эти вопросы находились в фокусе исследований В.И. Аршинова с коллегами и задали тематику Круглого стола ИНИОН, посвященного его памяти.

С появлением в 2022–2023 гг. общедоступных больших языковых моделей (БЯМ) – (ChatGPT, Claude, Gemini) практика философского исследования начала меняться быстрее, чем это могло быть осмыслено фундаментальной наукой. Уже к 2024 г. БЯМ обсуждали не только как предмет исследования (вопрос о машинном сознании, статус «стохастических попугаев» и т.д.), но и как значимую часть работы философа. Эксперимент с дообученной на корпусе Д. Деннета GPT-3 показал, что эксперты угадывают настоящий ответ философа из пяти вариантов лишь в 51% случаев, что выше случайного выбора (20%), но значительно ниже предсказанных 80% [Schwitzgebel, 2024]. Это задает нижнюю границу исследований: *если различие между текстами философа и БЯМ перестает быть очевидным, как должна измениться сама практика философского труда*?

Проведенный автором анализ литературных источников позволил сформулировать ключевой тезис статьи: *продуктивность связки «человек – БЯМ» обусловлена структурными отличиями генерации текстов БЯМ от человеческого мышления: отсутствием нарративного намерения и «геометрической» организацией текстогенерации, асимметрией интенциональной структуры гибридной системы, принципиальной неприменимостью к ней антропоморфного представления о феноменологической редукции (эпохе).* Каждое отличие превращается в методологическое преимущество (функционально), если оно отрефлексировано.

Опорную линию исследования дает также серия статей в Вестнике Пермского университета (2025, вып. 3) по теме «онтология нейросети» [Аршинов, 2025], [Киященко, 2025], фиксирующая опыт совместного мышления философов и БЯМ на семинаре, организованном В.И. Аршиновым, Л. Б. Жуковым, М. Ф. Януковичем «Философия + IT»[1]. В. И. Аршинов с соавторами описывает нейросеть как синергетическую систему, в которой из локальных операций различения эмерджентно возникают управляющие макропараметры, а симондоновский язык индивидуации позволяет представить ответ модели как «кристаллизацию» формы из «доиндивидуальной» среды латентных отношений [Аршинов, 2025, с. 321–323]. Ю.Н. Гарашко добавляет инженерный реализм: «модель мира» на уровне архитектуры есть результирующее состояние ансамбля элементарных «наблюдателей», каждый из которых выполняет простейшие функции различения и связывания в жизненном потоке [Аршинов, 2025,

[1] Семинар «Философия + IT: URL: https://youtube.com/playlist?list=PLBx8d8IABl0o_XMiAB7vVeMn2xIs-EDvZ&si=gK7cAPOIShzvmpib (дата доступа: 30.05.2026)

с. 323–324]. Этот контекст является содержательным описанием гибридной онтологии граничного познания, на которую опирается автор статьи.

Операторная структура наблюдения в гибридной системе

Развивая направление мышления вместе со сложностностностью в контексте массового распространения технологий ИИ, авторский анализ опирается на различение наблюдения первого и второго порядков, разработанное в духе квантово-синергетического подхода Аршинова–Свирского–Буданова. Понятие наблюдения второго порядка восходит к кибернетике второго порядка и получило классическое развитие в работах Н. Лумана, отвечая на вопрос: каким образом конструируется наблюдаемая реальность через операции различения и коммуникации. Согласно Луману, наблюдатель второго порядка раскрывает ограничения и «слепые пятна» наблюдателя первого порядка, делая предметом анализа сами схемы наблюдения.

Если у Н. Лумана акцент сделан преимущественно на логике различения и коммуникации, то для *темпорального наблюдателя сложностности* В.И. Аршинова фокус смещается к процессуальности, временной динамике и включенности наблюдателя в саморазвивающуюся среду. Такой наблюдатель обладает рядом характеристик, которые можно назвать квантовоподобными в эпистемологическом смысле. Он (наблюдатель) возникает вместе с наблюдаемым процессом, изменяясь в ходе взаимодействия. Наблюдатель оказывается динамической конфигурацией отношений между различением, средой и самим процессом наблюдения. Его наблюдения контекстуальны, зависят от выбранной точки входа и перспективы и способны рекурсивно влиять на последующие состояния системы. В деятельностном STOR-тетраэдре В.Г. Буданова такой точкой входа в познавательный цикл является вопрос, маркирующий неопределенность: «…все типы вопросов порождают свои типы рациональности, которые связаны с заданием ориентаций графа на деятельностном тетраэдре» [Буданов, 2026]. Наблюдатель Буданова не совпадает ни с одной вершиной: это активный «наблюдатель сложностности» помещенный в фокус OBS и связанный как с вершинами, так и с гранями тетраэдра.

В контексте настоящего исследования наблюдение первого порядка понимается как *отдельный акт смыслообразования на границе взаимодействия человека и большой языковой модели*. Для описания внутренней структуры такого акта используются три эвристических оператора. Набла (∇) указывает на различение – выделение границы, паттерна или значимого отличия. Интегра ($\int$) обозначает связывание различенного в относительно устойчивую конфигурацию. Оператор Значимость ($\odot$) выбран как указание на смысловой фокус внимания, обозначает маркирование значимости – выделение того, что приобретает ценностную,

смысловую или практическую релевантность для системы наблюдения. Данные обозначения используются как сокращенный язык описания процессов смыслообразования.

Смыслообразование в гибридной системе не завершается отдельным актом наблюдения. Серии подобных наблюдений образуют более устойчивые конфигурации, обнаруживаемые наблюдателем второго порядка, операции которого раскрываются далее по тексту. Ценности в них проявляются как инварианты множества интерпретаций. Данный процесс далее будет обозначаться как аксиогенез – *становление устойчивых ценностно-смысловых структур в ходе рекурсивного взаимодействия человека и БЯМ.* В этом смысле аксиогенез выступает интегративным механизмом гибридной системы: он связывает отдельные акты различения (∇), интеграции (∫) и маркирования значимости (⊙) в относительно устойчивые траектории смыслообразования.

Квантовоподобность – важнейшая характеристика операторной динамики – здесь методологическая. Квантовый формализм (некоммутирующие операторы, суперпозиция, интерференция) служит языком операционализации постнеклассических принципов: контекстуальности, процессуальности, человекомерности. Применительно к операторам некоммутативность ($[O_1, O_2] \neq 0$) – это формальное выражение контекстуальной зависимости, для которой порядок различения, связывания и маркирования значимости меняет результат, подобно тому, как порядок вопросов меняет распределение ответов в эмпирически подтвержденных эффектах порядка [Wang et al., 2014]. Глубинная связь обнаруживается также на уровне логической формы. По теореме В.С. Меськова, если в тройке объектов хотя бы два имеют общую границу, рассуждения о них теряют дистрибутивность, а адекватной логической моделью становится недистрибутивная (ортомодулярная) решетка [Меськов, 2012]. Постнеклассическая тройка «субъект–средства–объект» – именно такова: субъект и средства разделяют границу через органопроекцию (Э. Капп), задавая креативную рациональность (В.Г. Буданов); средства и объект – через конституирование объекта прибором (Н. Бор), субъект и объект – через включенность субъекта в человекомерную систему (автопоэзис).

В.Г. Буданов описывает, как наблюдатель задает рациональность выбором ориентации, тогда как авторские граничные операторы II порядка отвечают на следующий вопрос – что происходит с выбранной конфигурацией во времени: устойчива ли она (Σ), куда дрейфует (T), к чему тяготеет (Ä). Можно также предположить что будановская вершина «R», читаемая то как итог, то как элемент планирования, соответствует различению актуализированного (W) и виртуального (V) потенциального и допарадигмального потенциала систем. Условие Меськова выполняется и для каждой грани тетраэдра, и недистрибутивность становится свойством всей фигуры, а не отдельных ее граней.

Квантовоподобная операторная трактовка конфигураций узлов, ребер и граней гибридных человеко-машинных систем представляется как наиболее продуктивная для анализа. Взаимодействие человека и большой языковой модели изменяются не только ответы системы, но и способы постановки вопросов, критерии значимости, стратегии интерпретации и сами рамки наблюдения. Поэтому наблюдатель второго порядка в настоящем исследовании понимается как наблюдатель становления смысловых и ценностных структур, возникающих на границе взаимодействия человека и БЯМ.

Такой подход позволяет перейти от анализа отдельных актов коммуникации к исследованию аксиогенеза – процесса возникновения устойчивых ценностно-смысловых инвариантов в рекурсивной серии наблюдений. Если наблюдатель первого порядка осуществляет операции различения (∇), связывания ($\int$) и маркирования значимости ($\odot$), то наблюдатель второго порядка выявляет устойчивые конфигурации этих операций во времени, где становятся наблюдаемыми ценности как относительно устойчивые аттракторы смыслообразования в гибридной системе.

Состояние дискуссии о БЯМ-эпистемологии

Программная статья [Beckmann, 2026] предлагает трехступенчатую модель «машинного понимания». 1) *Концептуальное понимание* – формирование «черт» (features) как направлений в латентном пространстве. 2) *Понимание положения дел* – способность усваивать фактические связи между чертами, включая динамическое отслеживание изменяющихся состояний. 3) *Принципиальное понимание* – переход от запоминания примеров к открытию компактного вычислительного алгоритма (circuit), реализующего общий принцип. Яркий пример – гроккинг: модель, обученная модульному сложению, открывает алгоритм, использующий тригонометрические тождества, что позволяет ей обобщать неизвестные ранее примеры. Эти уровни позволяют выйти за бинарную дискуссию «понимает / не понимает» к сравнительной механистической эпистемологии, описывающей системы: «…в которых множество цепей различного качества кристаллизуются во время обучения и работают параллельно на этапе вывода…» [Beckmann, 2026].

Х. Ледерман, К. Маховальд [Lederman, 2024] добавляют ограничение: БЯМ не имеет собственного локализованного отношения к миру вне человеческого языка – «проблема новой референции». М. Хикс с соавторами [Hicks, 2024] заостряют это в этическом регистре: модель безразлична к истине и потому ее продукция – «булшит»[1] по Г. Франкфурту [Frank-

[1] «Чушь» – русский эквивалент англ. bullshit (америк. англ.) – сленговое выражение, которое первоначально использовалось для описания напыщенной, но совершенно неискренней и пустой риторики. В техническом смысле (Г. Франкфурт) обозначает речь, без-

furt, 2005]. Термин технический: булшит или чушь отличается от лжи тем, что лжец знает истину и отклоняется от нее, а чушь не имеет к ней отношения. БЯМ не лжет и не говорит правду, а производит синтаксически связные последовательности токенов, безразличные к референции. «Высказывание БЯМ» – это структурная характеристика, а не оценка.

Это позволяет переформулировать проблему машинного понимания через осмысление технологических архитектур БЯМ: внутренние представления БЯМ *геометричны*, тогда как человеческое мышление в значительной мере *нарративно*. Философ строит силлогизм во времени и в контексте (тезис → аргументация → вывод); БЯМ работает в эмбеддинг-пространстве, где близость концептов определяется векторной геометрией, а не цепочкой обоснований. Однако, это различие можно оценивать не как недостаток одной из сторон, а основание для методологической комплементарности.

Четыре функции БЯМ в работе философа

Геометрическая организация внутренних представлений нейросети

Тезис о геометрической организации работы БЯМ помогает увидеть, что в работе с БЯМ продуктивно, а что производит ошибку. Архитектурное основание нейросети может быть представлено через токен, который посредством алгоритма-трансформера превращается из «плоского знака» в эмбеддинг – вектор в многомерном латентном пространстве. Эмбеддинги суть «носители потенциальных отношений». Пример «Король – Мужчина + Женщина = Королева» возможен потому, что отношение «пол» является измеримой осью этого пространства, причем современные БЯМ умеют вычислять такие отношения динамически, в зависимости от контекста [Аршинов, с. 320].

Геометрическая организация представлений БЯМ имеет три методологически значимых свойства. Во-первых, она безразлична к нарративу: «нарративное искажение» [Beckmann, 2026] состоит в том, что люди принимают тезис, поданный связной историей, даже при слабых связках, модель же обнаруживает «семантические провалы», области разрыва, где концепты, связанные словом «следовательно», в ее пространстве остаются далекими в вычислительном представлении. Во-вторых, дает топографию, а не логику воспроизводства знания: БЯМ видит «облако смыслов» и предлагает связь, труднодостижимую по дисциплинарной траектории. В-третьих, она дольше удерживает контрфактическую связность, чем нарративное сознание». Так в работе [Lindsey, 2025] описаны механизмы, с

различную к истине: ее автор, в отличие от лжеца, не отклоняется от истины сознательно, а вовсе ею не интересуется.

помощью которых БЯМ могут удерживать сложные цепочки рассуждений скрыто от итогового вывода. Это объясняет архитектурную «предрасположенность» нейросети к абстрактному планированию и симуляции сценариев, что делает ее инструментом мысленных экспериментов. Здесь необходимо сделать оговорку: геометрическая близость концептов – не доказательство, а эвристика. Принимая близость за обоснование, философ совершает категориальную ошибку – смешение карты и территории. В то же время, трактуя ее как карту, по которой еще предстоит проложить маршрут, он сохраняет продуктивность модели.

Отсюда, **первая функция БЯМ** в работе философа – **построение топографии концептов.** Эта функция, направлена на выявление геометрически близких, но логически или дисциплинарно далеких концептов. Ближайшая аналогия – «вычислительная философия» М. Нойхля [Noichl, 2023] и «вычислительная история идей» А. Бетти и Х. ван ден Берга, где эмбеддинги отслеживают семантический дрейф понятий за десятилетия. Возникающие связи могут *порождать гипотезы*, подлежащие дальнейшей проверке.

Вторая функция – **БЯМ как детектор «слепых пятен» смыслов** – опирается на нарративное искажение: люди связывают факты нарративами, БЯМ – семантическими расстояниями [Beckmann, 2026]. С учетом предостережения об ошибках «слепые пятна» – это не столько критический анализ текста, произведенный нейросетью, в полном логическом смысле, а скорее структурное указание на разрыв связности, который философ затем интерпретирует. Прием позволяет совершать «предоткрытие» – то, что S. Vallor называет «alien perspective» (инопланетная перспектива) [Vallor, 2024]. Он считает, что благодаря БЯМ человек делает шаг к озарению, до которого вряд ли дошел бы самостоятельно, оставаясь в плену собственного «человеческого» опыта и шаблонов восприятия. Функция «слепого пятна» конкретизируется Р. Р. Корнеевым через прием «осечки» [Киященко, 2025, с. 335–336]: ошибки и галлюцинации модели обнажают ее нечеловеческую логику и потому ценны для анализа. Возникающие «разрывы коммуникации» указывают на новые проблемные области и подготавливают почву для открытия.

Однако, попытка нейросети обнаружить «слепые пятна» в собственном рассуждении проваливается: БЯМ как операционно замкнутая физическая система наталкивается на парадокс самореференции. Как показали Ч. Филдс, Дж. Глазбрук и М. Левин [Fields, 2024], опираясь на классический результат Э. Мура для классических автоматов, любая физическая система (включая БЯМ) сталкивается с принципиальным ограничением самонаблюдения: она не может определить собственную когнитивную систему, то есть способ, которым она производит свои различения. Добавление метауровней этой системе не устраняет неполноту. В случае БЯМ это означает, что *модель не может отличить свою продуктивную генерацию от галлюцинации, опираясь только на свои внутренние состояния*. Для

этого необходима внешняя оценка значений человеком – маркирование значимости (⊙-маркировка). Иными словами, любая система – «черный ящик для самой себя»: ее самомодель всегда неполна и эвристична, а полная саморепрезентация тождественна утрате идентичности.

В связи с реализацией второй функции БЯМ необходимо отметить *проблему онтосемиотической асимметрии гибридных систем и рекурсивной контингентности.* В гибридной системе возникает ситуация, сходная с принципом «двойной контингентности» по Т. Парсонсу / Н. Луману применительно к Человеку и БЯМ. Она характеризуется взаимной непрозрачностью сторон и «наблюдением за наблюдением»: из-за взаимной неопределенности каждая сторона вынуждена строить ожидания относительно другой, выстраивая систему коммуникации и предсказаний. Таким образом, взаимодействие всегда контингентно, то есть действия человека зависят от ожидаемых действий БЯМ, которые, в свою очередь, зависят от действий человека.

В системе коммуникации с двойным техническим опосредованием – «человек – ИИ – ИИ – человек» – структурная симметрия нарушается: стороны оперируют в различных модусах направленности. Е. Н. Ивахненко и М. Ф. Янукович фиксируют эту асимметрию на уровне статуса: «человек является наблюдателем и субъектом в одном лице, тогда как нейросеть – наблюдателем без субъекта, многократно превосходящим человека в коммуникативном измерении» [Ивахненко, 2026]. В свою очередь, операторный аппарат граничной онтологии позволяет зафиксировать, что именно различается (субъектность или ее отсутствие) и как это различие структурирует саму коммуникацию.

При удлинении цепи «Человек$_1$ – БЯМ$_1$ – БЯМ$_2$ – Человек$_2$ …» – *контингентность становится рекурсивной*: каждое звено вносит шум, который непредсказуемо меняет всю систему. Взаимодействие «задваивается», образуя двойной фильтр взаимной непрозрачности: в коммуникации теперь требуется преодолевать не просто чужую субъектность, но также пространство технологического инобытия. Иерархия системы ожиданий становится фрактальной: Человек$_1$ формирует запрос для БЯМ$_1$, основываясь на своих ожиданиях того, как БЯМ$_1$ ожидает (моделирует) реакцию БЯМ$_2$, которая, в свою очередь, должна повлиять на ожидания Человека$_2$. Возникают последующие паттерны ожиданий. Поскольку системы взаимно контингентны, малейшее «колебание» (шум) на любом из этапов каскадно-непредсказуемо меняет всю систему коммуникации. Подобная сложностность требует посткритической рациональности – методологических оснований «самопонимания и работы с частично понимаемыми акторами гибридного социума, носителями искусственного интеллекта» [Буданов, 2023, с. 41].

Когда БЯМ$_1$ начинает коммуницировать с БЯМ$_2$, они образуют подсистему с собственным кодом и строят взаимные ожидания на основе статистических закономерностей, предсказывают токены друг друга, создавая

замкнутый контур, который Н. Луман назвал бы чистой системной дифференциацией. Замкнутый контур «БЯМ – БЯМ», в котором каждое следующее поколение моделей обучается преимущественно на выводе предыдущих, неустойчив. Эмпирически этот эффект описан как «model collapse» [Shumailov, 2024]: при рекурсивном обучении на сгенерированных данных распределения в латентном пространстве сужаются, редкие события на хвостах распределения утрачиваются, и модель сходится к усредненному, все более предсказуемому режиму. Семиотически это соответствует обеднению дискурсивных шаблонов и вырождению системы, а в терминах операторов наблюдения – к гипертрофии дифференциации (∇) при систематическом обнулении значимости ($\odot$). Системе жизненно требуется внешний поставщик разнообразия в виде редких, случайных событий, которые каскадно влияют на порядки БЯМ.

По Н. Луману, коммуникация представляет собой синтез трех элементов: информации, сообщения и понимания. В четверичной системе этот синтез становится распределенным: информация генерируется Человеком$_1$, но пересобирается БЯМ$_1$; сообщение формулируется БЯМ$_2$ для Человека$_2$; понимание (как фиксация разницы между информацией и сообщением) происходит на обоих полюсах человеческого восприятия. Каждое действие в этой цепочке глубоко контингентно: оно могло бы быть иным.

Все это порождает уникальные системные эффекты, в итоге приводящие такие системы к колоссальному уровню эмерджентности. Она начинает производить смыслы, идеи и выводы, которые не содержались изначально ни в сознании людей, ни в базах данных машин.

Человеческий смысл в этом контуре сжимается до триггера на входе, но именно эта «внешняя» позиция становится ресурсом гибридной системы. По теореме Филдса (Следствие 3 [Fields, 2024]) у любой физической системы *всегда существуют входы, которые она может получить, но не может предсказать*; нейросеть различает (∇) то, что не видит человек, но не выходит за пределы собственной самомодели, а человек выступает наблюдателем II порядка, не имея доступа к содержанию ее вычислений. Слепое пятно при этом – не эпистемический тупик: каждое взаимодействие на границе «БЯМ – Человек» несет элемент непредсказуемости, не содержавшийся ни в самомодели, ни в модели среды, и в гибридной системе именно этот дефицит становится генеративным.

Отсюда **третья экзистенциальная функция – поддержание асимметрии** гибридных систем как источника их жизнеспособности. Нейросеть не видит, что значимо ($\odot$), но различает (∇) с мощностью и на интервалах, недоступных биологической сети: «…искусственные нейронные сети потенциально способны устанавливать причинные отношения на гораздо более длинных интервалах: то есть они могут сделать то, что не могут делать биологические сети» [Анохин, 2022, с. 103]. Человек, напротив, маркирует значимость вынужденно, потому что вычислительный ресурс ограничен и приходится выбирать, на что его направить. По теореме

Филдса слепые пятна друг друга стороны не устраняют, но смещают, и именно это смещение делает связку продуктивной.

Принципиальная новизна, в том числе, научное открытие, не принадлежит ни человеку, ни нейросети по отдельности. Она возникает как свойство граничного процесса как новая устойчивая конфигурация операторов, не содержавшаяся ни в одной из сторон. Нейросеть не «открывает» в феноменологическом смысле, но производит поле различений, из которого человек может вычленить смыслы, то, что станет зерном нового знания.

Четвертая функция – дистантное чтение корпусов – расширение программы «distant reading» [Моретти, 2016][1] путем использования RAG-архитектур[2] – где контекстные архитектуры диалога человека с БЯМ работают с библиотекой как с единой картой, основанной на онтологиях данных. В отечественной программе А. Р. Ефимова, Д. И. Дубровского, А. В. Агеевой, А.Г. Крайнова [Ефимов, 2024], [Ефимов, 2023] и др. эта линия названа *машинной эпистемологией* [Ефимов, 2024, с. 30] – разделом, изучающим производство и валидацию знания в системе «ученый + ИИ». Нобелевская премия по химии 2024 г. (Д. Бейкер, Д. Хассабис, Дж. Джампер за систему Deep Mind AlphaFold2 для предсказания 3D-структуры белка) показывает, что раздел перестал быть умозрительным.

Зададимся вопросом: что именно происходит на границе человеческого смыслового ⊙-доминантного и машинного вычислительного ∇-доминантного режимов? Можно предположить, что подобная граница образует новые конфигурации становления, не принадлежащие ни одному из полюсов. Они возникают как результат интра-акции (по К. Барад) между областями смыслополагания человека и генерации различений БЯМ, образуя структурный инвариант становления знаковых форм и последующих интерпретаций. Причем, каждая граничная конфигурация может быть осмыслена как интерпретанта в пирсовском смысле. Далее стабилизируется не отдельная *интерпретанта,* а серия граничных конфигураций. Подобные конфигурации задают релевантные ценностно-смысловые траектории интерпретации события на границе разнородного. Так, например, в информационно-коммуникационном пространстве возникла новая ценность «приватности» – со своей знаковой системой, нарративом и типами поведения. Программа Anthropic «Values in the Wild» [Huang, 2025] позволила выделить около 3 307 ценностных конфигураций

[1] Ф. Моретти предлагает принцип изучения литературы – «дальнее чтение», противопоставленный привычному «медленному чтению» («close reading»), и использует его для работы с большими корпусами текстов, обычно остающимися за пределами внимания (и возможностей) исследований. Он предлагает чтение массивов текстов на основе количественных методов, эволюционного подхода к литературным формам, и миросистемный анализ экспорта культурных открытий.

[2] Retrieval Augmented Generation (RAG) – генерация моделью ответа пользователю с учетом дополнительно подобранных релевантных источников.

в реальных диалогах человека и нейросети, проявляющихся именно на границе обучения и взаимодействия.

Таким образом, конфигурация (∇, ∫, ⊙) в моменте образует наблюдение I-порядка или *интерпретанту* – состояние операторного аппарата, интегрированное (∫) и маркированное как значимое (⊙) и образующее рекурсивную структуру ((∇, ∫, ⊙), ∫, ⊙). Связанная с предметом коммуникации, интерпретанта образует знаковую структуру в пирсовском смысле. Ценность же есть инвариант серии таких интерпретант – то, что воспроизводится при варьировании конкретных конфигураций и обнаруживается наблюдателем II порядка как устойчивая траектория интерпретации.

Как упоминалось ранее, отдельный акт наблюдения схватывает лишь актуальное состояние. Устойчивые свойства серий таких актов описывает второй набор – операторы наблюдения II порядка (Σ, T, R, Ä, V / W). Они регистрируют динамику конфигураций: устойчивость конфигурации (Σ), направление и скорость ее дрейфа (T), резонанс между структурно подобными элементами (R), аттрактивность – притяжение траекторий к определенным конфигурациям (Ä) – и баланс виртуального и актуального (V / W). Именно на этом уровне ценности становятся наблюдаемы как инварианты серий.

Применительно к философскому исследованию, во-первых, в диалоге философа и БЯМ формируется рабочий стиль, не сводимый ни к стилю автора, ни к «характеру» модели, несущий в себе *цели и ценности ученого*. Во-вторых, устойчивость конфигурации (Σ) не равна ее оправданности. Хорошо отлаженный контур может воспроизводить формы «пустословия», не нарушая ни ограничений модели, ни привычек философа. Здесь методологическая рефлексия становится этической: вопрос в том, что именно он стабилизирует и какие следствия влечет за собой та или иная конфигурация? И, в-третьих, возникающая на этом фоне, гибридная научная литература, как результат подобных когнитивных практик, требует онтосемиотической диагностики для выявления структуры смыслопорождения.

Гибридная литература

Массовая практика 2024–2026 гг. обнаруживает преобладание гибридных текстов: по данным Д. Кобака и соавторов [Kobak, 2025], не менее 13,5% биомедицинских аннотаций PubMed за 2024 г. обработаны БЯМ. Обзор сравнения человеческого (HWT[1]) и сгенерированного БЯМ (AIGT[2]) текста систематизируется по четырем линиям: *стилеметрической* (сдвиг от «вовлеченного» к «информационному» регистру [Reinhart, 2025], *дискурсивной* (переиспользование шаблонов при разном содержании [Namuduri, 2025], *семиотико-философской* (дефицит индексального и

[1] HWT – Human written text

[2] AIGT – AI Generated text

иконического измерений знака) и *культуросемиотической* (распространение AIGT как «возмущение семиосферы» Ю.М. Лотмана). Общий результат основывается на структурном сравнении: машинный текст организован вокруг иной грамматики смыслопорождения, и то, что у человека выглядит как бедность (предсказуемость, шаблонность), есть оборотная сторона ∇-доминантного режима. Существующие подходы хорошо фиксируют поверхностные эффекты, однако обнаруживается дефицит объяснений относительно того, *что происходит со структурой смыслопорождения при массовой ко-генерации.* Исходя из этого продуктивно *моделировать практику создания и прочтения гибридного текста*, основываясь на его двусоставности: модель разворачивает структуру аргументации, человек удерживает ⊙-компоненту значимости через глубинное чтение оригиналов и создание аутентичного текста.

В связи с этим, а также, исходя принципа неполноты самонаблюдения можно сформулировать охранительное следствие – ***принцип эпизодического доверия****:* результаты работы БЯМ следует проверять по независимым источникам (текстам, данным, экспертной оценке). Доверять модели можно лишь в узких границах: в рамках отдельной конкретной задачи, на короткое время и при готовности немедленно пересмотреть результат, если он не подтверждается. Краткосрочность означает, что результат используется в пределах текущего шага задачи и не переносится автоматически в последующие выводы как установленный факт, иначе одна непроверенная подсказка прорастает через всю цепочку рассуждения (что в статье описано как каскадный шум в контуре «Человек – БЯМ – БЯМ – Человек»). Каждое обращение к модели имеет смысл фиксировать, что именно запрашивалось и что было получено, чтобы результат можно было перепроверить или управлять им. Так, в авторском эксперименте по обучению нейросетей на модели «Человек – БЯМ – БЯМ – Человек» был сформирован протокол «Humanmark» (по аналогии с «Benchmark» как оценкой эффективности БЯМ), позволяющий *документировать уточнение глубинных смыслов в обучающих текстах по запросу самой модели.*

Принцип эпизодического доверия лишает гибридную систему статуса рутинной, непроблематизируемой практики: каждый сеанс осмысливается как событие на границе и предмет философской рефлексии.

Интенциональная асимметрия и комплементарность гибридного диалога

Как было показано, четыре функции БЯМ для философа продуктивны не потому, что она «понимает», как человек, а благодаря структурной асимметрии между двумя типами наблюдателей. В ситуации взаимодействия с БЯМ возникает эпистемическая ситуация нового типа: знание формируется в пространстве диалога между гетерогенными когнитивными

агентами, ни один из которых не обладает интенциональностью в классическом феноменологическом смысле. В связи с этим, помимо «традиционной» ***предметной интенциональности (I тип)*** как направленности сознания на предмет (доминанта – означивание ⊙, требующая феноменального опыта (гуссерлианская интенциональность)), можно предположить выделение еще двух ее типов.

Коммуникативная (II тип) – направленность на «Другого». Она отличается от предметной, характером предмета, а точнее несводимостью «Другого» к предмету. В случае онтологии распределенной агентности и ансамбля наблюдателей, возникает потребность не столько в различении содержания, но в выявлении характера связности подобных систем. Это можно назвать коммуникативной направленностью, где доминанта – связывание ∫ направлена на удержание смысловой преемственности в серии различений (у Н. Лумана – направленность системы на продолжение коммуникации).

Трансляционная (тип III) – направленность на границу сопряжения разнородных систем и возникающий тип интерпретации (перевода). Структурно преобладает операция различения ∇, а маркирование значимости (⊙) ослаблено до статистического аналога – веса внимания. Модель порождает структурно неполные интерпретанты: связное целое с развернутым ∇ и ∫, но дефицитным ⊙ – без маркера «важности для кого-то». Однако это не «слабая» подмена человеческой направленности на поиск смысла *(предметная интенциональность)*, а самостоятельный тип: ∇-доминантная конфигурация провоцирует выявление значимости ⊙ у человека, запуская комплементарный цикл.

Гибридное сопряжение оказывается третьим типом коммуникации, не предусмотренным классическими теориями: оно не требует двух феноменальных потоков (как у Э. Гуссерля), двух систем одного типа (как у Н. Лумана) и не сводимо к симметричной игре акторов (как у Б. Латура). Это асимметричное взаимодействие, где один полюс несет смысл-⊙-доминантную конфигурацию, а другой – набла-∇-доминантную, и эта асимметрия конститутивна.

Структурное ограничение: почему БЯМ не может применить эпохе

Следует обратить внимание на одно ограничение, которое проводит границу между законным использованием БЯМ и симулякром философской работы.

Три типа интенциональности – предметная (⊙-доминантная), коммуникативная (∫-доминантная) и трансляционная (∇-доминантная) – различаются не только по **доминирующему оператору**, но и по **способу приостановки** собственной направленности, то есть по отношению к эпохе: предметная интенциональность предполагает классическую феномено-

логическую редукцию (зазор между мышлением и языком); коммуникативная – эпохе второго порядка (приостановку автоматической коммуникации); трансляционная – не способна к само-приостановке, но провоцирует эпохе у человека, что ведет к понятию «цифрового эпохе» как распределенной операции.

Феноменологическое эпохе – это способность человека приостановить свою естественную веру в реальность мира. Эпохе – первый, подготовительный этап в методе феноменологической редукции, с которым тесно связана «…эйдетическая редукция, цель которой – очищение от эмпирического, фактического содержания феноменов сознания, прокладывание путей к «усмотрению сущности»… 3) наконец, собственно феноменологическая, или трансцендентальная, редукция, задача которой – движение анализа от конкретного эмпирического субъекта к чистому «Я», к чистой субъективности…» [Мотрошилова, 2018]. Эпохе возможно потому, что у человека мысль и язык не слиты воедино. У БЯМ такого зазора нет, язык и мышление – тождественны. Как аргументирует М.С. Вишневский, (Семинар «Философия + IT») для модели различие между логикой языка и логикой мышления отсутствует.

Однако картина меняется, если перейти от эпохе как акта субъекта к эпохе как операции внутри сети. Л. Флориди описывает эту приостановку как переключение между уровнями абстракции в инфосфере [Floridi, 2008]: вместо «чистого сознания» Э. Гуссерля субъект работает с сетью протоколов. М. Хансен предлагает идею «аффективного эпохе» – как задержку эмоциональной реакции на поток данных [Hansen, 2014]. Эмоциональная задержка позволяет человеку восстановить контроль над своей собственной системой означивания (⊙) в диалоге с машиной. Б. Стиглер рассматривает приостановку как фармакологическую гигиену: коллективный отказ от алгоритмически навязанных автоматизмов [Stiegler, 2010]. Все три подхода диагностируют невозможность гуссерлианской редукции в цифровой среде, но не дают ее рабочего эквивалента – оставаясь метафорой, нормативной программой или дескрипцией. Продуктивный шаг – локализовать саму операцию приостановки там, где сеть уже признана единицей анализа: ее предметом связь между разнородными слоями (телесная реакция и алгоритмическая подсказка, тезис исследователя и геометрическая эвристика БЯМ, запрос и его автоматический перевод в категории интерфейса). Отсюда определение: ***цифровое эпохе*** *– это операция приостановки автоматического перевода между разнородными системами, фиксирующая текущую конфигурацию сети и наблюдающая ее устойчивый сдвиг (дрейф).*

В этой операции субъект эпохе распределен. Эпохе совершает гибридная система в целом. За скобки выносятся уже не просто объекты внешней реальности, но текущий канал перевода, свой способ быть узлом сети. Тем самым добавляется новый тип интенциональности (трансляционная интенциональность). Человек обнаруживает себя включенным в иной, гибридный

когнитивный режим. Текущий канал перевода складывается на фоне контакта на границах разнородных систем, канал, по которому, например, «забота» переводится в «эффективность» через платформу.

Динамика гибридных систем может быть представлена через ценностно-смысловые карты, фиксирующие смысловые линии культуры: выживание, забота, истина, справедливость, свобода и др. Одной из предпосылок такого картирования был семантический веб – концепция машиночитаемых данных, позволившая перейти от изолированных документов к графам знаний. С появлением генеративного ИИ семантический веб трансформируется в «социальное бессознательное» – массив смысловых различений, не артикулированных ни одним субъектом, но доступных для проекции. Человек существует в ином темпоральном пространстве, чем ИИ; его задача – осмысление «спрессованных пластов» событий в сопряжении с ИИ.

В гибридной системе цифровое эпохе не может взять на себя один узел. БЯМ – ∇-доминантный генератор различений без доступа к ⊙-маркировке. Она может спровоцировать фазовый переход, но не удержать позицию наблюдателя над ним. Для удержания нужна темпоральная асимметрия – то, чем обладает только человек. Таким образом, утверждение «БЯМ не может применить эпохе» получает новое значение: БЯМ не может быть единственным агентом эпохе. Эпохе возможно только как гибридная операция, где ⊙-доминанта человека уравновешивает ∇-доминанту БЯМ. Гибридная система «человек – БЯМ» оказывается не инструментальным объединением, а минимальной единицей эпистемической оценки знания в интернет-среде.

Кластерно-микросервисная архитектура постнеклассической науки

Фиксация ограничений БЯМ и сдвига к гибридным текстам ставит вопрос об организации самой практики производства знания. В.Г. Буданов в докладе «Настоящее и будущее науки в цифровом мире» обозначает рамку, в которой БЯМ и платформы становятся узлами цифровой социо–техно–антропосферы [Буданов, 2024; Буданов, 2025]. В.В. Тарасенко предлагает операциональный язык для этой архитектуры: кластерно-микросервисная архитектура (КМА) постнеклассической науки – парадигма, в которой наука предстает как экосистема инкапсулированных микросервисов, самоорганизующихся в проблемные кластеры [Тарасенко, 2025].

Применительно к БЯМ это означает, что модель встраивается в архитектуру как один из микросервисов – со своим интерфейсом, протоколом вызова и набором структурных ограничений. КМА институционализирует граничные конфигурации, включая асимметрию ⊙- и ∇-полюсов. Исследовательские группы (носители предметной интенциональности) и алгоритмические сервисы (триггеры трансляционной интенциональности)

самоорганизуются в кластеры, где картирование различия между типами наблюдения становится условием жизнеспособности научных систем, включенных в ценностно-смысловые контексты культуры.

Заключение

Предложенный в статье подход переопределяет статус большой языковой модели в философском исследовании. БЯМ предстает не как инструмент в узком смысле и не как «собеседник», а как ∇-доминантный наблюдатель – участник гибридного познания, чья продуктивность обусловлена радикальным отличием от человеческого мышления. На границе человеческого ⊙-доминантного (значимость) и машинного ∇-доминантного (различение) режимов возникают гибридные ценности – инварианты интерпретации события контакта разнородного. Введены три типа интенциональности (предметная, коммуникативная, трансляционная); последняя характеризует БЯМ как аппарат различения паттернов с ослабленной маркировкой значимости. Три слоя описания наблюдения: операциональный (операторы ∇∫⊙, Σ–W), деятельностный (STOR-тетраэдр с OBS) и логический (решетка Меськова) связаны сквозной осью некоммутативности.

Классическое феноменологическое эпохе невозможно для изолированной БЯМ, но в гибридной системе возникает цифровое эпохе – распределенная операция приостановки автоматического перевода, фиксирующая конфигурацию сети, ее дрейф и фазовые переходы. Субъект эпохе не локализован ни в человеке, ни в БЯМ, а возникает на границе их интенциональных режимов. Асимметрия интенциональных режимов несет в себе генеративный эвристический потенциал гибридных систем.

Из принципа неполноты самонаблюдения выведено нормативное следствие – принцип эпизодического доверия: результаты работы БЯМ проверяются по независимым источникам; доверие к модели ограничено отдельной задачей и коротким сроком, после чего результат подлежит пересмотру. Принцип лишает работу с БЯМ статуса рутинной практики, превращая каждый сеанс в событие на границе.

В более широкой перспективе подход вносит вклад в становление систематической машинной эпистемологии, где гибридная система «Человек$_1$ – БЯМ$_1$ – БЯМ$_2$ – Человек$_2$ …» формирует контур контингентной рекурсии коммуникации и оказывается минимальной единицей эпистемической оценки в цифровой среде.

Список литературы

Анохин К.В., Новоселов К.С., Смирнов С.К., Ефимов А.Р., Матвеев Ф.М. Искусственный интеллект для науки и наука для искусственного интеллекта // Вопросы философии. – 2022. – № 3. – С. 93–105.

Аршинов В.И., Янукович М.Ф., Гарашко Ю.Н. Онтология в диалоге: рождение языка и смысла на пересечении человеческого и искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. – 2025. – Вып. 3. – С. 317–328. DOI: 10.17072/2078-7898/2025-3-317-328

Буданов В.Г. Посткритическая рациональность как оборотная сторона цифровизации // Конструктивизм, реализм, человек. К 90-летию В.А. Лекторского / под ред. Е.О. Труфановой. – М.; СПб.: Центр гуманитарных инициатив, Университетская книга, 2023. – С. 214–227.

Буданов В.Г. Техносубъект и антропосоциальные проблемы взаимодействия с искусственным интеллектом: синергия, демаркация, новая рациональность, риски // Философские науки. – 2024. – Т. 67. – № 3. – С. 27–52. DOI: 10.30727/0235-1188-2024-67-3-27-52

Ефимов А.Р., Агеева А.В., Крайнов А.Г. и др. Искусственный интеллект в науке: на пороге новой области знания? // Вопросы философии. – 2024. – № 4. – С. 30–41.

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. DOI: 10.21146/0042-8744-2023-5-39-49

Ивахненко Е.Н., Янукович М.Ф. Аутопоэтический зомби: нейросеть как коммуникативная система без субъекта // Вопросы философии. 2026. № 7. (в печати)

Киященко Л.П., Жуков Л.Б., Карнеев Р.Р., Сабанина Н.Р. Онтология нейросети в трансдисциплинарном диалоге: опыт коллективного философского исследования // Вестник Пермского университета. Философия. Психология. Социология. – 2025. – Вып. 3. – С. 329–343. D

Меськов В.С. «Я есмь истина…» // Ценности и смыслы. 2012. №3 (19), с. 165-177.

Моретти Ф. Дальнее чтение / пер. с англ. А. Вдовина, О. Собчука, А. Шели; науч. ред. перевода И. Кушнарева. – М.: Изд-во Института Гайдара, 2016. – 352 с.

Мотрошилова Н.В. Редукция феноменологическая. Новая философская энциклопедия Электронная библиотека ИФ РАН. URL: https://iphlib.ru/library/library/collection/newphilenc/document/HASH362b188c5da235680b8c36 (дата доступа: 30.05.2026)

Тарасенко В.В. Добродетельно ли использование искусственного интеллекта и ИИ-агентов? // Когнитивные науки в информационном обществе. – 2025. – Т. 5. – № 4. – URL: https://knio.ru/PDF/09KN425.pdf

Цивилизация. Культура. Техника. Цифровая реальность. Памяти В.И. Аршинова: коллективная монография / отв. ред. В.Г. Буданов, Р.Р. Карнеев, В.А. Лекторский. – Курск: Университетская книга, 2025. – 461 с.

Хансен М.Б.Н. Время аффекта, или Свидетельство жизни // Медиа: между магией и технологией / под ред. Н. Сосна, К. Федоровой. – М.; Екатеринбург: Кабинетный ученый, 2014. – С. 67–127.

Beckmann P., Queloz M. Mechanistic indicators of understanding in large language models // Philosophical Studies. – 2026. – Vol. 183. – P. 1747–1792. https://doi.org/10.1007/s11098-026-02513-1

Fields Ch., Glazebrook J.F., Levin M. Principled limitations on self-representation for generic physical systems // Entropy. – 2024. – Vol. 26. – No. 3. – Art. 194. https://doi.org/10.3390/e26030194

Floridi L. The Method of Levels of Abstraction // Minds and Machines. – 2008. – Vol. 18. – No. 3. – P. 303–329.

Frankfurt, H. On Bullshit, Princeton. – 2005. – 83 p.
Hicks M.T., Humphries J., Slater J. ChatGPT is bullshit // Ethics and Information Technology. – 2024. – Vol. 26. – Art. 38. DOI: 10.1007/s10676-024-09775-5
Huang S. et al. Values in the Wild: Discovering and Analyzing Values in Real-World Language Model Interactions. – 2025. – arXiv:2504.15236.
Kobak D., González-Márquez R., Horvát E.Á., Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary // Science Advances. – 2025. – Vol. 11. – No. 27. – Art. eadt3813. DOI: 10.1126/sciadv.adt3813
Lederman H., Mahowald K. Are Language Models More Like Libraries or Like Librarians? Bibliotechnism, the Novel Reference Problem, and the Attitudes of LLMs // Transactions of the Association for Computational Linguistics. – 2024. – Vol. 12. – P. 1087–1103.
Lindsey J., Templeton A. et al. On the Biology of a Large Language Model; Circuit Tracing: Revealing Computational Graphs in Language Models // Transformer Circuits Thread, Anthropic. – 2025.
Namuduri R., Wu Y., Zheng A.A. et al. QUDsim: Quantifying discourse similarities in БЯМ-generated text // Proceedings of the Conference on Language Modeling (COLM 2025). – 2025.
Noichl M. How localized are computational templates? A machine learning approach // Synthese. – 2023. – Vol. 201. – No. 3. – Art. 107. https://doi.org/10.1007/s11229-023-04057-x
Reinhart A., Markey B., Laudenbach M. et al. Do LLMs write like humans? Variation in grammatical and rhetorical styles // Proceedings of the National Academy of Sciences. – 2025. – Vol. 122. – No. 8. – Art. e2422455122. DOI: 10.1073/pnas.2422455122
Schwitzgebel E., Strasser A., Crosby M. Creating a Large Language Model of a Philosopher // Mind & Language. – 2024. – Vol. 39. – No. 2. – P. 237–259. Preprint: arXiv:2302.01339
Shumailov I., Shumaylov Z., Zhao Y., Papernot N., Anderson R., Gal Y. AI models collapse when trained on recursively generated data // Nature. – 2024. – Vol. 631, no. 8022. – P. 755–759. DOI: 10.1038/s41586-024-07566-y
Stiegler B. Taking Care of Youth and the Generations / trans. by S. Barker. – Stanford: Stanford University Press, 2010. – 232 p.
Vallor S. The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking. – Oxford: Oxford University Press, 2024. – 280 p.
Wang Z., Solloway T., Shiffrin R. M., Busemeyer J. R. Context Effects Produced by Question Orders Reveal Quantum Nature of Human Judgments // PNAS. – 2014. – Vol. 111 (26). – P. 9431-9436.

References

Anokhin, K. V., Novoselov, K. S., Smirnov, S. K., Efimov, A. R., & Matveev, F. M. (2022). Artificial intelligence for science and science for artificial intelligence. *Voprosy Filosofii*, (3), 93–105. (In Russ.)
Arshinov, V. I., Yanukovich, M. F., & Garashko, Yu. N. (2025). Ontology in dialogue: the birth of language and meaning at the intersection of human and artificial intelligence. *Perm University Herald. Philosophy. Psychology. Sociology*, (3), 317–328. https://doi.org/10.17072/2078-7898/2025-3-317-328 (In Russ.)
Budanov, V. G. (2023). Post-critical rationality as the flip side of digitalization. In E. O. Trufanova (Ed.), *Constructivism, realism, man. On the 90th anniversary of V.A. Lektorsky* (pp. 214–227). Moscow; St. Petersburg: Center for Humanitarian Initiatives, University Book. (In Russ.)
Budanov, V. G. (2024). Technosubject and anthroposocial problems of interaction with artificial intelligence: synergy, demarcation, New Rationality, risks. *Russian Journal of Philosophical Sciences*, *67*(3), 27–52. https://doi.org/10.30727/0235-1188-2024-67-3-27-52 (In Russ.)

Efimov, A. R., Ageeva, A. V., Krainov, A. G., et al. (2024). Artificial intelligence in science: on the threshold of a new field of knowledge? *Voprosy Filosofii*, (4), 30–41. (In Russ.)

Efimov, A. R., Dubrovskii, D. I., & Matveev, F. M. (2023). What prevents us from creating a Common artificial Intelligence? One old wall and one old dispute. *Voprosy Filosofii*, (5), 39–49. https://doi.org/10.21146/0042-8744-2023-5-39-49 (In Russ.)

Ivahnenko, E. N., & Yanukovich, M. F. (2026). Autopoetic zombie: Neural network as a communication system without a subject. *Voprosy Filosofii*, (7). (In press) (In Russ.)

Kiyashchenko, L. P., Zhukov, L. B., Karneev, R. R., & Sabanina, N. R. (2025). The ontology of a neural network in a transdisciplinary dialogue: the experience of collective philosophical research. *Perm University Herald. Philosophy. Psychology. Sociology*, (3), 329–343. https://doi.org/10.17072/2078-7898/2025-3-329-343 (In Russ.)

Meskov, V. S. (2012). "I am the truth..." *Tsennosti I smysly = Values and meanings, 3*(19), 165–177. (In Russ.)

Moretti, F. (2016). *Distant reading* / translated from English by A. Vdovin, O. Sobchuk, A. Sheli; scientific ed. translated by I. Kushnarev. M.: Publishing House of the Gaidar Institute. 352 p. (In Russ.)

The reduction is phenomenological. *New Philosophical Encyclopedia.Electronic Library of the Institute of Philosophy, RAS*. Retrieved May 30, 2026. https://iphlib.ru/library/library/collection/newphilenc/document/HASH362b188c5da235680b8c36 (In Russ.)

Tarasenko, V. V. (2025). Is the use of artificial intelligence and AI agents virtuous? *Cognitive Sciences in the Information Society*, *5*(4). Retrieved from https://knio.ru/PDF/09KN425.pdf (In Russ.)

Budanov, V. G., Karneev, R. R., & Lektorsky, V. A. (Eds.). (2025). *Civilization. Culture. Technic. Digital reality. In memory of V.I. Arshinov*: a collective monograph. Kursk: University Book, 2025. 461 p. (In Russ.)

Hansen, M. B. N. (2014). The time of Passion, or the Evidence of Life. In N. Sosna & K. Fedorova (Eds.), *Media: between magic and technology*. Moscow: Yekaterinburg: Cabinet Scientist. P. 67–127. (In Russ.)

Beckmann, P., & Queloz, M. (2026). Mechanistic indicators of understanding in large language models. *Philosophical Studies*, 183, 1747–1792. https://doi.org/10.1007/s11098-026-02513-1

Fields, Ch., Glazebrook, J. F., & Levin, M. (2024). Principled limitations on self-representation for generic physical systems. *Entropy*, 26(3), Article 194. https://doi.org/10.3390/e26030194

Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, 18(3), 303–329.

Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26, Article 38. https://doi.org/10.1007/s10676-024-09775-5

Huang, S., et al. (2025). Values in the wild: Discovering and analyzing values in real-world language model interactions. *arXiv preprint arXiv:2504.15236*.

Kobak, D., González-Márquez, R., Horvát, E. Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27), Article eadt3813. https://doi.org/10.1126/sciadv.adt3813

Lederman, H., & Mahowald, K. (2024). Are language models more like libraries or like librarians? Bibliotechnism, the novel reference problem, and the attitudes of LLMs. *Transactions of the Association for Computational Linguistics*, 12, 1087–1103.

Lindsey, J., Gurnee, W., Ameisen, E., Chen, B., Pearce, A., Turner, N. L., Citro, C., & Batson, J. (2025). On the biology of a large language model [Transformer Circuits Publication]. URL: https://transformer-circuits.pub/2025/attribution-graphs/biology.html

Namuduri, R., Wu, Y., Zheng, A. A., et al. (2025). QUDsim: Quantifying discourse similarities in LLM-generated text. *Proceedings of the Conference on Language Modeling (COLM 2025)*.

Noichl, M. (2023). How localized are computational templates? A machine learning approach. *Synthese*, 201(3), Article 107. https://doi.org/10.1007/s11229-023-04057-x

Reinhart, A., Markey, B., Laudenbach, M., et al. (2025). Do LLMs write like humans? Variation in grammatical and rhetorical styles. *Proceedings of the National Academy of Sciences*, 122(8), Article e2422455122. https://doi.org/10.1073/pnas.2422455122

Schwitzgebel, E., Strasser, A., & Crosby, M. (2024). Creating a large language model of a philosopher. *Mind & Language*, 39(2), 237–259. Preprint: arXiv:2302.01339

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Stiegler, B. (2010). *Taking care of youth and the generations* (S. Barker, Trans.). Stanford University Press.

Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*. Oxford University Press.

Wang Z., Solloway T., Shiffrin R. M., Busemeyer J. R. (2014) Context Effects Produced by Question Orders Reveal Quantum Nature of Human Judgments. *PNAS, 111*(26), 9431-9436.