

УДК: 165.2

DOI: 10.15372/PS20260202

EDN: VCIMAC

А.А. Сухно, В.В. Гулин

**РОЛЬ МЕТОДОВ МАШИННОГО
ОБУЧЕНИЯ В ЕСТЕСТВОЗНАНИИ:
РАСШИРЕНИЕ КОГНИТИВНЫХ СПОСОБНОСТЕЙ
И «КОМПЕНСАЦИЯ ПРЕДВЗЯТОСТИ»**

В статье рассматривается, какой подход может быть использован для решения проблемы теоретического обоснования методов машинного обучения (ML) в естественных науках. Авторы указывают, что текущая стратегия философско-эпистемологических и прикладных исследований, связанных с ML (борьба с предвзятостью, минимизация субъективных допущений), не в состоянии решить проблему «черного ящика», что затрудняет интерпретацию результатов и снижает их научную ценность. В статье предлагается альтернативный подход к теоретическому обоснованию ML, опирающийся на расширение вычислительных возможностей. Используя концепцию «саморасширения» (extending ourselves) Пола Хамфриса, авторы показывают, как вычислительные технологии позволяют обойти ограничения человеческого мышления и моделировать сложные явления, недоступные «традиционным» математическим методам. Выдвигается идея о *механизме «компенсации предвзятости»*, который может нивелировать влияние субъективных факторов в рамках естественно-научного исследования с использованием ML. Особое внимание уделено сравнению ML с компьютерными симуляциями, где влияние допущений/предвзятости может быть компенсировано за счет анализа глобальной динамики модели, тогда как в ML эта проблема остается нерешенной. Это влечет за собой необходимость отделить проблему «черного ящика» в области ML от «эпистемической непрозрачности», которая является общей для машинного обучения и компьютерных симуляций. Специально подчеркивается, что «черный ящик» в машинном обучении возникает не из-за «непрозрачности» модели, из-за сложности вычислительных операций, а из-за непроясненности связей модели с реальными физическими процессами («целевой системой»). Таким

образом, авторы демонстрируют, что применение ML в естествознании требует переосмысления существующих методологических предпосылок. Разработка механизмов компенсации предвзятости в области ML становится ключевой задачей для устранения проблемы «черного ящика» и успешной интеграции ML в научную практику.

Ключевые слова: машинное обучение; естествознание; эпистемическая непрозрачность; компьютерные симуляции; конструкционные допущения; черный ящик; когнитивные способности; компенсация предвзятости

A.A. Sukhno, V.V. Gulín

**THE ROLE OF MACHINE
LEARNING METHODS IN NATURAL SCIENCE:
EXTENDING COGNITIVE CAPABILITIES
AND “BIAS COMPENSATION”**

The article discusses which approach can be used to solve the problem of theoretical justification of machine learning methods in the natural sciences. The authors point out that the current strategy of philosophical, epistemological and applied research related to ML (countering bias and minimizing subjective assumptions) fail to solve the “black box” problem, which makes it difficult to interpret the results and reduces their scientific value. The article suggests an alternative approach to the theoretical justification of ML based on the extension of computational powers. Using Paul Humphreys’ concept of “extending ourselves,” the authors show how computing technologies can overcome the limitations of human thinking and model complex phenomena that are inaccessible to “traditional” mathematical methods. The idea of a “*bias compensation*” mechanism is put forward, which can neutralize the influence of subjective factors in the framework of natural science research using ML. Special attention is paid to comparing ML with computer simulations, where the influence of assumptions/bias can be compensated by analyzing the global dynamics of the model, whereas in ML this problem remains unresolved. This entails the need to separate the “black box” problem in ML from the “epistemic opacity” that is common to both machine learning and computer simulations. It is specifically emphasized that the “black box” in machine learning arises not from the “opacity” of the model or the complexity of computational operations, but from the lack of clarity of the model’s connections with real physical processes (the “target system”). Thus, the authors demonstrate that the application of ML in natural science requires a rethinking of existing methodological prerequisites. The development

of mechanisms of bias compensation in the field of ML is becoming a key task in order to overcome the “black box” problem and successfully integrate ML into scientific research.

Keywords: machine learning; natural science; epistemic opacity; computer simulations; construction assumptions; black box; cognitive capabilities; bias compensation

Введение

Интеграция естествознания и машинного обучения (Machine Learning, ML) открывает новые горизонты в анализе данных и моделировании сложных систем. Однако это не только приносит новые возможности, но и влечет за собой риски. Речь идет о проблеме теоретического обоснования применения методов машинного обучения в естествознании, которая неразрывно связана с понятием «черного ящика» / «эпистемической непрозрачности» (black box / epistemic opacity). Этот термин обозначает ситуацию, когда внутренние механизмы работы алгоритмов остаются непонятными для исследователей, что затрудняет интерпретацию результатов и снижает доверие к ним. Чем дальше, тем больше кажется, что для современной эпистемологии ML эта проблема превратилась в неразрешимый парадокс, сравнимый с «парадоксом непостижимой/необоснованной эффективности математики» Ю. Вигнера в естествознании [20, р. 44].

Мы полагаем, что одной из причин возникновения этой проблемы является преобладающая стратегия интеграции ML в естествознание – борьба с предвзятостью, стремление минимизировать субъективные допущения в алгоритмах. Это имеет место как в прикладных исследованиях, где влияние допущений снижается за счет «физической согласованности» (physical consistency), внедрения в алгоритмы ML научных знаний¹, так и в эпистемологии ML, где ведется поиск способов обосновать их использование [10; 11; 26]. Такой под-

¹ В явной форме мысль о том, что в алгоритмах ML допущения (assumptions) / предвзятость (bias) заменяются на научные знания, выражена у Дж.Э. Карниадакиса и др. [15, р. 424]. Но, по сути, любые инициативы в прикладных исследованиях, ориентированные на интеграцию методов машинного обучения с физическими теоретическими моделями, которые позволяют «сократить пространство поиска решений в ML-моделях» [27, р. 11] или «имеют больше шансов представить причинно-следственные связи... чем модели, основанные исключительно на дан-

ход подразумевает рассмотрение методов ML как очередного математического инструмента, дополняющего существующие методы моделирования – аналитические и геометрические/качественные подходы, где минимизация допущений вполне оправдана. В результате методы ML воспринимаются как продолжение традиционной линии развития естествознания, основанной на создании все более совершенных математических моделей.

Однако есть основания считать, что применение методов машинного обучения в естественных науках означает переход к принципиально новой фазе их развития, требующей пересмотра традиционных методологических предпосылок. В данной статье мы предлагаем альтернативный подход к теоретическому обоснованию методов ML в естествознании, основанный на следующих положениях.

♦ *Расширение когнитивных способностей через вычислительные технологии.* Опираясь на концепцию Пола Хамфриса о «саморасширении» (extending ourselves) посредством вычислительных технологий [13], мы исследуем, как компьютерные симуляции и методы машинного обучения способствуют расширению возможностей человеческого мышления. Это расширение позволяет преодолеть присущие человеческой интуиции и аналитическим методам ограничения при моделировании сложных природных явлений.

♦ *Механизм «компенсации предвзятости».* В рамках этой перспективы борьба с предвзятостью и минимизация субъективных допущений не являются оптимальными стратегиями для теоретического обоснования методов ML. Вместо этого мы вводим понятие механизма «компенсации предвзятости», благодаря которому влияние субъективных факторов на научное исследование «нейтрализуется» внутри используемых при исследовании алгоритмов.

ных» [16], направлены в конечном счете на снижение влияния допущений/предвзятости. Эта идея в своей базовой форме содержится еще в статье Т. Митчелла 1980 г., где вводится понятие «предвзятости» (bias) как основополагающее для обучающихся алгоритмов. В числе прочего Митчелл указывает, что хотя программе для обобщения за пределы обучающей выборки необходима «предвзятость», т.е. наличие «других источников информации», помимо «согласованности с обучающими примерами» [19, p. 185], при этом «можно ограничить число рассматриваемых обобщений, обратившись к знаниям из области решаемой задачи (task domain)» [Ibid., p. 188]. То есть исходный контекст интеграции знаний в алгоритмы ML – именно улучшение обобщаемости за счет снижения влияния предвзятости.

♦ *Различие между ML и компьютерными симуляциями.* Мы подчеркиваем фундаментальные различия в том, как механизм компенсации предвзятости действует в методах машинного обучения и в компьютерных симуляциях. Несмотря на использование вычислительных технологий в обоих подходах, они различаются по своей методологии, способам введения и корректировки допущений, а также по этапам вычислительного процесса, на которых эти допущения реализуются.

♦ В последующих разделах мы подробно рассмотрим каждый из этих пунктов, показывая, что такой подход может обеспечить теоретическую основу для применения методов ML в естественных науках.

**Развитие научного познания через
«вычислительную науку»:
конструкционные допущения
и их автоматизированная корректировка**

Как было сказано выше, развитие естествознания во многом определялось как совершенствование способов математического моделирования физических процессов. От «тихой научной революции», когда произошел скачок в развитии аналитических методов и физические законы стали выражаться посредством аппарата дифференциальных уравнений [25], мы перешли к качественным/топологическим методам, столкнувшись с явлениями нелинейной динамики, которые едва ли описываются аналитически. Но и здесь при создании фазового портрета динамических систем в пространстве состояний нас постигает «родовое проклятие» человеческого мышления, а именно невозможность работать с многомерными фазовыми пространствами, вследствие ограничений человеческой интуиции².

² Как пишет Мануэль Деланда, комментируя этот переход к топологическим методам, «мы меняем сложность изменений состояния объекта на сложность моделируемого пространства. Иными словами, мгновенное состояние объекта, неважно насколько сложное, становится единственной точкой, и это значительное упрощение, но пространство, в которое вложен объект (object's state is embedded), становится более сложным» [6, p. 6].

Таким образом, мы оказались в ситуации, где, по сути, исчерпаны возможности совершенствования способов математического моделирования для более точного описания физической реальности. И на первый взгляд кажется, что дальнейшее продвижение вперед невозможно: мы сталкиваемся либо с ограничениями аналитических методов, либо с ограничениями человеческой интуиции. И нам действительно ничего не остается, кроме как обратиться к численным методам, использование которых сопровождается необходимостью делать допущения/предположения (*assumptions*) [2, p. 182], а потому вызывает обоснованные подозрения у ученых, обеспокоенных все более разрастающейся «потерей физики» с каждым новым допущением/предположением [2, p. 182; 4, p. 234].

Встает вопрос: как преодолеть эти ограничения и продолжить развитие естествознания? Один из возможных путей – это не только совершенствовать методы математического моделирования, но и *расширять возможности самого человеческого мышления*. В истории науки человеческое мышление всегда было «константой», основой для разработки новых методов и теорий. Однако сейчас мы сталкиваемся с его объективными пределами, и дальнейшее движение вперед станет возможным только в результате их преодоления.

Британский философ науки Пол Хамфрис предложил концепцию «саморасширения» (*extending ourselves*), или «расширения вычислительных способностей» (*extensions of computational powers*). Он проводит аналогию с расширением возможностей восприятия через изобретение в XVII в. оптических приборов, которые позволили увидеть то, что было недоступно невооруженному глазу [13, p. 50]. Для понимания масштаба происходящих изменений в методологии естественно-научного исследования необходимо «отложить в сторону логический анализ структуры теории, оказавший столь большое влияние на философию науки, переключить внимание с проблем представления на проблемы вычислений» [*Ibid.*, p. 53], где ключевыми факторами являются «математика, а не логика, вычисления, а не представление, машины, а не мышление» [*Ibid.*].

Хамфрис видит в применении численных методов перспективу научной практики, которую он называет «вычислительной наукой» (computational science). Она не зависит от теоретических моделей и их ограничений, связанных с применением аналитических методов, которые снижают их применимость³. В связи с этим он вводит понятие, которое интегрирует эти теоретические модели в «вычислительные шаблоны» (computational templates) – специальные исследовательские режимы, основанные на численных методах. Эти шаблоны не являются ни теориями, ни законами, ни моделями в традиционном смысле, но обладают гибкостью и независимостью от конкретного предмета исследования [13, p. 67].

Ключевой особенностью вычислительных шаблонов является наличие «конструкционных допущений» (construction assumptions) и *набора корректировок* (correction set), которые позволяют их пересматривать при возникновении проблем в процессе моделирования. Например, в уравнении диффузии тепла допущение об идеально изолированном проводнике позволяет использовать граничное условие $u/r = 0$. Если изоляция оказывается несовершенной, это допущение пересматривается и граничное условие модифицируется для учета тепловых потерь. Аналогично если коэффициент теплопроводности k зависит от температуры, уравнение превращается в нелинейное⁴.

Идея заключается в том, что обычно перед тестированием «шаблона» мы уже предвидим расхождения между данными и его требованиями и уже знаем, какие его допущения могут оказаться не вполне адекватными для конкретного случая. «Благодаря этому ожиданию у нас часто есть план улучшения шаблона еще до начала

³ Там же: «...Как только мы выходим за рамки высокоидеализированных систем, число математических теорий, применимых только с использованием аналитических методов, становится очень небольшим» [13, p. 53]. Или: «...Чистые, абстрактные представления формальных схем скрывают тот факт, что подавляющее большинство теоретических шаблонов практически неприменимы напрямую даже к довольно простым физическим системам» [Ibid., p. 64].

⁴ Строго говоря, вычислительный шаблон у Хамфриса состоит из шести компонентов, но для наших целей достаточно упоминания первых трех: уравнения («теоретического шаблона») с начальными и граничными условиями, конструкционных допущений и набора корректировок [13, p. 102–113].

тестирования» [13, р. 77]. Это означает, что точка зрения, согласно которой «неверные предсказания логически оставляют все открытым для пересмотра, а решение мотивируется лишь прагматическими соображениями» [Ibid., р. 81], оказывается несостоятельной.

Однако корректировка таких допущений «вручную» часто является трудоемким процессом, если модель становится избыточно сложной, что может проявляться в большом количестве параметров, нелинейных зависимостях или неочевидной структуре взаимосвязей переменных (например, климатическое моделирование взаимодействия атмосферы и океана). Чем больше факторов необходимо учесть, тем сложнее становится оценка влияния каждого из них на точность модели. В таких условиях ручная корректировка может стать неподъемной задачей, требующей огромных усилий и ресурсов. И здесь на помощь приходят компьютерные симуляции, которые расширяют возможности научного исследования. Они представляют собой программную реализацию вычислительных шаблонов и позволяют автоматизировать процесс корректировки допущений.

Возьмем приведенное выше уравнение теплопроводности: традиционно допущение об идеально изолированной среде приводит к граничному условию $u/r = 0$. Если же изоляция оказывается несовершенной, исследователь вынужден пересматривать это допущение в ручном режиме и модифицировать граничное условие для учета тепловых потерь. Аналогично при зависимости коэффициента теплопроводности k от температуры уравнение становится нелинейным, усложняя решение. С помощью компьютерных симуляций эти корректировки могут быть осуществлены автоматически. Симулятивные модели позволяют вводить переменные параметры, такие как степень изоляции или температурная зависимость k , и компьютер способен быстро просчитать множество сценариев с различными комбинациями этих параметров. Это не только экономит время и ресурсы, но и открывает возможность исследовать более сложные и реалистичные модели, учитывая нелинейности и неоднородности системы. Таким образом, симуляции расширяют вычислительный шаблон, позволяя автоматизировать перебор в пространстве решений модели

и корректировать допущения без непосредственного вмешательства исследователя⁵.

Эта особенность компьютерных симуляций привела философа науки Хуана Мануэля Дюрана к мысли, что каждую из них можно рассматривать как своего рода «супермодель» – «конгломерат (amalgam) компьютерных моделей разного типа с различными масштабами, входными параметрами и протоколами» [9, р. XX]⁶.

Таким образом, расширение вычислительных возможностей через компьютерные симуляции означает переход к новой фазе развития естествознания. Если раньше «необходимо было разрабатывать математические техники, позволяющие создавать более сложные модели, то во многих случаях в вычислительной науке математика остается неизменной, и именно технология должна развиваться» [13, р. 120]. Вычислительный шаблон, созданный человеческим мышлением, внедряется в техническое устройство и получает возможность корректировки допущений (предвзятости исследователя) через автоматический перебор в пространстве решений вычислительной

⁵ Здесь важно уточнить, какие именно конструкционные допущения могут быть пересмотрены таким образом. Хамфрис приводит примеры допущений, которые можно разделить на две категории: те, которые могут корректироваться автоматически (например, коэффициент теплопроводности, плотность вещества, величина времени в уравнении диффузии), и те, корректировка которых требует участия исследователя (например, выбор уравнения для описания теплопроводности, идеализация изолированного проводника, использование аналогии тепла как жидкости) [13, р. 73–76]. Возможность автоматической корректировки допущений первого типа ограничена рамками компьютерных симуляций, тогда как вторые требуют «ручного» вмешательства и/или обращения к методам машинного обучения.

⁶ Мануэль Деланда наглядно показывает, что стоит за автоматической корректировкой допущений в симуляции: «...Можно выполнить несколько запусков одной и той же симуляции, изменяя величины, назначенные определенным параметрам, и отслеживая процесс развертывания, чтобы проверить, достигает ли он одного и того же конечного состояния. Поскольку значения параметров представляют собой допущения (assumptions), сделанные относительно среды, в которой происходит симулируемый процесс, их изменение в разных запусках эквивалентно проверке устойчивости результата к изменениям в допущениях» [7, р. 34–35]. То есть при таком «замедленном» рассмотрении симуляция обретает черты метода исследования модели, похожего на проведение экспериментов с разными условиями.

модели. В связи с этим Хамфрис подчеркивает: «то, что считается вычислимым, выходит далеко за рамки того, что на самом деле поддается вычислению с помощью ограниченных возможностей человека, и это не постоянная граница, а функция технического прогресса» [Ibid., p. 124].

В этом контексте Хамфрис указывает, что человек также может рассматриваться как конкретное вычислительное устройство. Однако симуляция, производимая таким устройством (*human simulation*), «слишком медленная, чтобы представлять большой интерес на практике» [13, p. 110]. Тем самым британский философ дает нам понять, что при должном уровне развития техники способность выполнять вычисления больше не связывается с конкретным носителем и освобождается от физических ограничений человеческого мозга. Вычислительный процесс становится функциональной характеристикой, которая может быть реализована на различных уровнях – от органического до квантового.

Эпистемическая непрозрачность: от объяснения к прогнозу

Расширение вычислительных возможностей, достигнутое благодаря компьютерным симуляциям, значительно превзошло традиционные численные методы и открыло новые горизонты для исследования сложных систем. Однако у этого прогресса есть обратная сторона – проблема *эпистемической непрозрачности*. Рассмотрение этой проблемы позволяет глубже понять механизм, посредством которого в симуляциях происходит компенсация конструктивных допущений, заложенных в модель, и объясняет, почему вычислительная мощь симуляций столь эффективно «позволяет снизить степень идеализации» [14, p. 115].

Эпистемическая непрозрачность в контексте компьютерных симуляций означает, что динамическая взаимосвязь между начальным и конечным состояниями симуляции недоступна для полного понимания и анализа. Как отмечает Пол Хамфрис, «большинство этапов процесса недоступны для прямого контроля и верификации» [14, p. 618]. Причину этого Хамфрис видит в том, что «вычислительный

процесс происходит слишком быстро, чтобы люди могли следить за ним в деталях» [13, р. 148]. Если бы этот процесс можно было замедлить настолько, чтобы человек мог рассмотреть каждый его шаг, то даже сложные вычисления стали бы эпистемически прозрачными [Ibid., р. 150].

Таким образом, отсутствие возможности пошагового разбиения вычислительного процесса нарушает принцип *модульности* (modularity), который традиционно считается основой рационального подхода в науке [17, р. 190–195]. В модульных системах каждый этап процесса может быть отдельно проанализирован, что позволяет определить роль каждого компонента в достижении конечного результата и при необходимости скорректировать его. В случае компьютерных симуляций такой подход оказывается неприменимым. Хуан Мануэль Дюран подчеркивает этот аспект, утверждая, что если бы исследователь мог остановить компьютерное моделирование и заглянуть внутрь, он не смог бы реконструировать предыдущие состояния процесса или предсказать будущие состояния на основе текущих [9, р. XIX], поскольку остановка дает возможность увидеть лишь статическое состояние («модуль»), а не динамическую последовательность шагов, необходимую для понимания процесса.

Эта проблема эпистемической непрозрачности часто отождествляется с проблемой «черного ящика» в методах машинного обучения [8, р. 1090; 21, р. 85–86; 29]. Оба термина указывают на отсутствие возможности полного понимания внутреннего механизма работы системы, *пошагового разбиения вычислительного процесса*, когда исследователю известны только входные и выходные данные, но не процесс преобразования между ними (ср., например, «модульность» по Ленхарду [17, р. 190–195] или Хамфрису [13, р. 150] и определение «черного ящика» у Лопеса-Рубио и Ратти [18, р. 1342] или у Зедника [29]).

Таким образом, искушение приравнять проблему «черного ящика» к «эпистемической непрозрачности» оказалось настолько велико, что поддаться ему стало своеобразным трендом. Так, собственно, в прикладных исследованиях и эпистемологии ML сформировалась стратегия теоретического обоснования методов машинного обучения, которая заключается в борьбе с предвзятостью (bias)

и минимизации допущений в алгоритмах ML [10; 11; 15, p. 424; 26]. И в самом деле, чем меньше допущений используется в алгоритмах ML, тем ниже «непрозрачность» в работе модели, тем более точно мы представляем промежуточные шаги, взаимосвязи между ее «входом» и «выходом».

Однако дело в том, что полное устранение допущений/предвзятости технически невозможно. В любом случае в алгоритме должна присутствовать «индуктивная предвзятость» (inductive bias), которая является условием эффективного применения методов ML за пределами обучающего набора данных [19, p. 185]. Поэтому приходится мириться с существованием «черного ящика» в тех областях естественно-научных исследований, где применяется ML: мол, есть такая «прогнозно ориентированная» наука, которая представляет собой альтернативу науке механистической, направленной на причинно-следственное объяснение [8, p. 1091; 18, p. 3148; 21, p. 98; 22, p. 913–914; 24, p. 159].

С нашей точки зрения, подобный подход следует рассматривать как отказ от решения проблемы «черного ящика» и считать его в целом ошибочным. Мы настаиваем, что применение методов ML в естествознании имеет принципиальные отличия от использования компьютерных симуляций. И первым указанием на это служит то обстоятельство, что проблема эпистемической непрозрачности в компьютерных симуляциях, в отличие от проблемы «черного ящика» в машинном обучении, *имеет решение*.

Динамика компьютерных симуляций и эмерджентные свойства: преодоление эпистемической непрозрачности

Это решение проблемы эпистемической непрозрачности становится возможным благодаря *динамическому характеру* реализации вычислительной модели в компьютерных симуляциях.

На эту особенность обратил внимание немецкий философ Штефан Хартманн, предложив уточнение к подходу Хамфриса. Он дополнил определение Хамфриса, подчеркнув, что о компьютерных симуляциях нельзя говорить только лишь как о «реализованном на

компьютере способе исследования свойств математических моделей в тех случаях, когда аналитические методы недоступны» [14, р. 501]. Не менее важен тот факт, что они «предназначены для воспроизведения *эволюции реальной системы во времени* (курсив наш. – Авт.). Другими словами, симуляция представляет собой процесс, имитирующий другой процесс» [12, р. 82]. Это отличие имеет фундаментальное значение для понимания роли симуляций в научном исследовании.

Пол Хамфрис в своей монографии о вычислительной науке [13] охотно принял эту поправку. Более того, в свою очередь, он указал, что выполнение вычислительного процесса в режиме реального времени «кардинально отличает его от лежащей в его основе вычислительной модели» [13, р. 109], которая, по существу, носит вневременной характер. Хамфрис отмечает: «Переход от абстрактных логических и математических представлений к протекающим во времени вычислительным процессам – это то, что делает эти процессы поддающимися усилению (*amplification*)» [14, р. 619]. Таким образом, динамическая реализация модели позволяет исследователям наблюдать за процессом в его развитии, что открывает новые возможности для понимания сложных систем.

В чем здесь идея? По сути, именно вычисления в реальном времени позволяют исследовать пространство решений модели методом проб и ошибок (*trial-and-error procedures*) или случайного блуждания в конфигурационном пространстве (*random walk in configuration space*), отображая многочисленные связи между вычислительным шаблоном и его возможными решениями [13, р. 150]. Например, мы «можем наблюдать, какие структуры остаются стабильными со временем» [Ibid., р. 113], или, в случае мультиагентных систем, «обнаружить *эмерджентные паттерны более высокого порядка* (курсив наш. – Авт.)» [Ibid.], которые невозможно выявить, «если результаты таких моделей отображать численно или статично» [Ibid.].

Понятие эмерджентных свойств (*emergent properties*) становится ключом к решению проблемы эпистемической непрозрачности. Именно с его помощью Мануэль Деланда теоретически обосновывает применение компьютерных симуляций в естествознании. Благодаря

своему динамическому характеру компьютерные симуляции формируют для вычислительных моделей пространства состояний или пространства возможностей, связанные с бесчисленными взаимодействиями физических систем с окружением. Это возможно, «поскольку взаимодействия... могут быть смоделированы и варьироваться множеством способов до тех пор, пока уникальные особенности (singular features) пространства возможностей не станут видимыми» [7, р. 21].

Эти уникальные эмерджентные структуры нельзя было вывести напрямую из «вневременной» вычислительной модели. Возможности этой модели для корректировки конструкционных допущений, заложенных в ее основе, ограничены. В то время как в симуляциях для исследования пространства возможностей в реальном времени фактически развертывается процесс программной реализации различных допущений.

В результате при разных конструкционных допущениях могут быть обнаружены *структуры более высокого порядка*, которые делают задачу отслеживания каждого этапа вычислительного процесса менее актуальной. Возможность оценивать влияние допущений и корректировать их относится к структурам более «низкого» уровня – локальным взаимодействиям. Как отмечает Й. Ленхард, комментируя открытие аттрактора Лоренца, «интуитивно понятные визуализации могут позволить получить представление о сложной динамике, даже если взаимное влияние (*mutual interplay*) допущений, заложенных в модели, не полностью понятно (курсив наш. – Авт.) [17, р. 52].

Таким образом, выбор масштаба рассмотрения явления – ключевой аспект, делающий симуляцию в определенном смысле независимой от произведенных допущений. Дело в том, что «стабильное эмерджентное свойство “безразлично” к локальным изменениям во взаимодействиях, которые его порождают. Это объективное безразличие ведет к объективной *объяснительной нерелевантности* деталей взаимодействий: их включение в объяснение было бы избыточным, поскольку множество различных взаимодействий дадут один и тот же результат» [7, р. 14].

Это означает, что субъективные допущения могут влиять на локальные процессы в компьютерной симуляции, но в значительной мере исчерпывают свое влияние на уровне глобального функциони-

рования модели⁷. Именно благодаря тому, что в реальном времени вычислений проявляются эмерджентные свойства системы, происходит своего рода компенсация влияния допущений/предвзятости: они как бы «растворяются» в эволюции симулятивной модели, осуществляющей автоматизированный поиск в пространстве решений.

Динамический характер компьютерных симуляций и возможность наблюдать эмерджентные свойства позволяют преодолеть ограничения, связанные с эпистемической непрозрачностью [5]. Симуляции становятся мощным инструментом для исследования сложных систем, где детализация локальных взаимодействий менее важна, чем наблюдение глобальных эмерджентных свойств. Таким образом, исчезает потребность знать все этапы вычислительного процесса, *и именно это позволяет выделить прогностические задачи как основное направление научных исследований* с применением компьютерных симуляций. Ленхард формулирует это так: «Имитационное моделирование существенно меняет математическое моделирование: понимание, основанное на теории, и эпистемическая прозрачность отходят на второй план, тогда как на первый план выходит понимание, основанное на симуляции, ориентированное на вмешательство и прогнозирование, а не на теоретическое объяснение» [17, р.115].

Между симуляциями и машинным обучением: информация, методология и связь с реальностью

Можем ли мы применить те же рассуждения к проблеме «черного ящика» в алгоритмах машинного обучения? То есть, по сути, просто

⁷ Согласно Ленхарду, «допущения модели, эффекты которых трудно или даже невозможно оценить, могут быть проверены, изменены и модифицированы с помощью итеративно-исследовательских процедур» [17, р. 133], и «симулятивное моделирование может обеспечить ориентацию на поведение моделей, основанную на опыте итерационных вычислений, который проявляется в изменении параметров» [Ibid., р. 119]. Это делает регулировку параметров (adjustment of parameters) «ключевым компонентом симулятивного моделирования» [Ibid., р. 200], поскольку она «объединяет разнородные части, которые не имеют общей теоретической основы. Настройка выполняется *целостно на основе поведения глобальной модели* (курсив наш. – Авт.). То, как функционируют отдельные части, часто остается непрозрачным» [Ibid.].

ее игнорировать, как эпистемическую непрозрачность – в компьютерных симуляциях? К сожалению, нет. Попробуем объяснить последовательно нашу позицию.

Безусловно, между компьютерными симуляциями и методами машинного обучения существует определенное «родовое» сходство. Оба подхода используют технические устройства для выполнения вычислительных операций, а иногда – и одни и те же инструменты, такие как нейросети или генетические алгоритмы⁸. Однако в контексте естествознания эти методы представляют собой во многом противоположные исследовательские подходы. Именно это различие делает невозможным сведение методов машинного обучения лишь к их прогностической функции в ущерб объяснительной, как это происходит в случае компьютерных симуляций.

На это фундаментальное различие указывали разные авторы, подходя к его анализу с разных сторон. Однако, если рассматривать их подходы в совокупности, то можно увидеть единую линию рассуждений, подобно тому как в известной притче слепые ощупывали слона с разных сторон и каждый предлагал свою версию того, с чем имеет дело.

Л. фон Рюден и соавторы [23] анализируют указанное различие с точки зрения способов обработки информации и отмечают, что симуляции и методы ML при проведении исследований организуют информационный поток и его сжатие противоположными способами. Симуляции подразумевают «развертывание локальных взаимодействий из компактно представленной исходной модели в расширенное пространство данных» – это так называемый «подход сверху вниз»

⁸ Видимо, весь вопрос в том, как они используются. С одной стороны, они могут быть полезны для автоматизированного поиска решений в пространстве возможностей физической системы, когда у нас есть базовые знания о физических явлениях. Например, генетические алгоритмы подходят для моделирования механизмов саморепликации молекул РНК [7, р. 55–56], а рекуррентные нейронные сети и «самоорганизующиеся карты» (self-organizing maps) – для изучения нейронных механизмов, ответственных за память у млекопитающих [Ibid., р. 103]. С другой стороны, их можно применять для выявления скрытых зависимостей в данных, когда основные закономерности и модели нам неизвестны. Например, программа пекинских разработчиков DLGA-PDE представляет собой комбинацию нейросети и генетического алгоритма, которая используется для открытия дифференциальных уравнений из набора данных [28].

(top-down approach). В отличие от этого, инструменты машинного обучения осуществляют «поиск закономерностей в изначально большом пространстве данных, которые в конечном итоге представлены в сжатой форме в виде окончательной гипотезы», – это «подход снизу вверх» (bottom-up approach) [Ibid., p. 551].

Х. Дюран акцентирует внимание на методологическом различии между симуляциями и методами ML (Big Data): перед нами два «диаметрально противоположных источника, из которых исследователи собирают данные» [9, p. 169]. Дюран пишет: «В то время как компьютерные симуляции генерируют большое количество данных из имеющейся модели, большие данные (Big Data) воссоздают структуру определенного явления на основе обширных наборов данных» [Ibid.]. Таким образом, основное различие заключается в «характере наших выводов о целевой системе (target system)»: «Компьютерные симуляции, как правило, являются результатом имитационной модели (дедуктивный подход. – *Авт.*), тогда как большие данные основаны преимущественно на индуктивных выводах» [Ibid.].

Ф. Боге, по всей видимости, уже учитывая это методологическое различие [3], концентрирует внимание на том, как выстраиваются связи между вычислительным алгоритмом и реальностью (target, «целевой системой»). Он пишет, что компьютерная симуляция «начинается с концептуализации целевой системы (target) и на этой основе предсказывает “гипотетические данные”» [Ibid., p. 59]. Что касается машинного обучения, то оно «начинается с концептуализации данных и на их основе также предсказывает новые гипотетические данные» [Ibid.]. В результате связь с «целевой системой» в машинном обучении должна быть установлена *post factum*, тогда как в компьютерных симуляциях она делается (или по крайней мере пытается быть сделанной) заранее.

Сопоставим эти позиции:

- компьютерные симуляции и методы машинного обучения различаются по способу преобразования информации между моделями и данными. В компьютерных симуляциях информация *расширяется* от модели к данным, а в методах машинного обучения *сжимается* от данных к модели;
- эти различия в информационных процессах отражаются в различных методологиях познания. Компьютерные симуляции ис-

пользуют дедуктивный подход (*от модели к данным*), а методы машинного обучения – индуктивный (*от данных к модели*);

- в результате различий в методологии компьютерные симуляции и методы машинного обучения по-разному устанавливают связь с реальностью. В компьютерных симуляциях связь с реальным миром устанавливается *заранее* через концептуальную модель, тогда как в методах машинного обучения – *после* обработки данных и формирования модели.

Таким образом, сначала мы понимаем, как информация преобразуется и обрабатывается в компьютерных симуляциях и методах машинного обучения [23]. Затем мы осознаем, что эти различия в информационных процессах связаны с разными методологическими подходами к научному познанию – дедукцией и индукцией [9]. Наконец, мы видим, как эти методологические различия влияют на процесс моделирования и установления связи с реальным миром [3].

Что это означает в плане применения методов машинного обучения в естествознании для прогнозирования? Можем ли мы игнорировать проблему «черного ящика», как это было сделано с эпистемической непрозрачностью в случае компьютерных симуляций?

«Черный ящик»: от прогноза к объяснению

Итак, в методах машинного обучения мы собираем большие объемы данных и используем специальные алгоритмы для выявления паттернов (закономерностей) без предварительной теории. На основе этих выявленных паттернов делаются прогнозы, и уже *post factum* мы пытаемся установить их связь с реальностью – понять, каким образом эти прогнозы опираются на реальные закономерности, скрытые в исходных данных.

В противоположность этому, в компьютерных симуляциях используется предварительная теория, из которой мы получаем дополнительные данные. В этих данных могут быть обнаружены недоступные ранее паттерны (эмерджентные свойства), но они основаны на уже существующей концептуальной модели. Таким образом, в машинном обучении мы из данных получаем другие данные (прогнозы) и затем пытаемся найти теоретические основания им доверять, тогда как в симуляциях теория предшествует данным и направляет процесс моделирования.

Следовательно, акцент в машинном обучении делается на обнаружении паттернов и предсказании, часто без полного понимания внутренних механизмов, лежащих в основе открытия этих паттернов. Однако такое непонимание уже не может быть проигнорировано, как в случае с симуляциями; напротив, его необходимо устранить, чтобы обеспечить научную ценность прогнозов. Без объяснения того, как были выявлены эти закономерности, прогнозы, полученные методами машинного обучения, остаются недостаточно обоснованными с научной точки зрения.

Здесь за точными формулировками следует снова обратиться к Ф. Боге и его весьма проницательной статье о двух «типах непрозрачности» [3]. Один из них, который автор называет «*h*-непрозрачностью» (*how-opacity*), является общим и для компьютерных симуляций, и для методов машинного обучения (глубоких нейросетей). Этот тип касается того, *как* алгоритм обучается – каким образом модель изменяет свои параметры автоматически в ответ на входные данные [Ibid., p. 59–60]. Между тем так называемая «*w*-непрозрачность» (*what-opacity*) является «визитной карточкой» машинного обучения. Она отражает, *что именно* алгоритм изучает, т.е. какие скрытые паттерны он извлекает из данных и на их основе делает успешные предсказания [3, p. 61–63]⁹.

⁹ Боге приводит два примера из физики элементарных частиц и биологии, чтобы показать, как именно проявляется *w*-непрозрачность. Нейросети на основе «низкоуровневых» данных формируют сложные признаки (*complex features*), такие как масса частицы или вторичная структура белка [3, p. 62–63]. Однако сами зависимости, паттерны/закономерности, на основе которых были сформированы эти признаки, остаются скрытыми: мы не знаем, что именно модель выявила в данных. Это ограничивает нашу способность использовать такие модели для научных открытий: понимание, *какие аспекты импульсов частиц* связаны с их массой или *какие взаимодействия аминокислот* определяют структуру белка, могло бы привести к новым знаниям [Ibid.]. Таким образом, *w*-непрозрачность отражает эту проблему: модель успешно формирует высокоуровневые признаки, но скрытые зависимости остаются недоступными. Можно привести более простой пример – с обучением AlphaZero [1], где программа анализирует огромное количество партий, а шахматные диаграммы представляют визуально те паттерны, которые она «запоминает» как ключевые для принятия стратегических решений. При этом по большей части мы совершенно не представляем, что именно эти диаграммы, которые программа идентифицирует как устойчивые паттерны, означают с точки зрения оценки шахматной позиции.

Таким образом, именно эта «и-непрозрачность» (скрытые паттерны в наборе данных, недоступные для понимания человеческим мышлением) не позволяет игнорировать задачу объяснения, т.е. установления причинно-следственных связей, в пользу простого прогнозирования. Нам необходимо каким-то образом установить те закономерности, которые мы вывели с помощью методов машинного обучения, обнаружить их реальную ценность и связь с целевой системой. Иначе прогнозы не будут иметь научной ценности.

Иными словами, это ключевой фактор, который, с нашей точки зрения, делает целесообразным разделение проблем эпистемической непрозрачности и «черного ящика» на терминологическом уровне. Согласившись, что в машинном обучении существует проблема эпистемической непрозрачности, роднящая его с компьютерными симуляциями, мы должны признать, что существует и более серьезная проблема, для которой пока нет общепринятого решения.

♦ Эпистемическая непрозрачность устраняется, если становятся понятны алгоритмические операции, с помощью которых в этом пространстве производятся вычисления. В этом смысле Хамфрис прав, полагая, что она сводится к скорости вычислений: если их достаточно замедлить, проблема эпистемической непрозрачности исчезает.

♦ «Черный ящик» не исчезает, даже если разобраться в каждой алгоритмической операции, использованной при обучении. Мы не знаем, чему именно обучается машина при взаимодействии с наборами данных; у нас есть только модель, но мы не понимаем, каким образом она была получена и что именно побудило машину выбрать именно ее.

Конструкционные допущения и их влияние на прогнозы

Это также означает, что не существует механизма компенсации предвзятости в машинном обучении, аналогичного тому, который имеет место в компьютерных симуляциях:

- в компьютерных симуляциях допущения/предвзятость входят *в структуру вычислительной модели, реализуемой в техническом устройстве*. Эпистемическая непрозрачность не позволяет полностью контролировать их воздействие, но глобаль-

ный уровень функционирования модели открывает возможность рассмотрения ее эмерджентных свойств независимо от влияния допущений на локальном уровне;

- в машинном обучении допущения/предвзятость (inductive bias) используются для *построения специальных алгоритмов* [15, р. 424], которые самостоятельно, независимо от человеческого мышления со всеми его ограничениями осуществляют моделирование физического процесса.

В этом смысле выбор перспективы исследования зависит от того, на каком этапе вычислительного процесса мы вводим конструкционные допущения.

В компьютерных симуляциях эти допущения входят в структуру модели до начала вычислений и явно отражены в ее концептуальной основе. Это позволяет исследователю иметь определенный контроль над влиянием допущений на результаты модели и при необходимости корректировать их в процессе моделирования. Эпистемическая непрозрачность, возникающая из-за сложности вычислений, может быть компенсирована за счет анализа глобальной динамики модели и ее эмерджентных свойств.

В методах машинного обучения конструкционные допущения вводятся на этапе разработки алгоритма и глубоко интегрированы в процесс обучения модели. Поскольку модель самостоятельно извлекает паттерны из данных, эти допущения могут оставаться неявными и трудно поддающимися анализу после завершения обучения. Это затрудняет понимание того, как именно допущения влияют на выявленные закономерности и полученные прогнозы.

Таким образом, выбор того, на каком этапе вводятся конструкционные допущения, имеет решающее значение для возможности компенсации предвзятости в модели. В машинном обучении без явного понимания и контроля над этими допущениями проблема «черного ящика» становится еще более острой. Мы сталкиваемся не только с непрозрачностью алгоритмических операций, но и с невозможностью понять, какие именно паттерны/закономерности были выявлены и, соответственно, почему модель делает те или иные предсказания.

Как именно работает механизм компенсации предвзятости в методах машинного обучения – вопрос, неотделимый от решения проб-

лемы «черного ящика». Без осознания конструкционных допущений и контроля над ними на этапе разработки и обучения модели мы не можем гарантировать научную обоснованность полученных прогнозов. Пока что можно констатировать, что для дальнейшего продвижения в области эпистемологии ML необходимо избавиться от «наивного подхода»: разделение научной практики на объяснение и прогноз, которое может быть оправданным в случае компьютерных симуляций, оказывается неприменимым к методам машинного обучения в естествознании.

Заключение

В настоящей статье мы рассмотрели фундаментальные различия между методами машинного обучения и компьютерными симуляциями как инструментами естествознания. С самого начала мы исходили из того, что доминирующая стратегия интеграции методов ML, основанная на борьбе с предвзятостью и минимизации субъективных допущений, не приводит к удовлетворительному решению проблемы «черного ящика». В этой стратегии методы ML рассматриваются как очередной математический инструмент, дополняющий существующие подходы к моделированию физических процессов, и в итоге ML воспринимается как продолжение традиционной эволюции естественных наук через совершенствование математических моделей. Поэтому, следуя интуиции Пола Хамфриса, мы использовали альтернативный подход, основанный на расширении возможностей мышления через совершенствование вычислительных технологий.

♦ Компьютерные симуляции благодаря своему динамическому характеру и способности моделировать эволюцию систем во времени предоставляют исследователям возможность наблюдать эмерджентные свойства сложных систем. Это расширение когнитивных способностей позволяет преодолевать ограничения, присущие человеческой интуиции и аналитическим методам при моделировании сложных природных явлений.

♦ В компьютерных симуляциях возможна компенсация предвзятости, когда влияние субъективных факторов нивелируется на глобальном уровне модели. Эпистемическая непрозрачность на локальном уровне компенсируется за счет понимания глобальной динамики

системы и наблюдения эмерджентных свойств. Таким образом, симуляции предоставляют механизм, который позволяет учитывать допущения/предвзятость и эффективно ими управлять, что приводит к созданию устойчивых и надежных моделей.

♦ В методах машинного обучения отсутствует аналогичный механизм «компенсации предвзятости». Поскольку ML-модели самостоятельно извлекают скрытые закономерности из больших объемов данных без предварительной теоретической модели, конструктивные допущения остаются неявными и трудно поддающимися анализу после обучения. Это усложняет понимание того, *как именно* допущения влияют на выявленные паттерны и полученные прогнозы.

♦ Традиционное разделение научной практики на «объяснение» и «прогнозирование», применимое в случае компьютерных симуляций, не подходит для методов ML в естествознании. В машинном обучении объяснительная составляющая не может быть отодвинута на второй план, так как без нее прогнозы остаются без достаточного обоснования и не могут считаться надежными в научном смысле.

Таким образом, применение методов ML в естественных науках не является просто очередным этапом развития математических моделей, а представляет собой качественно новый этап, требующий переосмысления фундаментальных методологических принципов. Решение проблемы «черного ящика» и разработка механизма компенсации предвзятости в ML являются обязательными условиями для дальнейшего прогресса в этой области. Необходимы новые подходы, которые позволят установить связь моделей ML с физической реальностью, интегрируя их в естествознание без потери объяснительной силы.

Литература

1. *Acquisition of Chess Knowledge in AlphaZero*. URL: <https://storage.googleapis.com/uncertainty-over-space/alphachess/index.html> (дата обращения: 20.04.2025).
2. *Blakseth S.S., Rasheed A., Kvamsdal T., San O.* Deep neural network enabled corrective source term approach to hybrid analysis and modeling // *Neural Networks*. 2021. Vol. 146, No. 15. P. 181–199.
3. *Boge F.J.* Two dimensions of opacity and the deep learning predicament // *Minds and Machines*. 2022. Vol. 32. P. 43–75.
4. *Brunton S.L., Kutz J.N.* Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge: Cambridge University Press, 2019. 492 p.

5. *Chen Y., Zhang D.* Integration of knowledge and data in machine learning // *Neural Networks*. URL: <https://www.alphaxiv.org/abs/2202.10337> (дата обращения: 20.04.2025).
6. *DeLanda M.* *Intensive Science and Virtual Philosophy*. Reprint ed. London: Bloomsbury Academic, 2013. 232 p.
7. *DeLanda M.* *Philosophy and Simulation: The Emergence of Synthetic Reason*. London: A&C Black, 2011. 226 p.
8. *Duede E.* Deep learning opacity in scientific discovery // *Philosophy of Science*. 2022. Vol. 90. P. 1089–1099.
9. *Duran J.M.* *Computer Simulations in Science and Engineering: Concepts – Practices – Perspectives*. Dordrecht: Springer, 2018. 209 p.
10. *Enni S.A., Herrie M.B.* Turning biases into hypotheses through method: A logic of scientific discovery for machine learning // *Big Data & Society*. 2021. Vol. 8, No. 1. P. 1–13.
11. *Hagendorff T., Fabi S.* Why we need biased AI: How including cognitive biases can enhance AI systems // *Journal of Experimental and Theoretical Artificial Intelligence*. 2024. Vol. 36, No. 8. P. 1885–1898.
12. *Hartmann S.* The world as a process: Simulation in the natural and social sciences // *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View* / Ed. by R. Hegselmann, U. Muller, K. Troitzsch. Dordrecht: Kluwer Academic, 1996. P. 77–100.
13. *Humphreys P.* *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press, 2004. 182 p.
14. *Humphreys P.* The philosophical novelty of computer simulation methods // *Synthese*. 2009. Vol. 169, No. 3. P. 615–626.
15. *Karniadakis G.E., Kevrekidis I.G., Lu L., Perdikaris P., Wang S., Yang L.* Physics-informed machine learning // *Nature Reviews Physics*. May 2021. P. 422–440.
16. *Karpatne A., Atluri G., Faghmous J.H., Steinbach M., Banerjee A., Ganguly A., Shekhar S., Samatova N., Kumar V.* Theory-guided data science: A new paradigm for scientific discovery from data // *IEEE Transactions on Knowledge and Data Engineering*. URL: <https://arxiv.org/pdf/1612.08544> (дата обращения: 20.04.2025).
17. *Lenhard J.* *Calculated Surprises: A Philosophy of Computer Simulation*. Oxford: Oxford University Press, 2019. 256 p.
18. *López-Rubio E., Ratti E.* Data science and molecular biology: Prediction and mechanistic explanation // *Synthese*. 2021. Vol. 198, No. 4. P. 3131–3156.
19. *Mitchell T.* *The Need for Biases in Learning Generalizations*. New Brunswick, NJ: Rutgers University, Department of Computer Science, 1980.
20. *Napoletani D., Panza M., Struppa D.* The agnostic structure of data science methods // *Lato Sensu: Revue de la Société de Philosophie des Sciences*. 2021. Vol. 8, No. 2. P. 44–57.
21. *Nickles T.* Whatever happened to the logic of discovery? From transparent logic to alien reasoning // *Current Trends in Philosophy of Science: A Prospective for the Near Future* / Ed. by W.J. González. Cham: Springer, 2022. P. 81–102.
22. *Pietsch W.* Aspects of theory-ladenness in data-intensive science // *Philosophy of Science*. 2015. Vol. 82, No. 5. P. 905–916.

23. *Rueden L., von, Mayer S., Sifa R., Bauckhage C., Garcke J.* Combining machine learning and simulation to a hybrid modelling approach: Current and future directions // *Lecture Notes in Computer Science: Advances in Intelligent Data Analysis XVIII*. Cham: Springer International Publishing, 2020. P. 548–560.

24. *Srećković S., Berber A., Filipović N.* The automated laplacean demon: How ML challenges our views on prediction and explanation // *Minds and Machines*. 2021. Vol. 32, No. 1. P. 159–183.

25. *Stan M.* From metaphysical principles to dynamical laws // *The Cambridge History of Philosophy of the Scientific Revolution*. URL: https://www.academia.edu/41468384/From_metaphysical_principles_to_dynamical_laws (дата обращения: 20.04.2025).

26. *Sterkenburg T., Grünwald P.* The no-free-lunch theorems of supervised learning // *Synthese*. 2021. Vol. 199, No. 3-4. P. 9979–10015.

27. *Willard J., Jia X., Xu S., Steinbach M.* Integrating scientific knowledge with machine learning for engineering and environmental systems // *ACM Computing Surveys*. 2022. Vol. 55, No. 4. P. 1–37.

28. *Xu H., Zhang D., Zeng J.* Deep-learning of parametric partial differential equations from sparse and noisy data // *Physics of Fluids*. URL: <https://arxiv.org/pdf/2005.07916> (дата обращения: 20.04.2025).

29. *Zednik C.* Solving the black box problem: A normative framework for explainable artificial intelligence // *Explainable AI Conference Proceedings*. URL: <https://arxiv.org/pdf/1903.04361> (дата обращения: 20.04.2025).

References

1. *Acquisition of Chess Knowledge in AlphaZero*. Available at: <https://storage.googleapis.com/uncertainty-over-space/alphachess/index.html> (date of access: 20.04.2025).

2. *Blaketh, S.S., A. Rasheed, T. Kvamsdal & O. San.* (2021). Deep neural network enabled corrective source term approach to hybrid analysis and modeling. *Neural Networks*, Vol. 146, No. 15, 181–199.

3. *Boge, F.J.* (2022). Two dimensions of opacity and the deep learning predicament. *Minds and Machines*, 32, 43–75.

4. *Brunton, S.L. & J.N. Kutz.* (2019). *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge, Cambridge University Press, 492.

5. *Chen, Y. & D. Zhang.* Integration of knowledge and data in machine learning. *Neural Networks*. Available at: <https://www.alphaxiv.org/abs/2202.10337> (date of access: 20.04.2025).

6. *DeLanda, M.* (2013). *Intensive Science and Virtual Philosophy*. Reprint ed. London, Bloomsbury Academic, 232.

7. *DeLanda, M.* (2011). *Philosophy and Simulation: The Emergence of Synthetic Reason*. London, A&C Black, 226.

8. *Duede, E.* (2022). Deep learning opacity in scientific discovery. *Philosophy of Science*, 90, 1089–1099.
9. *Duran, J.M.* (2018). *Computer Simulations in Science and Engineering: Concepts – Practices – Perspectives*. Dordrecht, Springer, 209.
10. *Enni, S.A. & M.B. Herrie.* (2021). Turning biases into hypotheses through method: A logic of scientific discovery for machine learning. *Big Data & Society*, Vol. 8, No. 1, 1–13.
11. *Hagendorff, T. & S. Fabi.* (2024). Why we need biased AI: How including cognitive biases can enhance AI systems. *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 36, No. 8, 1885–1898.
12. *Hartmann, S.* (1996). The world as a process: Simulation in the natural and social sciences. In: R. Hegselmann, U. Muller & K. Troitzsch (Eds.). *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Dordrecht, Kluwer Academic, 1996, 77–100.
13. *Humphreys, P.* (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford, Oxford University Press, 182.
14. *Humphreys, P.* (2009). The philosophical novelty of computer simulation methods. *Synthese*, Vol. 169, No. 3, 615–626.
15. *Karniadakis, G.E., I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang & L. Yang.* (2021). Physics-informed machine learning. *Nature Reviews Physics*, May, 422–440.
16. *Karpatne, A., G. Atluri, J.H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova & V. Kumar.* Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*. Available at: <https://arxiv.org/pdf/1612.08544> (date of access: 20.04.2025).
17. *Lenhard, J.* (2019). *Calculated Surprises: A Philosophy of Computer Simulation*. Oxford, Oxford University Press, 256.
18. *López-Rubio, E. & E. Ratti.* (2021). Data science and molecular biology: Prediction and mechanistic explanation. *Synthese*, Vol. 198, No. 4, 3131–3156.
19. *Mitchell, T.* (1980). *The Need for Biases in Learning Generalizations*. New Brunswick, NJ, Rutgers University, Department of Computer Science.
20. *Napoletani, D., M. Panza & D. Struppa.* (2021). The agnostic structure of data science methods. *Lato Sensus: Revue de la Société de Philosophie des Sciences*, Vol. 8, No. 2, 44–57.
21. *Nickles, T.* (2022). Whatever happened to the logic of discovery? From transparent logic to alien reasoning. In: W.J. González (Ed.). *Current Trends in Philosophy of Science: A Prospective for the Near Future*. Cham, Springer, 81–102.
22. *Pietsch, W.* (2015). Aspects of theory-ladenness in data-intensive science. *Philosophy of Science*, Vol. 82, No. 5, 905–916.
23. *Rueden, L. von, S. Mayer, R. Sifa, C. Bauckhage & J. Garcke.* (2020). Combining machine learning and simulation to a hybrid modelling approach: current and future directions. In: *Lecture Notes in Computer Science: Advances in Intelligent Data Analysis XVIII*. Cham, Springer International Publishing, 548–560.

24. *Srećković, S., A. Berber & N. Filipović.* (2021). The automated Laplacean demon: How ML challenges our views on prediction and explanation. *Minds and Machines*, Vol. 32, No. 1, 159–183.

25. *Stan, M.* From Metaphysical principles to dynamical laws. In: *The Cambridge History of Philosophy of the Scientific Revolution*. Available at: https://www.academia.edu/41468384/From_metaphysical_principles_to_dynamical_laws (date of access: 20.04.2025).

26. *Sterkenburg, T. & P. Grünwald.* (2021). The no-free-lunch theorems of supervised learning. *Synthese*, Vol. 199, No. 3-4, 9979–10015.

27. *Willard, J., X. Jia, S. Xu & M. Steinbach.* (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, Vol. 55, No. 4, 1–37.

28. *Xu, H., D. Zhang & J. Zeng.* Deep-learning of parametric partial differential equations from sparse and noisy data. *Physics of Fluids*. Available at: <https://arxiv.org/pdf/2005.07916> (date of access: 20.04.2025).

29. *Zednik, C.* Solving the black box problem: A normative framework for explainable artificial intelligence. In: *Explainable AI Conference Proceedings*. Available at: <https://arxiv.org/pdf/1903.04361> (date of access: 20.04.2025).

Информация об авторах

Сухно Алексей Андреевич – независимый исследователь.
volyakvlasti@mail.ru

Гулин Вячеслав Владимирович – Северо-Восточный федеральный университет им. М.К. Аммосова; Московский государственный университет им. М.В. Ломоносова (119192, Москва, Мичуринский проспект, 1).
kornet104@gmail.com

Information about the authors

Sukhno, Alexey Andreevich – independent researcher.
volyakvlasti@mail.ru

Gulin, Vyacheslav Viktorovich – Ammosov North-Eastern Federal University; Lomonosov Moscow State University (1, Michurinskiy Ave., Moscow, 119192, Russia).
kornet104@gmail.com

Дата поступления 11.02.2026.
Принята к публикации 12.05.2026.