

Научная статья

DOI: 10.15593/perm.kipf/2025.3.04

УДК 004.8:004.5



ИИ: ГРАНИ «МАШИННОСТИ» И «ЧЕЛОВЕЧНОСТИ»

С.В. Оболкина

Институт философии и права Уральского отделения РАН, г. Екатеринбург, Российская Федерация

О СТАТЬЕ

Поступила: 13 апреля 2025 г.
Одобрена: 09 июня 2025 г.
Принята к публикации: 31 июля 2025 г.

Ключевые слова:

машина, искусственный интеллект, естественный интеллект, нейросеть, логика, математика, метамашинный интеллект.

АННОТАЦИЯ

Рефлексия сопряжения искусственного интеллекта и человека, поиск специфики человеческого интеллекта в ситуации все активнее проявляющейся конкуренции между ними – одна из самых актуальных проблем современности, которая выступает тематическим контекстом данной работы. Главный тезис статьи: искусственный интеллект не противопоставлен естественному человеческому интеллекту и является экстерииоризацией некоторых его компетенций. Главным в методологическом подходе исследования выступает неразрывная связь философских спекуляций по поводу искусственного интеллекта и конкретики инженерного праксиса в этой сфере. Внимание к самой машине является обязательным. Показано, что в технический артефакт изначально преобразуется только тот модус когитации, который можно «машинизировать», то есть выстроить в качестве особых структур, обладающих вынуждающей силой вывода. В работе рассмотрены формальная логика и математика в качестве искусственной части естественного интеллекта человека. Этот «внутренний» искусственный интеллект обладает возможностями экстерииоризации процесса обучения и выстраивания технических артефактов. Предложен краткий анализ «промежуточной» формы между «внутренним» и «внешним» искусственным интеллектом: логических машин, машин Г.В. Лейбница и С.Н. Корсакова. Показаны эпистемологические основания и условия формирования компьютерной метафоры мозга, а также парадигмы инженерного коннекционизма. Далее два основных модуса технического искусственного интеллекта рассмотрены в оптике модусов машины: в качестве классической и неклассической машины. На основе представления о специфике неклассической машины аргументируется тезис о том, что нейросети также выступают экстерииоризацией – но не работы мозга как таковой, а одного из модусов естественного интеллекта, связанного с коллективным опытом. В заключение сделаны краткие выводы о мета-машинных компетенциях человеческого интеллекта.

© **Оболкина Светлана Викторовна** – кандидат философских наук, старший научный сотрудник отдела истории и философии науки, ORCID: <https://orcid.org/0000-0002-6644-104X>, e-mail: Obol2007@mail.ru.

© **Svetlana V. Obolkina** – Candidate of Philosophical Sciences, Senior Researcher, Educational Department of Philosophy, ORCID: <https://orcid.org/0000-0002-6644-104X>, e-mail: Obol2007@mail.ru.

Финансирование. Исследование не имело спонсорской поддержки.
Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.
Вклад автора. 100 %.



Эта статья доступна в соответствии с условиями лицензии Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

AI: THE FACETS OF "MACHINE-NESS" AND "HUMANITY"

S.V. Obolkina

Institute of Philosophy and Law of the Ural Branch of the Russian Academy of Sciences,
Yekaterinburg, Russian Federation

ARTICLE INFO

Received: 13 April 2025
Revised: 09 June 2025
Accepted: 31 July 2025

Keywords:

machine, artificial intelligence, natural intelligence, neural network, logic, mathematics, metamachine intelligence.

ABSTRACT

The article is devoted to reflecting the problems that arise in the current situation of interaction and clear competition between artificial intelligence and humans. The search for the specifics of human intelligence in these conditions is an important task of philosophy. The main thesis of the work is that AI is not opposed to human intelligence; AI is the exteriorization of certain competencies of natural intelligence. The basic methodological approach entails that philosophical reflections on AI cannot ignore engineering practice in this field; attention to the machine is required. The author shows that only the way of thinking that can be "mechanized" transforms into a technical artifact. This means creating special mental structures that have the coercive power of inference. The article considers formal logic and mathematics as an artificial part of natural human intelligence. This "internal" artificial intelligence has the properties of exteriorization: it is associated with learning and the creation of technical artifacts. The author offers a brief analysis of the "intermediate" form between "internal" and "external" artificial intelligence. These are logical machines as well as those of G.V. Leibniz and S.N. Korsakov. The paper also shows which epistemological foundations contributed to the formation of the computer-brain metaphor and the connectionist paradigm in AI. Two main types of technical artificial intelligence are considered - classical and non-classical machines. Based on the idea of the specifics of non-classical machines, the author develops the thesis that neural networks act as a kind of natural intelligence responsible for collective experience. Finally, brief conclusions are drawn about the meta-machine capabilities of human intelligence.

Введение

Недавняя новость из сферы ИИ [1] описывает довольно курьезную ситуацию: компания Anthropic, создавшая одного из самых популярных ИИ-помощников по написанию текстов, настоятельно рекомендует соискателям на должности в самой Anthropic не использовать ИИ для написания резюме. То есть практическая реализации идеи тождества человеческого и машинного интеллектов сопровождается некой базовой интуицией их принципиального различия. В данной статье мы постараемся реализовать концептуальную рефлексии проблемы тождества и различия этих интеллектов в исторической динамике. «Машинность» и «искусственность» интеллекта сначала появляется как *модус естественного интеллекта* (далее ЕИ). Далее часть ЕИ *экстериоризируется* в качестве технического устройства. ИИ – это все еще человеческий интеллект, но принципиально не во всей его полноте. Данный взгляд на ИИ не претендует на статус решения всех злободневных проблем, связанных с ИИ, но обращает внимание на некоторые «выпавшие детали» его традиционного анализа.

В развитии наших доводов мы в первую очередь сосредоточимся на понятии машины и «машинности» (параграф «Машина»). Внимание к инженерному праксису выступает принципиальной методологической установкой нашего подхода к ИИ. Далее будет предложен анализ логики и математики как ИИ, а также «переходного этапа» к «машинизации» этих компетенций (параграф «Естественный искусственный интеллект»). В параграфе «Мозг и машина» будет рассмотрено, какие компетенции ЕИ экстериоризированы в парадигме технического коннекционизма. В заключение отметим роль такой компетенции ЕИ, которая не имеет и не может получить своего «машинного» аналога.

Машина

Современная философия часто склонна максимально расширять область значения слова «машина» – вплоть до признания тождества понятий «машина» и «реальность». Такой подход свойствен, например, философии Ж. Делеза и Ф. Гваттари. По ряду причин, говорить о которых

здесь нет возможности (этому посвящены отдельные исследования), мы не считаем валидным такое максимальное онтологическое обобщение. Однако не поддерживаем и противоположную позицию, согласно которой машина – это только техническое устройство. Последняя установка является одним из проявлений убеждения о чуждости машины по отношению к человеку. Между этими крайними взглядами располагается концепция Ж. Симондона, которая сохраняет за машиной право на собственную природу и при этом реализует оптику непрерывности: машина воспринимается посредником между природой и человеком; машина есть результат трансдуктивного процесса творчества, который приводит к технической индивидуации (конкретизации), когда гетерогенные среды подчиняются единству в акте изобретения [2].

Машина – это среда непрерывности не только между природой и человеком, но и в онтологическом смысле: машина является практическим медиатором к неналичеству состоянию реальности в пределах локальной ситуации. Само слово «машина» предполагает такое положение вещей: μηχανή означало уловку, возможность нарушить естественный ход вещей – причем без указания на природу этой уловки; важна лишь «хитроумность» изобретения. Поэтому уже изначально μηχανή относилось и к техническому объекту, и к сложному концептуальному построению (например, дьявольским хитростям) [3].

Понятие «механизм» раскрывает специфику машины. В теории машин механизм понимается как система материальных тел, предназначенных для преобразования движения одного или нескольких тел в требуемые движения остальных. Русские ученые-механики заложили основы теории механизма, активно развивая концепт степеней свободы кинематической пары. А.П. Малышев вслед за П. И. Сомовым дает определение механизма: «... механизмом называется такая кинематическая цепь, в которой все точки описывают определенные траектории, т.е. каждое звено которой не имеет больше одной степени свободы» [4, с. 10]. При этом исследователь подчеркивает, что «в природе фактически движение происходит только с одной степенью свободы. В самом деле, если точка,двигающаяся в плоскости, обладает двумя степенями свободы, то это значит, что она может занять какое угодно положение в этой плоскости. <...> в нашем пространстве есть движение только с одной степенью свободы» [4, с. 9]. В чем же тогда заключается специфика машины? Исследователь поясняет: «... получаемая в механизме одна степень свободы обеспечивается только теми условиями связи, которые нам известны и которыми мы распоряжаемся. Если этого не будет, то не будет и механизма, хотя бы в цепи была только одна степень свободы под влиянием каких-нибудь условий связи, назначенных природой» [4, с. 10]. То есть подчеркивается, что фиксация степеней свободы в инженерном творчестве – это ситуация анте-, а не постфактум; это ограничение благодаря выстраиванию заданных условий связи.

На основе этого классического понимания механизма можно предложить обобщенное определение машины: «Машина – это устройство, в котором осуществляются предусмотренные проектом детерминированные дискретные и конечные процессы, их целью являются предопределенные действия для решения поставленных задач» [5, с. 30]. Что в полной мере относится и к некоторым «уловкам» ментальной природы.

Естественный искусственный интеллект

Ж. Симондон говорит о когнитивной схеме Декарта по аналогии с «простой машиной»: «В картезианском механизме фундаментальная работа простой машины аналогична функционированию логической мысли, способной быть строгой и продуктивной» [6, р. 17]. Логическая связность в строгом – формальном – смысле является искусственным режимом когитации, который не тождествен просто сосредоточенности внимания, но для которого характерно

помещение интенциональной предметности в уже существующую систему самодвижения мысли. Для этой системы характерны дискретность элементов и строгий детерминизм: «если, то». Дискурсивность осуществляется в силу того, что элементы ограничивают степени свободы друг друга, чтобы траектория движения мысли образовывала стабильные и воспроизводимые паттерны. Именно благодаря их воспроизводимости эта когнитивная система может быть «отчуждена» от интеллекта конкретного человека и передана другому без потери функциональности. Этот искусственный режим сознания связан с особой языковой работой, получившей название «логика» и «математика».

В любом куррикулуме с помощью этих форм знания разум приобретает «уловку» для максимально эффективного решения задач – тех, которые могут быть сформулированы на соответствующем языке (языках). Неслучайно, к примеру, что в рамках образования математика прямо характеризуется как ИИ: «В указанном языковом смысле математику следует считать искусственной составляющей естественного интеллекта, развиваемой самим интеллектом для оптимизации своей деятельности» [7, с. 9]. А следующее обобщение автора можно отнести и к логике: «Можно считать, что математика создает искусственный интеллект, который развивается параллельно естественному интеллекту и в определенном смысле оптимизирует работу последнего» [7, с. 60]. Сознательно усилим это положение: логика – это не только (и не столько) про правильность, сколько про «машинность».

Во-первых, логическое умозаключение является не единственным способом прийти к правильному выводу – можно вспомнить, к примеру, про феномен интуитивного озарения. Во-вторых, как подчеркивает А.Г. Кислов, представление «о почти религиозной святости либо логических законов (что устарело), либо методов (обычно – теоретико-множественных) построения логических систем» можно характеризовать как "незрелое"» [8, с. 145]. Подобным же образом устарела и пифагорейская «святость» математики. Нравится нам это или нет, но логика и математика сегодня – это в первую очередь инструменты такой когнитивной активности, в которой мыслительные действия – операции – могут *автоматически* вести к правильным выводам и/или отсекают неправильные. Таким же образом работает с энергией человек, выстраивающий механизм из сцеплений шестеренок.

Логико-математическая и техническая форма знания часто выступают единым узусом целерационального действия. В. Г. Горохов показывает, как логика и математика становились метаязыком систем автоматического регулирования: релейно-переключающие схемы выстраивались как аналогии логических операций. А современное воплощение логики в технике показано на примере реализации логических элементов «не», «или», «и» с помощью молекулярных переключателей. Исследователь подчеркивает: «... констелляция техники и логики запрограммирована в самой природе технической теории». [9, с. 99]. Что закономерно: основная задача и логики, и техники – выстроить (то есть произвести искусственным / искусным образом) контролируемые русла энергии. Наиболее ранняя форма логики – силлогистика Аристотеля – как раз об этом.

Будучи «изъятими» из стихии свободно текущей мысли и сформулированными в качестве строгих конструкций, силлогизмы Аристотеля, по его же замыслу, выступали способом выявить и взять под контроль ту *вынуждающую силу*, которая заставляет принимать определенные положения в качестве истинных. «... силлогизм же есть высказывание, в котором при утверждении чего-либо из него необходимо вытекает нечто отличное от утвержденного ...»; «под словами "вытекает благодаря этому"» подразумевается «автоматичность», так как «оно не нуждается ни в каком постороннем термине, чтобы следовать с необходимостью» [10, с. 10].

Метод сведения к противоречию (+ закон исключения третьего) – это базовый «механизм», реализующий ограничение степеней свободы в «само-движении» мысли. Сам Аристотель назвал его «органом», то есть «инструментом», но вполне можно скорректировать это название в пользу слова «машина», если учитывать доводы К. Маркса. Маркс критиковал подход, согласно которому «простые механические средства как рычаг, наклонную плоскость, винт, клин и т.д. называют машинами» [11, с. 377], и подчеркивал, что за процессами «рабочих машин» стоит *двигатель*. «Двигатель» приобретает и определенным образом выстроенная мысль/высказывание; фразы «необходимо вытекает», «следовать с необходимостью» – рефрены Первой аналитики.

Детерминированность, дискретность, вынуждающая сила вывода и т.д. – принципы работы интеллекта в особом – искусственном (искусном) – режиме. В одной из ключевых работ по ИИ, прежде чем сосредоточиться на машинных алгоритмах, авторы рассуждают об «оптимизации» и «алгоритмизации» мышления человека с помощью байесовских и т.п. методов. Так мыслится история «рациональных агентов», которыми могут выступать люди или машины [12].

Выше уже говорилось об экстерниоризации логики и математики в другое сознание благодаря процессу обучения. Но с самого начала развития этих когнитивных тактик предпринимались усилия и по их экстерниоризации в техническое устройство: это арифмометры и логические машины. Они наглядно передают процесс перевода рассуждений в движения спиц, реек и шестеренок. «Если установленный силлогизм верен, то вращение рукоятки, продолжающей ось А, вызовет медленное вращение всех связанных ремнями колес», – так В.В. Шилов описывает конструкцию одной из логических машин – машины Пасторе [13]. Таким образом, «внутренний» ИИ становился «внешним» задолго до появления феномена компьютерного ИИ. Который, однако, не имел бы шанса появиться без этой долгой подготовительной стадии.

В ходе этой эволюции ИИ особенно важны не столько мощные жесты, сколько малые смещения смысла. Одно из таковых выступило катализатором создания «внешнего» ИИ, и именно оно ответственно за идею об «отсутствии» исторических перспектив у человеческого интеллекта.

Речь о нововведении стоиков, которые предложили само слово «логика» (приоритет отдается Хриссипу) для обозначения науки о *законах мыслительной деятельности*. То есть целое оказалось сведенным к части; *весь* интеллект оказался сведен к его искусственному режиму (модусу). Рационализм Нового времени явился следующим важным пунктом утверждения этой позиции.

Невозможно обойти вниманием роль исследований Г.В. Лейбница в процессе машинной экстерниоризации «внутреннего» ИИ. Его «умная машина» должна была стать итогом исследований по математизации логики, универсальной нотации, искусственного языка и исчислению умозаключений. Замена обычных рассуждений машинным оперированием знаками, по мысли Лейбница, дало бы человечеству возможность «наведения порядка в мыслях» в ситуации спора или непонимания. Н. Винер, который в числе многих других высоко оценивал *calculus ratiocinator* Лейбница, говорил о том, что «... тот же самый умственный толчок, который привел к развитию математической логики, одновременно привел к гипотетической или действительной механизации процессов мышления» [14, с. 57]. Кроме того, метафизика Лейбница, помноженная на математику, привела философа к разработке темы двоичного исчисления: «Я пояснил эти предметы (*ista non nihil*) образованием чисел с помощью 0 и 1, обнаруженным мною, которое является прекраснейшим изображением непрерывного творения вещей из ничего и зависимости его от Бога» [15, с. 219]. Динамика взаимодействия 0 и 1 у Лейбница не была связана с необходимостью состояний «выкл» и «вкл», выступая концептуальным требованием метафизики творения *ex nihilo*. Однако продуктивность этого хода мысли трудно переоценить в контексте истории развития и компьютера и ИИ на его основе.

Но, пожалуй, еще более важным этапом на этом пути была идея о том, что «наведение порядка в мыслях» – не цель, а *средство* «умной машины». Ада Лавлейс, известная миру как первый программист, должна быть по достоинству оценена как мыслитель, подчеркивающий важность в первую очередь «науки об операциях». Единственная научная работа Ады является, как известно, примечаниями к статье, описывающей аналитическую машину Ч. Беббиджа [16]. И в первом же примечании (Примечание А) акцентируется то, что *не выделяли ни сам Бэббидж, ни популяризаторы его изобретения (они видели в этих машинах лишь инструмент улучшения праксиса математических вычислений)*: «наука об операциях, особенно в том виде, в каком она вытекает из математики, сама по себе является наукой и имеет свою собственную абстрактную истину и ценность». Ада сумела аналитически выявить в тех процессах, которые составляли устройство аналитической машины Бэббиджа, неочевидное, но важнейшее для дальнейшего развития «умных машин» обстоятельство: между теми «математическими процессами», которые «проходят через человеческий мозг» и теми, которые реализуются «посредством неодушевленного механизма», есть промежуточный этап: «рассуждения, связанные с операциями».

Хотелось бы подчеркнуть роль еще одной версии «умной машины», которая явилась важнейшим этапом на пути создания «внешнего» ИИ в его современном модусе. Это машина сравнения идей Г.Н. Корсакова. В отличие от Лейбница Корсаков не стремится создать машину для исправления «несовершенств» мышления человека. Его цель, как и у Бэббиджа, вполне практическая: осуществлять с помощью перфокарт и механизмов рутинные операции обработки больших объемов статистических данных [17]. Исследователи отмечают, что именно С.Н. Корсаков первым предложил вариант машины – предшественницы современного ИИ в виде нейросетей, поскольку она была призвана автоматизировать поиск данных по степени важности определенных критериев (весовых коэффициентов, как сказали бы мы сегодня).

Идея создания машин, «экспроприрующих» у человека функцию интеллекта, сразу же вызвала резкую критику. Императорская академия наук отклонила проект Корсакова со словами: «Г-н Корсаков потратил слишком много разума на то, чтобы научить других обходиться без разума» [18, с. 564]. Похожую реакцию демонстрирует и сатира Д. Свифта: в своем «Путешествии Гулливера» он изображает Большую Академию в городе Лагардо, в которой досужие умы занимаются ненужными проектами, в ряду которых стоит и «работа над проектом усовершенствования умозрительного знания при помощи технических и механических операций» [19, с. 97].

Обратим внимание, что ни вчера, ни сегодня никто не оспаривает положительную роль «внутреннего» ИИ (логики и математики). Конечно, «нечеловеческий» облик формальных структур логики (особенно математической логики) вызывает определенный дискомфорт, но ему далеко до неприятия экстерииоризации логико-математической компетенции в виде технического устройства. Что вполне можно понять: в машинной форме опции интеллекта неизбежно гипертрофируются и гипостазированы. К этому «чужому» культура испытывает амбивалентные, но выступающие сторонами одной медали чувства: демонизацию, с одной стороны, апологетику – с другой. Возможно, если бы изначально речь шла не о человеческом интеллекте как таковом, но о его *искусственной части*, мы имели бы совсем другой дискурс ИИ. Проверить это мы уже не можем, но можем констатировать: сведение части к целому в отношении интеллекта немало способствовало формированию прямой аналогии между компьютером и сознанием человека – что, в свою очередь, выступило толчком к мощному развитию современного ИИ.

Мозг и машина

Установка, согласно которой как таковой интеллект человека реализуется в логико-математической форме, неизбежно формировала понимание интеллекта в логике дефицитов. Которые требовали восполнения. Примером может выступать концепция «пределов рациональности» Г. Саймона, который рассуждает о том, что машинное мышление практически идентично мышлению человека; отличие заключается лишь в «предельности» интеллекта человека: «... помимо стратегий, единственной характеристикой человека, ярко проявляющейся при решении криптоарифметических задач, является ограниченная емкость его кратковременной памяти» [20, с. 47]. Согласно его доводам, человек стремится к «чистой рациональности», но ограничен по параметрам вычислительной мощности и объему памяти.

Когда в 40-х годах XX века начали утверждаться компьютерная метафора сознания и парадигма коннекционизма (от лат. «связь») в инжиниринге ИИ, сведение интеллекта к логико-математическим структурам и идея «дефицитов» интеллекта человека выступали путеводной звездой. У.С. МакКаллока и У. Питтса, описавших акт единичного вычисления, производимого мозгом и являющегося квантом активности одного нейрона, занимал вопрос, как мозг и передача импульсов в нем выстраивает логические выводы. Концепция «перцептрона» Ф. Розенблатта, в которой акт восприятия мозгом информации рассматривался уже в парадигме вероятностного подхода, исходила из соображений вычислимости (эффективности использования ресурсов «биологической машины» мозга). На пересечении линий развития вычислительной модели мозга и развития электронно-вычислительных машин родилась идея искусственной нейронной сети на основе компьютерных технологий. В 1947 году Винер писал: «... нам стало ясно, что сверхбыстрая вычислительная машина, поскольку вся она строится на последовательном соединении переключающих устройств, является идеальной моделью для решения задач, возникающих при изучении нервной системы» [14, с. 60].

Поскольку исследователи пришли к убеждению, что для организма кодировать каждый акт возбуждения каждого нейрона оказывается энергетически слишком затратным, инженерные решения в сфере ИИ также отошли от выстраивания формализмов, кодирующих каждый вычислительный акт. Формировалось принятие того обстоятельства, что работать придется с хаосом реального. Тем более что этому способствовала и эволюция логики: «... область "логического" не остановилась на рассмотрении форм взаимоотношений между знаками (логический синтаксис), а расширилась до анализа форм отношений знаков и реальности (логическая семантика), форм отношений носителей языка к знакам и форм взаимоотношений между самими носителями языка (логическая прагматика)» [8, с. 142]. Была поставлена задача, в определенном смысле обратная интенции сведения целого интеллекта к его части: если после стоиков человеческий интеллект выступал таковым в той мере, в которой был похож на построения логики, то теперь сама логика начала уподобляться «необработанному», естественному человеческому интеллекту. Но это все-таки была именно логика, и потому ее формализмы смогли стать механизмами машины. Начиная с трехзначной логики Я. Лукасевича, возникают системы неаристотелевской¹ логики; Л. Заде вводит новый математический аппарат, который переводит неоднозначные утверждения естественного языка на язык математических формул. Закладывался фундамент новой инженерной парадигмы ИИ, связанной с вероятностными алгоритмами и нечетким управлением.

Несомненные успехи этой парадигмы ИИ, усиленные нерелефлексированной убежденностью в том, что инжиниринг «умных машин» имеет дело со *все*м интеллектом, а не его частью, привело

¹ Сам Лукасевич называет эту логику нехрисипповой.

к уверенности: наука ИИ создала машину, которая по своему функционалу уже совпала с границами человеческого интеллекта и вот-вот выйдет за эти пределы. Такую эволюцию ИИ рисует, например, компания Сбер на одной из площадок Интернета [21]: 1) ИИ на основе правил; 2) контекстный ИИ; 3) специализированный ИИ; 4) думающий ИИ (тот, что победил в чемпионате мира по игре в Го); 5) генеративный ИИ (Chat GPT и др.), который, как сказано, «имеет глубокое понимание мира, включая абстрактные концепции и сложные человеческие эмоции». Далее постулируется плавное вхождение в сферу суперинтеллекта / самосознающего ИИ.

Подобные «классификации» вряд ли делают сферу ИИ прозрачнее. Скорее, наоборот, вносят еще больший дисбаланс в понимание сути и роли ИИ. Не только потому, что абсолютно не ясны основания деления данной классификации, но и в силу бесперспективности той полемики, которую она инициирует. Она сводится по преимуществу к вопросам, что *действительно* умеет тот или иной ИИ, свидетельствуют ли ответы Chat GPT о "глубоком понимании мира"? и т.д.

Наша тактика в поиске ответов на вопрос о сходстве и различии человеческого интеллекта и ИИ будет опираться на изложенные выше доводы о сущностной нерасторжимой связи технического ИИ и человеческого ЕИ. В контексте внимания к инженерным решениям мы предлагаем сопоставить существующие механизмы ИИ с компетенциями человеческого интеллекта. «Остаток» компетенций (если таковой будет обнаружен) будет выступать контекстом и основанием для решения вопроса о специфике человеческого ЕИ.

Классическая и неклассическая машина ИИ

На сегодняшний день рабочие версии ИИ представлены двумя базовыми инженерными моделями – механизмами:

1) ИИ на основе правил, то есть ИИ как вычислительная система на базе строгого детерминизма + словарь. Эта модель предполагает создание алгоритмов, которые бы осуществляли вычисления, соответствующие процессам «зрение», «чтение», «восприятие языка» и т.д.;

2) ИИ на основе самообучающихся нейросетей: вероятностные алгоритмы + базы данных. В этом случае инженеры уже не пытаются познать и воспроизвести глубинные структуры мышления, зрения, языка и т.д, а создают машины, которые самостоятельно добывают ответы на поставленные запросы. Для этого нейронная сеть сначала обучается разработчиками на специально подобранных примерах (обучающих датасетах), а далее она отпускается в некое пространство «дикого множества» знаков (изображений, звуков, слов и т.д.), в котором продолжает обучаться самостоятельно. (Стоит подчеркнуть, однако, что самостоятельность эта весьма условна. На практике любую нейросеть постоянно проверяют на адекватность и оснащают фильтрами по этическим и политическим соображениям).

Первая стратегия – это классическая машина, о которой сказано в самом начале. Вторая стратегия – это машина, которая на низком уровне организации остается все той же системой строго детерминированных процессов (классической), а на более высоком делает хаос элементом своего механизма.

В качестве одного из своих смыслов хаос – это неуправляемость. Этот момент выступает поводом для серьезных беспокойств и, соответственно, исследований. Речь о том, что нейросети для оптимизации своей работы формируют аналогии человеческих абстракций, однако, как показывает работа L. Bereska и E. Gavves, «они не обязательно должны соответствовать понятиям, интерпретируемым человеком» и, соответственно, «могут становиться все более абстрактными, кодируя информацию способами, противоречащими человеческой интуиции». С точки зрения ин-

женерии это выглядит как «полисеманτικότητα нейронов», когда «например, один и тот же нейрон может активироваться как при виде кошек, так и при виде автомобилей» [22].

Невозможность прямо управлять тем, как информация кодируется в ходе внутренней самореорганизации нейросети, означает потерю человеком полного контроля за процессами оптимизации. Как показано в работе J. Engels и др. [23], нейронные сети могут формировать многомерные пространства, поскольку внутренние абстракции мешают друг другу из-за того, что количество признаков² больше, чем фактическое количество нейронов. Опасения в отношении «самостийности» внутренних представлений нейронных сетей – не повод говорить о «самостийном» интеллекте машины. Непредусмотренные представления нейросети возникают в результате оптимизации как неизбежного требования экономии ресурсов. Но то, что для самой машины выступает снижением беспорядка, для человека часто оказывается воплощением хаоса. В силу чего множатся исследования, направленные на изменение сложившейся парадигмы анализа работы нейросетей. «В современном машинном обучении ИИ "выращивается", а не проектируется», – сетуют авторы проекта MIRI [24] и настойчиво предлагают, если можно так выразиться, сделать современную машину ИИ более *машинной*. На данный момент, подчеркивают исследователи, мы располагаем только методом наблюдения за поведенческими свойствами нейросети, тогда как необходимы методы, позволяющие внедрить в систему определенные внутренние свойства или проверить их наличие.

Но даже в случае успеха этой стратегии хаос, связанный с работой новой машины, сохраняется. Речь идет уже не об отсутствии контроля, а о необходимой составляющей LLM³: взаимодействие нейросети с «диким множеством». Речь о запросах пользователей или самостоятельная обработка нейросетью контента Интернета, ситуация вождения в реальных условиях (беспилотный транспорт) и т.п. – эти другие «необработанные» данные являются элементом механизма новой машины. Поэтому, анализируя машину, основанную на технологии искусственной нейронной сети, можно констатировать появление неклассической машины – вслед за неклассическими философией, логикой, научной картиной мира⁴ и т.д.

Сегодня ведется активная полемика в отношении того, является ли нейросеть как инженерная стратегия тем, о чем изначально мечтали создатели ИИ. Критики уверены, что нейросеть не «думает», поскольку уже не копирует мыслительный процесс человека. Алгоритмы, как мы говорили, умеют копировать *логические* умозаключения человека (тот самый «внутренний ИИ»), но попытки копировать *не искусственные* когнитивные процедуры ЕИ приводят к неподъемному усложнению алгоритмов. Поэтому алгоритмы обрабатывают большие данные, отбирая максимально вероятные комбинации знаков. Эта инженерная стратегия резко критикуется за «предательство» самой идеи думающей машины. В частности, эта дилемма составляет суть спора Н. Хомского и П. Норвига (инженера и директора по инновациям в компании Google). Отвечая на озвученные Хомским разочарования, Норвиг говорит: «... инженерный успех показывает, что что-то работает правильно, и поэтому является свидетельством (но не доказательством) научно успешной модели» [25]. Иными словами, это работает как интеллект, значит, это интеллект. Однако общим и у той, и у другой позиции является то, что они говорят обо *всем* интеллекте и выбирают ту или другую стратегию для его машинного «воссоздания». Однако вторая стратегия, как и первая, имитирует не весь интеллект, но вновь лишь его часть – уже другую. Спора, по су-

² Признаки (Features) – фундаментальные единицы, с помощью которых нейросети кодируют знания

³ LLM – Большая языковая модель – программа, распознающая естественный язык и генерирующая тексты.

⁴ В хронологически более точном смысле речь должна идти о постнеклассической науке, но в данном случае неклассичность означает как таковое противостояние с классической научной картиной мира.

ти, вообще быть не может. *Классическая машина – это экстериоризация нашей способности к логическим/математическим построениям. Неклассическая машина – экстериоризация нашей способности оперировать коллективным опытом.*

Социальная вовлеченность, характерная для человека с младенческого состояния, и принцип интерсубъективности формируют богатый арсенал знаний (явных и неявных) и нерелексируемых установок, которые единообразно воспроизводятся многими людьми. Не составляет особого труда обнаружить в открытой информационной среде и формализовать в качестве статистических закономерностей повторяемые запросы, реакции пристрастий и т.п. группой людей или одним человеком (что является в том числе основой технологий контекстной и таргетированной рекламы). Сам человек может не фиксировать некоторые формы активности именно в силу их привычности, рутинности, но у машины не «замыливается глаз» на повторяющиеся паттерны, и ее память ничего не «стирает». Собственно, именно поэтому специализированные (то есть обученные для конкретного вида деятельности) нейросети оказываются очень полезным и тонким инструментом. Примером может быть анализ черновых рукописей Пушкина сетью «Да Винчи» [26]. Или ИИ-сопровождение деятельности комбайнера, когда ИИ отвечает за рутинную операцию удерживания конфигурации кромки. Можно привести в пример и программу AlphaGo (Google DeepMind, 2015 г.), которая одержала победу над профессиональным игроком в го (Ли Седолем). Обычно ее приводят в пример в качестве довода о способности ИИ к «эксклюзивному думанию». Но стоит обратить внимание на механизм этой машины: эта нейросеть использовала два набора данных: опыт большого количества игроков в го и случайно сгенерированные законные ходы. То есть Ли Седоль, по сути, играл с коллективным разумом игроков в го.

Способность воспроизводить коллективные паттерны – это и ответ на вопросы по поводу того, как же нейросетям удастся создавать художественные образы. Все многообразие образов художественной литературы, визуальных и музыкальных жанров искусства выстраивается на основе небольшого количества снова и снова повторяющихся нарративов, стилистических приемов, лингвистических и музыкальных фраз – это своего рода «библиотека» коллективного художественного сознания. Ключевую роль повтора в генерировании художественного текста человеком анализирует, например, Г.В. Векшин [27].

Совсем недолго осталось ждать и полноценной ИИ-обработки различных контекстов и эмоций, поскольку и то, и другое являются принципиально повторяющимися паттернами и также представлены в «диком множестве» контента Интернета. Одна из разработок ИИ, проанализировав 6 миллионов видео из 144 стран, выявила только 16 типов мимики, соответствующих определенным контекстам [28]. Нравится нам это или нет, но ИИ «доберется» до любых паттернов коллективного опыта, поскольку они существуют в знаковой форме в открытом доступе.

Таким образом, этот модус ИИ выступает экстериоризацией еще одной *части* нашего интеллекта – той, которая ответственна за думание «сообща», за ситуации, в которых мы «сливаемся» в неразличимости типичного/коллективного. Это важная часть нашего интеллекта, поэтому нейросети нашли применение в самых разных сферах нашей жизни.

Интересно, что в «не-западной» традиции также выявлена связь ИИ с компетенциями человеческого интеллекта. В работе, посвященной преломлению этических проблем ИИ в новой китайской научной фантастике, автор рассказывает о фантастическом повествовании, в котором те же две парадигмы ИИ (на основе правил и машинное обучение) соотносятся с подходами в китайской философии. Эти два способа думания (архетипы) и два базовых варианта механизмов ИИ

дополняются третьим архетипом, для которого на практике пока нет соответствующих машинных механизмов, но они существуют в воображении писательницы (и программиста) Му Мин: «... повествование исследует возможности конфуцианства, моизма и шаманских традиций царства Чу в объяснении создания искусственного интеллекта» [29, с. 42]. Однако подчеркнем, что в этом интересном исследовании все-таки говорится о *сходстве* когнитивных практик и механизмов ИИ. Мы же говорим о *переносе*: о воплощении когнитивных техник в технические артефакты.

Остается ли в нашем интеллекте что-то «за вычетом» этих экстериозированных частей? В качестве заключения хочется подчеркнуть ряд моментов, которые не позволяют «Ахиллесу» ИИ догнать «черепаху» интеллекта человеческого.

В пределах заявленной методологии мы должны анализировать некие метакомпетенции по отношению к тем, которые человек превратил в механизмы ИИ. И таковые обнаруживаются сразу же в важнейшем элементе инжиниринга ИИ: в обучении-с-человеком (обучение нейросети, предшествующее самообучению). Стоит обратить внимание на те когнитивные процедуры, которые составляют данную мета-машинную компетенцию.

Думается, важнейшей из них является естественная для человека способность менять оптику вероятностного прогнозирования: человек выделяет типичные и нетипичные ситуации, тогда как алгоритмы «заточены» либо на поиск типичного (максимально вероятного), либо на поиск нетипичного (минимально вероятного, исключительного, уникального). Но само «переключение» этих режимов, как выяснилось в ходе развития экспертных систем, является принципиально *неформализуемым* знанием экспертов-людей и потому зоной принципиальных трудностей, с которыми столкнулись инженеры по знанию [30]. Не просто обучая машину, но *учась обучать машину* – а именно это сегодня наиболее активно развивающаяся область индустрии ИИ, – человек неизбежно усложняет и свой мета-машинный интеллект.

Кроме того, в мире, где «коллективная разумность» стала функционалом машины-нейросети, человек обречен на развитие индивидуальной креативности. Спустя очень небольшое время автора текста, который наполнен штампами и тривиальным умозаключениями, наверняка будут обвинять в использовании нейросетей. Или искать возможность заменить его этой нейросетью. Человек будет вынужден развивать себя «от противного» – от функционала ИИ.

И, наконец, любую машину требуется чинить. Возникнет пул ремонтных бригад ИИ. То есть людей, у которых по определению больше степеней свободы для действий, чем у машины. Потому что починка машины – это не то же самое, что восстановление изначального состояния. Это почти всегда создание «обходного маневра» по отношению к процедурам, составляющим механизм самой машины. «Изолента» всегда будет в ходу у инженеров и, думается, это отличная новость для человеческого интеллекта.

Выводы

Радикальное противопоставление человеческого интеллекта и ИИ (причем в пользу любого из них) при ближайшем рассмотрении не является философски валидным. Мы имеем только человеческий интеллект – машинный он или нет. В качестве функций машины экстериоризируются функции (части) естественного интеллекта. Ранее всего это опредмеченные в механизме структуры логики и математики; сегодня активно экстериоризируется исчисление статистически значимых (коллективных и индивидуальных) паттернов мыследеятельности. При этом речь идет принципиально о *части* интеллекта – как минимум потому что обучение, ремонт и концептуальное развитие машин требует и всегда будет требовать развития метафункционала всего человеческого интеллекта.

Список литературы

1. Cole, S. AI Company Asks Job Applicants Not to Use AI in Job Applications. Feb. 3. 2025 [Электронный ресурс] / S. Cole. – URL: <https://www.404media.co/anthropic-claude-job-application-ai-assistants/> (дата обращения: 23.06.2025).
2. Куртов, М. Жильбер Симондон «О способе существования технических объектов» [Электронный ресурс] / М. Куртов // Транслит. – № 9. – URL: <http://www.trans-lit.info/materialy/9-vypuski/zhilber-simondon-o-sposobe-sushhestvovaniya-tehnicheskikh-obektov> (дата обращения: 23.06.2025).
3. Зарипов, Я.В. Μηχανή: анализ семантики в диахронической перспективе [Электронный ресурс] / Я.В. Зарипов // Материалы Всероссийской школы молодых византистов NEANIA, 2015. – URL: <https://www.hse.ru/data/2015/10/03/1077230538/Зарипов%20Я.В.pdf> (дата обращения: 23.06.2025).
4. Малышев, А.П. Анализ и синтез механизмов с точки зрения их структуры [Электронный ресурс] / А.П. Малышев // Известия Томского Технологического Института [Известия ТТИ], 1923. – Т. 44. – URL: <http://earchive.tpu.ru/handle/11683/7368> (дата обращения: 23.06.2025).
5. Оболкина, С.В. Онтология машины. Блеск и нищета машинизма / С.В. Оболкина // Антиномии. – 2023. – Т. 23, вып. 3 – С. 20–41. DOI: 10.17506/26867206_2023_23_3_20
6. Simondon, G. Technical mentality / G. Simondon // Parrhesia. – 2009. – No. 7. – P. 17–27.
7. Селезнев, В.А. Элементы математического формализма для филологов: учеб. пособие / В.А. Селезнев. – Новосибирск: Изд-во НГТУ, 2000. – 68 с.
8. Кислов, А.Г. Социокультурная кроссидентификация логики / А.Г. Кислов // Научный ежегодник Института философии и права Уральского отделения Российской академии наук. – 2011. – Вып. 11. – С. 134–149.
9. Горохов, В. Логика и техника: от теории электрических цепей к наносистемотехнике / В. Горохов // Логические исследования. – 2012. – Т. 18. – С. 97–126.
10. Аристотель. Аналитики. Первая и вторая. – Москва: Гос. изд-во полит. лит., 1952. – 437 с.
11. Маркс, К. Капитал. Критика политической экономии. Т. 1 / К. Маркс. – Москва: ОГИЗ Госполитиздат, 1952. – 794 с.
12. Рассел, С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. – Москва: Вильямс, 2006. – 1408 с.
13. Шилов, В.В. История логических машин [Электронный ресурс] / В.В. Шилов // Информационные технологии. – 2004. – № 6. – URL: <https://nsu.ru/xmlui/bitstream/handle/nsu/9043/shilov.pdf;jsessionid=3CC455C99BC0551D2FAC1BC6B6C9B92C?sequence=1> (дата обращения: 23.06.2025).
14. Винер, Н. Кибернетика или управление и связь в живом и машине / Н. Винер. – 2-е изд. – Москва: Советское радио, 1968. – 328 с.
15. Лейбниц, Г.В. Письма и эссе о китайской философии и двоичной системе исчисления / Г.В. Лейбниц. – Москва: ИФ РАН, 2005. – 404 с.
16. Menabrea, L.F. Invented by Charles Babbage. With notes upon the Memoir by the Translator Ada Augusta, Countess of Lovelace [Электронный ресурс] / L.F. Menabrea // Bibliothèque Universelle de Genève, October, 1842, No. 82. – URL: https://translated.turbopages.org/proxy_u/en-ru.ru.eedb0aca-6851856e-002fb6df-74722d776562/https/web.archive.org/web/20080915134651/http://www.fourmilab.ch/babbage/sketch.html#NoteA (дата обращения: 23.06.2025).
17. Корсаков, С.Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / С.Н. Корсаков. – Москва: МИФИ, 2009. – 44 с.

18. Историко-математические исследования. – Вып. 14. – Москва: Гостехиздат, 1961. – 635 с.
19. Свифт, Д. Путешествия Гулливера [Электронный ресурс] / Д. Свифт. – 1727. – 178 с. – URL: <https://olehnik.ru/files/literatura/2/3.%20Свифт%20Путешествия%20Гулливера.pdf> (дата обращения: 23.06.2025).
20. Саймон, Г. Науки об искусственном / Г. Саймон. – Изд. 2-е. – Москва: Едиториал УРСС, 2004. – 146 с.
21. Колосова, А. Эволюция ИИ и что его ждет в будущем. 28 февраля 2025 [Электронный ресурс] / А. Колосова. – URL: <https://habr.com/ru/companies/sberbank/articles/886806/> (дата обращения: 23.06.2025).
22. Bereska, L. Mechanistic Interpretability for AI Safety – A Review [Электронный ресурс] / L. Bereska, E. Gavves // Transactions on Machine Learning Research (08/2024). – URL: <https://leonardbereska.github.io/blog/2024/mechinterpreview/> (дата обращения: 23.06.2025).
23. Not All Language Model Features Are Linear [Электронный ресурс] / J. Engels, I. Liao, E.J. Michaud, W. Gurnee, M. Tegmark // [PDF] – CoRR. – 2024. – URL: https://www.researchgate.net/publication/380847625_Not_All_Language_Model_Features_Are_Linear (дата обращения: 23.06.2025). DOI: 10.48550/arXiv.2405.14860
24. Machine Intelligence Research Institute (MIRI) [Электронный ресурс]. – URL: <https://intelligence.org/the-problem/> (дата обращения: 23.06.2025).
25. Norvig, P. On Chomsky and the Two Cultures of Statistical Learning [Электронный ресурс] / P. Norvig // Peter@Norvig.com. – URL: <http://norvig.com/chomsky.html#:~:text=But%20I%20observe%20that%20science,can%20progress%20on%20its%20own> (дата обращения: 23.06.2025).
26. Сурабекянц, С. Российские специалисты из Smart Engines расшифровали рукописи Пушкина при помощи ИИ. 06.02.2025 [Электронный ресурс] / С. Сурабекянц. – URL: <https://3dnews.ru/1117862/spetsialisti-rossiyskoy-kompanii-smart-engines-rasshifrovali-rukopisi-pushkina-pri-pomoshchi-ii> (дата обращения: 23.06.2025).
27. Векшин, Г.В. Очерк фоностилистики текста: Звуковой повтор в перспективе смыслообразования / Г.В. Векшин. – Москва: Моск. гос. ун-т печати, 2006. – 461 с.
28. Sixteen facial expressions occur in similar contexts worldwide / A.S. Cowen, D. Keltner, F. Schroff [et al.] // Nature. – 2021. – № 589. – С. 251–257. DOI: 10.1038/s41586-020-3037-7
29. Юань, Фэн. Воображение и этика искусственного интеллекта: исследования на примере современной китайской научной фантастики / Юань Фэн // Технологос. – 2024. – № 4. – С. 31–47. DOI: 10.15593/perm.kipf/2024.4.03
30. Оболкина, С.В. Экспертиза vs экспертная оценка: понятийные основания компетенций / С.В. Оболкина, Н.Г. Попова // Социология науки и технологий. – 2019. – Т. 10, № 4. – С. 38–49.

References

1. Cole S. AI Company Asks Job Applicants Not to Use AI in Job Applications. Feb.3. 2025, available at: <https://www.404media.co/ anthropic-claude-job-application-ai-assistants/> (accessed 23 June 2025).
2. Kurtov M. Zhil'ber Simondon «O sposobe sushchestvovaniia tekhnicheskikh ob'ektov» [Gilbert Simondon "On the Mode of Existence of Technical Objects"]. *Translit*, no. 9, available at: <http://www.trans-lit.info/materialy/9-vypuski/zhilber-simondon-o-sposobe-sushchestvovaniya-tehnicheskikh-obektov> (accessed 23 June 2025).
3. Zaripov Ya.V. Μηχανή: analiz semantiki v diakhronicheskoi perspective [Μηχανή: analysis of semantics in diachronic perspective]. *Materialy Vserossiiskoi shkoly molodykh vizantinistov NEANIA*, 2015, available at: <https://www.hse.ru/data/2015/10/03/1077230538/Zaripov%20Ia.V.pdf> (accessed 23 June 2025).
4. Malyshev A.P. Analiz i sintez mekhanizmov s tochki zreniia ikh struktury [Analysis and synthesis of mechanisms from the point of view of their structure]. *Bulletin of Tomsk Polytechnic University*, 1923, vol. 44, available at: URL: <http://earchive.tpu.ru/handle/11683/7368> (accessed 23 June 2025).

5. Obolkina S.V. Ontologiya mashiny. Blesk i nishcheta mashinizma [Ontology of the machine. The splendor and poverty of machinism]. *Antinomies*, 2023, vol. 23, iss. 3, pp. 20–41. DOI: 10.17506/26867206_2023_23_3_20
6. Simondon G. Technical mentality. *Parrhesia*, 2009, no. 7, pp. 17–27.
7. Seleznev V. A. Elementy matematicheskogo formalizma dlia filologov: Uchebnoe Posobie [Elements of mathematical formalism for philologists: Textbook]. Novosibirsk, Novosibirskii gosudarstvennyi tekhnicheskii universitet, 2000, 68 p.
8. Kislov A.G. Sotsiokul'turnaia krossidentifikatsiia logiki [Sociocultural cross-identification of logic]. *Nauchnyi ezhegodnik Instituta filosofii i prava Ural'skogo otdeleniia Rossiiskoi akademii nauk*, 2011, iss. 11, pp. 134 – 149.
9. Gorokhov V. Logika i tekhnika: ot teorii elektricheskikh tsepei k nanosistemotekhnike [Logic and technology: from the theory of electrical circuits to nanosystems engineering]. *Logical Investigations*, 2012, vol. 18, pp. 97–126.
10. Aristotel'. Analitiki. Pervaiia i vtoraiia [Analytics. First and Second]. Moscow, Gosudarstvennoe izdatel'stvo politicheskoi literatury, 1952, 437 p.
11. Marx K. Kapital. Kritika politicheskoi ekonomii. Tom 1 [Capital. Critique of Political Economy. Vol. 1]. Moscow, Ob"edinenie gosudarstvennykh knizhno-zhurnal'nykh izdatel'stv Gospolitizdat, 1952, 794 p., 377 p.
12. Russell S., Norvig P. Iskustvennyi intellekt: sovremennyi podkhod [Artificial Intelligence: A Modern Approach]. Moscow, Vil'iams, 2006, 1408 p.
13. Shilov V.V. Istoriia logicheskikh mashin [History of logical machines]. *Informatsionnye tekhnologii*, 2004, no. 6, available at: <https://nsu.ru/xmlui/bitstream/handle/nsu/9043/shilov.pdf;jsessionid=3CC455C99BC0551D2FAC1BC6B6C9B92C?sequence=1> (accessed 23 June 2025).
14. Wiener N. Kibernetika ili upravlenie i sviaz' v zhivom i mashine [Cybernetics or control and communication in the living and in the machine]. 2nd ed. Moscow, Sovetskoe radio, 1968, 328.
15. Leibniz G.V. Pis'ma i esse o kitaiskoi filosofii i dvoichnoi sisteme ischisleniia [Letters and essays on Chinese philosophy and the binary system of calculus]. Moscow, Institut filosofii Rossiiskoi akademii nauk, 2005, 404 p.
16. Menabrea L. F. Invented by Charles Babbage. With notes upon the Memoir by the Translator Ada Augusta, Countess of Lovelace. *Bibliothèque Universelle de Genève*, October, 1842, no. 82, available at: https://translated.turbopages.org/proxy_u/en-ru.ru.eedb0aca-6851856e-002fb6df-74722d776562/https/web.archive.org/web/20080915134651/http/www.fourmilab.ch/babbage/sket-ch.html#NoteA (accessed 23 June 2025).
17. Korsakov S. N. Nachertanie novogo sposoba issledovaniia pri pomoshchi mashin, sravnivaiushchikh idei [Outline of a new method of research using machines that compare ideas]. Moscow, Moskovskii inzhenerno-fizicheskii institut, 2009, 44 p.
18. Istoriko-matematicheskie issledovaniia [Historical and mathematical studies]. Iss. 14. Moscow, Gostekhizdat, 1961, 635 p.
19. Swift D. Puteshestviia Gullivera [Gulliver's Travels], 1727, 178 p., available at: <https://olehnik.ru/files/literatura/2/3.%20Свифт%20Путешествия%20Гулливера.pdf> (accessed 23 June 2025).
20. Simon G. Nauki ob iskustvennom [Artificial Sciences]. 2nd ed. Moscow, Editorial URSS, 2004, 146 p.
21. Kolosova A. Evoliutsiia II i chto ego zhdet v budushchem. 28 fevralia 2025 [Evolution of AI and what awaits it in the future. February 28, 2025], available at: <https://habr.com/ru/companies/sberbank/articles/886806/> (accessed 23 June 2025).
22. Bereska L., Gavves E. Mechanistic Interpretability for AI Safety - A Review. Transactions on Machine Learning Research (08/2024), available at: <https://leonardbereska.github.io/blog/2024/mechinterpreview/> (accessed 23 June 2025).
23. Engels J., Liao I., Michaud E.J., Gurnee W., Tegmark M. Not All Language Model Features Are Linear [PDF] – CoRR, 2024, available at: https://www.researchgate.net/publication/380847625_Not_All_Language_Model_Features_Are_Linear (accessed 23 June 2025). DOI: 10.48550/arXiv.2405.14860
24. Machine Intelligence Research Institute (MIRI), available at: <https://intelligence.org/the-problem/> (accessed 23 June 2025).
25. Norvig P. On Chomsky and the Two Cultures of Statistical Learning. *Peter@Norvig.com*, available at: <http://norvig.com/chomsky.html#:~:text=But%20I%20observe%20that%20science,can%20progress%20on%20its%20own> (accessed 23 June 2025).
26. Surabekianz S. Rossiiskie spetsialisty iz Smart Engines rasshifrovali rukopisi Pushkina pri pomoshchi II. 06.02.2025. [Russian specialists from Smart Engines deciphered Pushkin's manuscripts using AI. 06.02.2025.], available at: <https://3dnews.ru/1117862/spetsialisti-rossiyskoy-kompanii-smart-engines-rasshifrovali-rukopisi-pushkina-pri-pomoshchi-ii> (accessed 23 June 2025).
27. Vekshin G. V. Ocherk fonostilistiki teksta: Zvukovoi povtor v perspektive smysloobrazovaniia [Essay on the Phonostylistics of the Text: Sound Repetition in the Perspective of Meaning Formation]. Moscow, Moskovskii gosudarstvennyi universitet pečati, 2006, 461 p.
28. Cowen, A.S., Keltner, D., Schroff, F. et al. Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 2021, no. 589, pp. 251–257. DOI: 10.1038/s41586-020-3037-7
29. Yuan, Feng. Voobrazhenie i etika iskustvennogo intellekta: issledovaniia na primere sovremennoi kitaiskoi nauchnoi fantastiki [Imagination and Ethics of Artificial Intelligence: Case Studies in Contemporary Chinese Science Fiction]. *Technologos*, 2024, no. 4, pp. 31–47. DOI: 10.15593/perm.kipf/2024.4.03
30. Obolkina S.V., Popova N.G. Ekspertiza vs ekspertnaia otsenka: poniatiinye osnovaniia kompetentsii [Expertise vs. expert assessment: conceptual foundations of competencies]. *Sociology of Science and Technology*, 2019, vol. 10, no. 4, pp. 38–49.