УДК 165.2
DOI 10.15372/PS20250108
EDN UBWWEJ

А.А. Сухно, В.В. Гулин

РОЛЬ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЕСТЕСТВОЗНАНИИ: ПАРАДОКС «СКРЫТОЙ ДИНАМИКИ» И БОРЬБА С ПРЕДВЗЯТОСТЬЮ

Цель статьи заключается в формировании подхода к теоретическому обоснованию применения методов машинного обучения (machine learning, ML) в естественных науках. Основное препятствие на этом пути – проблема «черного ящика», или «эпистемической непрозрачности», которая состоит в отсутствии доступа ко всем элементам процесса познания, осуществляемого с помощью ML. При разработке подхода авторы формулируют критерий, которому должно соответствовать решение этой проблемы.

Авторы указывают, что причиной обращения к машинному обучению в естественных науках является ограниченная применимость традиционных аналитических и качественных методов в исследовании природы, поскольку человеческое мышление достигло своих пределов в процессе их использования ввиду сложности и многомерности изучаемых систем. Следовательно, решение проблемы «черного ящика» должно объяснять, как с помощью ML эти пределы можно преодолеть – как человеческое мышление может получить доступ к той области знания, которая ему недоступна вследствие его же собственных внутренних ограничений. В этой связи утверждается, что подход, стихийно сформировавшийся в рамках компьютерных наук, не может служить основой для решения указанной задачи. Такой подход предполагает внедрение существующих научных знаний в инструменты ML с целью борьбы с предвзятостью (bias), характерной для машинного обучения, – субъективными допущениями (assupmptions) исследователя, необходимыми для успешного обобщения за пределами обучающей выборки.

Авторы показывают, что внедрение научных знаний в инструменты ML имеет лишь прикладное значение и не решает проблему теоретического обоснования, поскольку не соответствует предлагаемому ими критерию: оно не преодолевает пределы человеческого мышления, а только согласует результаты ML с уже известными научными знаниями.

Ключевые слова: машинное обучение; черный ящик; эпистемическая непрозрачность; естествознание; предвзятость; пределы мышления; скрытая динамика; субъективные допущения

A.A. Sukhno, V.V. Gulin

THE ROLE OF MACHINE LEARNING METHODS IN NATURAL SCIENCE: THE PARADOX OF "HIDDEN DYNAMICS" AND THE FIGHT AGAINST BIAS

The article aims to develop an approach to the theoretical justification of the use of machine learning (ML) methods in natural sciences. The main obstacle on this path is the problem of the "black box", or "epistemic opacity", which is the lack of access to all elements of the cognitive process carried out through ML. In developing the approach, the authors formulate a criterion that must be met to solve the problem.

The authors point out that the reason for turning to machine learning in natural sciences is the limited applicability of traditional analytical and qualitative methods for studying nature, since human thinking has reached its limits in their use – because of the complexity and multidimensionality of the studied systems. Therefore, the solution to the "black box" problem must explain how ML can overcome these limits, i.e. how human thinking can access a domain of knowledge that is inaccessible to it due to its own internal limitations. In this regard, it is argued that the approach that spontaneously formed within computer science cannot serve as a basis for solving this task. Such an approach involves incorporating existing scientific knowledge into ML tools in order to fight against bias, which is characteristic of machine learning, i.e. the researcher's subjective assumptions that are necessary for successful generalization beyond the training set.

The authors show that incorporating scientific knowledge into ML tools has only applied value and does not solve the problem of theoretical justification, since it does not meet the criterion they propose – it does not overcome the limits

of human thinking, but merely aligns ML results with existing scientific knowledge.

Keywords: machine learning; black box; epistemic opacity; natural science; bias; limits of thinking; hidden dynamics; subjective assumptions

Введение

Интеграция методов машинного обучения (machine learning, ML) в естествознание показала свою практическую эффективность. Аналитические отчеты о научной деятельности, произведенной с помощью искусственного интеллекта и машинного обучения [45], специальные исследовательские группы[1] и программное обеспечение[2] – тому яркое свидетельство. По всей видимости, это единственный выход из ситуации, где «совокупный объем всех когда-либо созданных компьютерных алгоритмов ничтожно мал по сравнению с объемом всех человеческих знаний» [9, с. 26], что касается в том числе и естественных наук. То есть при «программировании вручную» невозможно использовать все существующие вычислительные мощности, представляющие собой в современном мире доступный и дешевый ресурс [9, с. 25–27], и поэтому необходимо создать ситуацию, когда «компьютеры не надо программировать: они программируют себя сами» [1, с. 13]. Эту задачу и выполняют «обучающиеся алгоритмы (learning algorithms) – те, что создают другие алгоритмы, обученные на основе данных» [1, с. 29].

Так постепенно складывается новая интеллектуальная среда – концептуальное «пространство (continuum) между механистическими моделями и моделями машинного обучения, где научное знание и машинное обучение объединены синергетическим образом» [47, p. 2]. Однако этот процесс не может считаться завершенным, пока не решены ключевые теоретические проблемы в этой области, и в первую очередь проблема «черного ящика».

---

[1] См., например: SciML Research Group at Brown University, 2024. URL: https://sites.brown.edu/bergen-lab/ (дата обращения: 15.06.2024); Texas A&M Institute Of Data Science Scientific Machine Learning Lab, 2024. URL: https://sciml.tamids.tamu.edu/ (дата обращения: 15.06.2024).

[2] См., например: Julia’s SciML, 2024. URL: https://juliapackages.com/u/sciml (дата обращения: 15.06.2024); SciML: Open Source Software for Scientific Machine Learning, 2024. URL: https://sciml.ai/ (дата обращения: 15.06.2024).

Постановка проблемы: теоретическое обоснование и прогностическая эффективность методов ML в естественных науках

Основные контуры этой проблемы впервые очертил британский философ Пол Хамфрис, хотя, что примечательно, изначально речь шла не о машинном обучении, а о компьютерных симуляциях [28]. Размышляя об основаниях вычислительной науки (computational science)[3], он сформулировал понятие «эпистемической непрозрачности» (epistemic opacity), которое впоследствии станет одним из ключевых в эпистемологии ML. Хамфрис его использовал для описания изменений в методологии научного исследования, датируемых 40-ми годами XX в., когда, с его точки зрения, вычислительная наука пересекла рубеж человеческого понимания: «Мы столкнулись с проблемой, которую можно назвать *антропоцентрическим затруднением (anthropocentric predicament)*, – как мы, люди, можем понять и оценить основанные на вычислениях научные методы, которые превосходят наши собственные способности и действуют способами, которые мы не можем полностью понять» [27, p. 134]. Особенно ярко эта проблема проявилась в связи с использованием в естественно-научных исследованиях компьютерных симуляций, где «большинство этапов процесса недоступны для прямого контроля и верификации» [28, p. 148], что «может привести к потере понимания, поскольку в большинстве традиционных статических моделей наше понимание основано на способности *разложить процесс между входными и выходными данными модели на отдельные этапы* (курсив наш. – *Авт.*)» [Ibid.][4].

В дальнейшем, когда применение в естествознании методов машинного обучения существенно расширилось, а сам термин «машин-

[3] «Новый вид научного метода», призванный решить проблему «разрешимости [теоретических, аналитических] моделей» (solvability of models), при использовании которого уже нельзя «ограничиться ресурсами человеческого разума» [28, p. 50–51].

[4] Или в другой, более обобщенной и чаще цитируемой формулировке: «Процесс по своей сущности эпистемически непрозрачен (essentially epistemically opaque) для *X* тогда и только тогда, когда вследствие природы *X* невозможно, чтобы *X* знал все эпистемически значимые элементы этого процесса» [27, p. 139].

ное обучение» прочно вошел в научный обиход, понятие «эпистемическая непрозрачность» можно встретить уже и у других авторов в связи с проблемой «черного ящика» в практике ML [12, p. 44; 18, p. 1090; 34, p. 40; 39, p. 85–86; 50; и др.]. Понятие «черный ящик» характеризует работу таких моделей, которые «точно связывают входные данные с выходными» [35, p. 1342], но «тем не менее *не могут явно отобразить (причинно-следственные) связи между ними* (курсив наш. – *Авт.*)» [Ibid.], что вполне соответствует исходному определению «эпистемической непрозрачности» у Хамфриса[5].

Так или иначе, но именно по причине проблемы «черного ящика» / «эпистемической непрозрачности» представление о возможности использования методов машинного обучения в области естествознания вызывает справедливые нарекания: «...Хотя общей конечной целью моделей науки о данных является создание работающих моделей (actionable models), процесс открытия знаний в научных областях на этом не заканчивается. Скорее, это перевод изученных паттернов и взаимосвязей в *интерпретируемые* теории и гипотезы, который ведет к развитию научного знания, например, путем объяснения или открытия физических причинно-следственных связей (mechanisms) между переменными. Следовательно, даже если модель “черного ящика” обеспечивает несколько более точную работу, но неспособна обеспечить механистическое понимание лежащих в ее основе процессов, ее *нельзя использовать в качестве основы для последующих научных разработок* (курсив наш. – *Авт.*)» [30].

То есть центральная проблема, которая встает перед научным сообществом, испытавшим «антропоцентрическое затруднение», – это *отсутствие фундаментального теоретического обоснования* у научно-исследовательской практики, использующей машинное обучение, как прямое следствие существования «черного ящика»

[5] Несмотря на бросающееся в глаза структурное сходство проблем «эпистемической непрозрачности» и «черного ящика» – невозможность пошагового разбиения вычислительного процесса на отдельные этапы между «входом» и «выходом», на будущих этапах нашего исследования нам придется их различать. И не последнюю роль здесь будет играть то обстоятельство, что понятие «эпистемической непрозрачности» у Хамфриса возникает именно в отношении компьютерных симуляций, а не методов машинного обучения.

(«эпистемической непрозрачности»). У нас так и нет общепризнанного объяснения, что является источником ее эффективности и насколько можно доверять полученным с ее помощью результатам. Именно поэтому применение методов машинного обучения сравнивается с «искусством»[6], а то и с «алхимией» или «магией», где определяющим является «разрыв между способностью наблюдателя интерпретировать принципы, лежащие в основе прогнозов, сделанных моделями глубокого обучения, и их точностью или эффективностью в определенном контексте» [14, p. 7][7].

Можно, конечно, пойти по пути наименьшего сопротивления и объявить отсутствие теоретического обоснования у методов ML в естествознании более-менее нормальным положением вещей. Мол, просто у нас теперь будут две относительно самостоятельные линии развития научных исследований: одна, «антропоцентрическая», направлена на поиск «объяснения» (explanation), на раскрытие причинно-следственных связей в природных процессах, а другая, «автоматизированная», – на «прогнозно-ориентированные исследования» (predictively oriented research) [42]. Одна ставит перед собой объяснительные цели и достигает их с помощью механистических моделей (mechanistic models), а другая разрабатывает прогностические модели и тем самым в конечном счете влияет на изменение целей научного исследования, порождая различные формы компромиссов между предсказанием и объяснением [35, p. 3148].

Эта же мысль может принимать иную форму. Например, одна линия развития будет касаться «контекста обоснования (justification)», а другая – «контекста открытия» [18, p. 1091], одна соответствует поисковым экспериментальным практикам (exploratory experimentation), которые имеют дело с «феноменологическим» уровнем научного знания, а другая – практикам, имеющим дело с уровнем

[6] «...Практическое применение ИИ и машинного обучения остается преимущественно искусством. Несмотря на достигнутый значительный прогресс, потребуются новые достижения в основаниях ИИ» [45, p. 99].

[7] Впервые публичное сравнение машинного обучения с алхимией прозвучало в выступлении А. Рахими в 2017 г. на конференции по нейронным системам обработки информации (NeurIPS) [41].

«теоретическим» [40, p. 913–914], и т.д. При этом основная идея об относительной независимости «объяснения» и «предсказания» внутри процесса научного познания так или иначе остается ведущей[8].

Конечно, можно было бы предположить, что эти две линии развития исследований будут дополнять, обогащать друг друга, обмениваясь достигнутыми результатами и совместно добиваясь поставленных целей. Но в текущей ситуации подобные «политкорректные» заявления выглядели бы скорее как дань вежливости, чем как глубокое теоретическое высказывание. Строго говоря, непонятно, как именно следует определять ценность создаваемых с помощью ML прогностических моделей для научно-исследовательской работы, т.е. – просто-напросто отсутствует *эпистемологический критерий*, в соответствии с которым такое «взаимообогащение» могло бы происходить.

Но если все-таки не пытаться выдать нужду за добродетель, то ключом к решению поставленной проблемы будет *более точное определение той роли*, которую играют методы машинного обучения в естествознании. Для этого в настоящей статье мы предпримем следующие шаги:

1) опишем ситуацию в естествознании, которая породила потребность в методах машинного обучения;

2) определим ключевой критерий, которому должно удовлетворять предлагаемое решение проблемы;

3) рассмотрим действующую стратегию интеграции машинного обучения и естествознания, какой она предстает в понимании ученых-компьютерщиков, и попытаемся определить, можно ли с ее помощью продвинуться в решении поставленной задачи теоретического обоснования методов ML в естествознании.

[8] Тем не менее существуют исключения (например, превращение предвзятости в теоретические гипотезы С. Энни и М. Херрие [20] или «агностическая наука» Д. Наполетани и др. [38] и т.п.), которые действительно предлагают решения проблемы теоретического обоснования методов ML. Полученные в их рамках результаты мы будем обсуждать на более поздних этапах нашего исследования.

Столкновение с пределами научного познания: проблема интегрируемости дифференциальных уравнений и «барьер интуиции»

Первый вопрос, на который необходимо получить ответ на пути решения поставленной задачи: в силу чего в естественных науках вообще оказались востребованы методы машинного обучения? Дело в том, что в настоящий момент естествознание, чье развитие во многом происходило как создание все новых способов математического моделирования физических систем, достигло *пределов*, которые носят *объективный* («*структурный*») характер. То есть они не могут быть преодолены путем дальнейшего создания более сложных математических моделей. Это, во-первых, проблема интегрируемости дифференциальных уравнений (сложность моделируемого физического процесса) и, во-вторых, проблема многомерности фазового пространства (сложность пространства, в котором моделируется процесс) (см. подробнее: [7]).

Сначала, в эпоху «тихой научной революции» XVIII в., произошел скачок в развитии аналитических методов в естествознании. Главным образом это касается использования аппарата дифференциальных уравнений для исследования физических систем, в том числе и для выражения законов природы [6, с. 101–102; 43]. Господство аналитических методов продолжалось вплоть до столкновения с проблемой интегрируемости дифференциальных уравнений, которая столь ярко проявила себя в «задаче *n* тел» [2, с. 76–79; 3, с. 41, 214]. Проблема носила именно объективный характер, поскольку не решалась посредством совершенствования самих аналитических методов: сложное, нелинейное поведение физических систем оказалось их объективным свойством, а не дефектом используемого математического метода [7, с. 80].

Подступиться к решению этой проблемы стало возможным через обращение к «качественным методам» [4]. Это означает, что предпринимается попытка описывать состояние физической системы минимально возможным набором величин (так называемые «степени свободы»), которые помещаются в собственное пространство (называемое «фазовым»), не имеющее прямого физического аналога. И затем в этом пространстве производится исследование физической

системы посредством анализа топологических особенностей ее фазового портрета (наличие замкнутых траекторий, точек притяжения и отталкивания и т.д.) [16, p. 6–7].

По сути, «качественные» методы открывают возможность геометрического исследования, если угодно, «геометрической расшифровки» векторного поля, которое возникает в фазовом пространстве. Таким образом, их отличие от способа моделирования, который предполагает определенное сходство между объектом в физическом пространстве и его математической моделью, состоит только в том, что человеческая интуиция на этот раз применяется не к физическому пространству, но к некоторому его *абстрактному эквиваленту, искусственно созданному пространству*. Это пространство открывает доступ к «имманентным» состояниям системы, т.е. – без внешнего взгляда наблюдателя, воплощенного в $(n + 1)$-м измерении модели физического пространства [16, p. 4–6], чтобы извлечь дополнительную информацию о ее свойствах. А это означает, что исследование физической системы посредством составления ее фазового портрета возможно постольку, поскольку можно свести n-мерное фазовое пространство, отражающее число степеней свободы физической системы, к *интуитивно доступным* – трехмерным или двумерным – геометрическим образам.

Однако в том-то и дело, что эта операция далеко не всегда осуществима. Например, чтобы создать фазовый портрет движения звезды в гравитационном поле («система Хенона – Хейлеса»), для описания траекторий движения достаточно двух пространственных и двух импульсных координат, т.е. требуется – всего четыре измерения, которые можно отобразить на трехмерной гиперплоскости [2, с. 78]. Но если мы попытаемся таким же образом решить задачу $n > 3$ тел, то свести результат к двумерным или трехмерным изображениям уже будет чрезвычайно трудно[9]. Дело в том, что несмотря на то что существуют приемы работы с четырехмерными геометрическими

[9] Задача n тел имеет аналитическое решение с помощью рядов [17; 46], которое не связано с ограничениями интуиции, но бесполезно с практической точки зрения [17, p. 69], т.е. не позволяет прогнозировать физические явления, ибо требует огромного количества слагаемых. Собственно, создание качественных методов и теории динамических систем было реакцией на невозможность предоставить аналитическое решение.

объектами [10], в случае пятимерного и более многомерного пространства те же самые приемы приводят к резкому увеличению количества графических образов, которые необходимо одновременно воспринимать как один и тот же визуальный объект. Строго говоря, с каждой новой размерностью количество таких изображений возрастает как факториал, что в перспективе требует таких интеллектуальных усилий, которые значительно превышают возможности человеческого мышления.

Таким образом, на этом этапе возникает обоснованное подозрение, что мы исчерпали все варианты совершенствования математических методов для описания физической реальности. Кажется, что дальнейшее продвижение уже невозможно, ибо непонятно, как преодолеть этот «барьер человеческой интуиции», которая просто не приспособлена для понимания полученных в рамках этого описания результатов[10].

Как ни бейся, но мы остаемся людьми со всеми нашими эволюционно-биологическими (или «трансцендентальными» в кантовском смысле) когнитивными ограничениями. Доступный нам опыт – это опыт наблюдения, развертываемый в физическом пространстве (или, по Канту, ограниченный условиями чистой интуиции, “Anschauung”, a priori). И даже если посредством качественных методов мы смогли осмыслить результаты, которые носят контринтуитивный характер (описание сложного, нелинейного поведения физических систем), то в любом случае нам необходимо свести эти результаты к интуитивно доступной форме – чтобы элементарно их осмыслить.

«Вилка» аналитических и качественных методов: теоретический парадокс применения ML

Фактически мы получили «вилку». С одной стороны, за счет качественных методов, «геометрической расшифровки» дифференциальных уравнений мы смогли получить информацию о физических

[10] Например, необходимость работать в 6- или 9-мерном пространстве возникает в задаче механики деформируемого твердого тела при описании явления пластичности, поскольку поверхность, определяющая границу между упругим и пластическим поведением среды, располагается в 9-мерном пространстве [5, с. 423].

системах, для которых нет корректного аналитического описания (т.е. нельзя использовать решение дифференциального уравнения для предсказания конкретных физических явлений). Но с другой стороны, мы можем это делать ровно до тех пор, пока результаты этой «геометрической расшифровки» нам интуитивно доступны. Иными словами, выбор математических методов в естествознании всегда предполагает наличие объективных ограничений для нашего мышления, с которыми необходимо считаться [7][11].

И чтобы обойти эти ограничения, нам пришлось бы, строго говоря, каким-то образом мыслить на уровне, который принципиально недоступен нашему мышлению, и пользоваться его результатами, что, очевидно, представляет собой неразрешимый парадокс. Или нет?... Собственно, это и есть тот момент, когда в игру вступает машинное обучение.

Ученые-компьютерщики (computer scientists), пытаясь прояснить причины использования ML в естествознании, не всегда заботятся о концептуальной строгости формулировок, делая акцент на конкретных программных технологиях и достигаемых с их помощью практических результатах. Но в целом есть несколько ключевых теоретических положений, которые можно выделить, несмотря на некоторые трудности в унификации используемого ими словаря описания.

Во-первых, основная причина использования ML в естествознании – это необходимость преодоления пределов научного познания, связанных со сложностью физических систем, поскольку «модели машинного обучения (например, нейронные сети) при наличии достаточного объема данных могут находить структуру и закономерности

[11] Когда Хамфрис пишет о неразрешимости большинства аналитических моделей и необходимости их включения в «вычислительные шаблоны», которые могут обеспечить их практическую применимость [28, p. 60–67], или о замещении их компьютерными симуляциями [28, p. 114–116], то он не указывает, *где именно проходят пределы человеческого мышления*, сводя их исключительно к «ограничениям вычислительных способностей» [28, p. 52]. Поэтому от его внимания ускользает «вилка» аналитических (сложность физических систем) и качественных (ограничения человеческой интуиции) методов, которая уже, в свою очередь, приводит к тому, что приходится обращаться к численным методам и формировать на их основе «вычислительные шаблоны».

в задачах, сложность которых исключает прямое программирование (explicit programming) точной физической природы системы» [47, p. 5].

Во-вторых, «такие задачи возникают вследствие того, что «физические модели многих сложных природных систем в лучшем случае “частично” известны как законы сохранения и не дают замкнутой системы уравнений, если только определяющие законы (constitutive laws) не были постулированы» [29, p. 436]. И эта ситуация, когда «физика известна частично, т.е. [известен] закон сохранения, но не определяющее соотношение (constitutive relationship)», «является наиболее общим случаем» [29, p. 423].

В-третьих, совокупность таких задач образует область «скрытой динамики» (hidden dynamics)[12], где «данных сколько угодно, а *физические законы или определяющие уравнения (governing equations) ускользают от понимания (remain elusive)* (курсив наш. – *Авт.*) [13, p. 229]... Все модели носят приближенный характер, и чем больше сложность, тем больше подозрений вызывают аппроксимации. Решение о корректности модели становится более субъективным, растет потребность в методах автоматизированного построения модели, которая проливала бы свет на фундаментальные (underlying) физические механизмы. Часто оказывается, что существуют скрытые переменные (hidden variables), которые имеют прямое отношение к динамике явления, но могли быть не измерены. Выявление этих скрытых эффектов – главная задача управляемых данными методов» [13, p. 234].

В-четвертых, конечные результаты естественно-научного исследования в области «скрытой динамики» включают в себя *достаточно большое количество допущений / предположений* (assumptions)[13], чтобы поставить под сомнение их объективность: «...“PBM” (“моделирование, используемое в физике”, physics-based modeling. – *Авт.*) игнорирует любую неизвестную физику (unknown physics), которая,

[12] Помимо монографии С. Брантона и Н. Куца, в указанном смысле понятие «скрытой динамики» используется, например, в статьях П. Ху и др. [26] и А. Яздани [49].

[13] В настоящей статье мы переводим «assumption» как «допущение», но в зависимости от контекста через символ «/» может прибавляться «предположение», чтобы придать термину нужный смысловой оттенок, ибо оригинальный термин объединяет оба значения.

например, не может наблюдаться напрямую или необъяснима. Основываясь на началах уже известной физики, “PBM” требует вывода одного или нескольких определяющих уравнений (governing equations) для системы. Этот вывод может включать в себя некоторые допущения, например то, что определяющими уравнениями учитывается только часть физики (partial physics). Чаще всего эти уравнения трудно решить аналитически. Поэтому чтобы решить их численно (за разумное время), мы делаем дополнительные допущения, что приводит к еще большей потере физики» [11, p. 182].

Таким образом, ответ на вопрос об основаниях применения ML в естествознании должен представлять собой, если угодно, разрешение теоретического парадокса: нужно объяснить, каким образом методы машинного обучения открывают доступ в область знания, которая принципиально недоступна человеческому мышлению. Если Брантон и Куц могут ограничиться указанием на «методы автоматизированного построения модели» как на спасательный круг, а потом сразу сослаться на конкретные практические результаты, достигнутые с их помощью [13, p. 234–235], то для эпистемологического исследования этого недостаточно: необходимо выяснить, насколько эти результаты в принципе заслуживают доверия.

И с точки зрения нашей перспективы дело не в том, чтобы определить, что могло бы явиться причиной убежденности различных когнитивных агентов (см., например, [19; 50]), и не в разработке способов добиться «прозрачности»/«интерпретируемости» алгоритмов ML (см.: например: [22; 34]). Задача состоит в том, чтобы выяснить, как осуществляется выход человеческого мышления за собственные пределы, как происходит пересечение с недоступным, «чужеродным» способом рассуждения/познания. По сути, от нас требуется описать *некий способ взаимодействия человеческого мышления с той областью знания, что по определению ускользает от человеческого понимания, «остается неуловимой» (remain elusive)*. И поэтому любое теоретическое обоснование, претендующее на действительное решение проблемы, должно принимать в расчет эту трудность, парадоксальную сложность поставленной задачи. И решение, которое уходит от сложностей и не соответствует этому критерию, бьет мимо цели.

Внедрение научных результатов в алгоритмы ML: ожидание «кумулятивного эффекта»

Т. Никлес в своей статье [39] задался вопросом о последствиях решения проблемы «черного ящика»». Он пришел к выводу, что в этом случае мы останемся в рамках подхода, ориентированного на человеческое мышление, и неизбежно утратим тот новаторский потенциал (эффект «чужеродного мышления», alien reasoning), который содержится в инструментах машинного обучения [39, p. 96]. Иными словами, чем меньше мы понимаем, тем больше знаем, и наоборот. Осознание этой дилеммы позволило Никлесу обнаружить круг в исследованиях «прозрачного»/«объяснимого» ML: «В будущем мы можем разработать машины с радикально отличающейся архитектурой или с интегрированной комбинацией архитектур, которая уменьшит непрозрачность и другие ограничения. Но *как это сделать, не прибегая к человеческим ограничениям в рассуждении* (курсив наш. – *Авт*.)?» [39, p. 97]. Это в числе прочего означает, что способ взаимодействия человеческого мышления с недоступной ему областью знания не может быть получен непосредственно в результате прикладных исследований с использованием ML, поскольку по большому счету является *условием самой возможности их проведения*.

Таким образом, для того чтобы методы машинного обучения и правда оказались применимыми в области естествознания, должен существовать *комплекс обстоятельств, который предшествует их применению* и, соответственно, имеет какое-то иное объяснение помимо (или вместо) того, какое может быть сформулировано на уровне ML. Именно по этой причине мы с определенной настороженностью относимся к результатам объяснимого/«прозрачного»/интерпретируемого ML с точки зрения их эпистемологической ценности. Дело в том, что они производятся с помощью тех же самых инструментов ML, а следовательно, методы, которые служат объяснению/интерпретации, сами должны быть точно также интерпретированы/объяснены[14].

[14] Например, чтобы обосновать использование линейного классификатора нелинейных моделей LIME (Linear Interpretable Model-agnostic Explanations), приходится проводить аналогии с математической идеализацией в «традиционном» научном исследовании [22, p. 548]. Еще по теме см. работу З. Липтона «Мифы об

Однако это не означает, что результаты прикладных исследований не будут информативны для решения эпистемологической задачи. Напротив, мы предполагаем, что в конечном счете именно анализ инструментов ML, применяемых в естествознании, может продемонстрировать, каким образом осуществляется взаимодействие человеческого мышления с недоступной для него областью знания («скрытой динамикой»). Только при этом мы должны не ограничиваться констатацией их эффективности и учитывать, что имеем дело только с *отражением, «технологической» проекцией* некоторых обстоятельств, относящихся к фундаментально-теоретическому уровню познания, которые и сделали возможным достигнутый с помощью инструментов ML практический результат. Мы полагаем, что какой бы трудной проблемой ни был парадокс «скрытой динамики», его решение так или иначе проявляет себя в сфере прикладных исследований естествознания с применением ML и, соответственно, прояснит, *что именно* позволяет инструментам ML быть эффективными. Иными словами, связи между физико-математическими моделями и методами ML на уровне программного обеспечения должны отражать связи между *их предпосылками/условиями* на эпистемологическом уровне[15].

Начать разбор этой темы следует с того, что у ученых-компьютерщиков выработалось собственное «рабочее» представление об ос-

интерпретируемости моделей», где вообще ставится под сомнение «прозрачность» линейных моделей, достижение которых во многом определяет содержание практики «объяснимого ML», по сравнению с моделями глубокого обучения [34, p. 42]. Но в целом эта практика и теоретическая значимость достигнутых с ее помощью результатов требуют отдельного изучения.

[15] У этого хода мысли, выражаясь юридическим языком, существует прецедент. В свое время Д. Чалмерс, разрабатывая решение для «трудной проблемы сознания», предпринял попытку описать структуру сознательного опыта через его *информационные корреляты*, связанные с понятием осведомленности (awareness). Он предположил, что связи между физическими состояниями, соответствующими осведомленности, сообщают нечто об организации сознательного опыта, поскольку они реализуют этот опыт, хотя последний к ним и не редуцируется [8, с. 298–303]. Если угодно, связь физико-математических моделей и инструментов ML, приносящая практический результат, в определенной перспективе можно рассматривать как *«программный» коррелят* связей человеческого мышления и недоступной ему области знания.

новных принципах интеграции методов ML и естествознания. Его суть в следующем: имеет место совместное решение научных задач, в рамках которого моделирование, используемое, например, в физике (physics-based modeling), и моделирование, управляемое данными (data-driven modeling), помогают преодолеть недостатки друг друга [11, p. 182; 30; 47, p. 2]. Как мы видели выше, в первом случае такими недостатками являются сложность с аналитическими выводом определяющих уравнений для физических систем и необходимость делать предположения при численных решениях [11, p. 182; 13, p. 234]. Но и машинное обучение, со своей стороны, также извлекает «выгоду» из этой ситуации: в числе ее недостатков указываются «зависимость от объема данных (data-hungry) и характер "черного ящика", плохая обобщаемость, врожденная предвзятость (inherent bias) и отсутствие строгой теории для анализа устойчивости модели» [11, p. 182][16].

Конкретные формулировки могут различаться, но в целом эта схема своеобразного симбиоза машинного обучения и естествознания транслируется из публикации в публикацию – независимо от поставленных задач и способов классификации: о нем пишут Дж. Уиллард и др. (physics-ML), С. Блэксет и др. (hybrid analysis and modeling), Г. Карниадакис и др. (physics-informed machine learning), А. Карпатне и др. (theory-guided data science) и другие исследователи.

Именно на врожденную предвзятость ML, inherent bias, или, точнее, inductive bias, поскольку она связана с возможностью индуктивных выводов, следует обратить внимание. Напомним, что как раз с субъективными допущениями («необходимостью делать предположения»), которыми сопровождается использование численных методов в естествознании, было связано обращение к инструментам ма-

[16] Или, как пишут в своем обзоре Дж. Уиллард и др., «применение машинного обучения (ML) в качестве "черного ящика" в научных областях имело ограниченный успех по ряду причин. Во-первых, несмотря на то что современные ML-модели способны фиксировать сложные пространственно-временные взаимосвязи, для обучения им требуется слишком много размеченных данных, которые редко доступны в реальных условиях. Во-вторых, ML-модели часто дают противоречивые с научной точки зрения (scientifically inconsistent) результаты. В-третьих, модели ML могут фиксировать взаимосвязи только в доступных обучающих данных и, следовательно, не могут быть обобщены на сценарии, не относящиеся к выборке (out-of-sample scenarios)» [47, p. 3].

шинного обучения. Но поскольку само машинное обучение от них также не застраховано, то, следовательно, текущая стратегия интеграции заключается в создании своего рода «кумулятивного эффекта», который, помимо борьбы с предвзятостью в методах естествознания, позволяет бороться и с предвзятостью в алгоритмах ML, т.е. с inductive bias.

«Окна» для внедрения научных знаний: inductive bias и область решаемой задачи

Родословную плохо переводимого на русский язык понятия «inductive bias»[17] традиционно возводят [20, p. 6; 24; 44, p. 9991] к работе Т. Митчелла 1980 г.: «Только если программа имеет другие источники информации (помимо “согласованности с обучающими примерами” – *Авт.*) или предвзятость (bias) в выборе одного обобщения среди других, она может разумным образом (non-arbitrarily) классифицировать элементы (instances), находящиеся за рамками обучающего набора. В этой статье используется термин “*предвзятость*” (*bias*) для обозначения любого *основания для выбора одного обобщения вместо другого, отличного от строгого соответствия наблюдаемым примерам обучения*» [37, p. 185].

Уже в 90-е годы подоспело формальное обоснование этой идеи – теорема “No free lunch” (NFL) [20, p. 6; 25; 32, p. 100–101; 44, p. 9991]. Она показала, что «для получения обобщения необходим некоторый набор предпочтений (preferences) (или inductive bias) в пространстве всех функций (которые бы выражали зависимости в данных. – *Авт.*), что не существует полностью универсального алгоритма обучения, что любой алгоритм обучения будет лучше обобщать на некоторых распределениях и хуже – на других» [23, p. 4].

В самом общем смысле термин «inductive bias» означает допущения/предположения (assumptions), которые позволяют разработчику модели машинного обучения сделать выбор в пользу «одного реше-

[17] Собственно, именно с данным обстоятельством связано то, что в настоящей статье мы используем это понятие без перевода, а русский вариант «bias» – «смещение», «склонность», «отклонение», «предвзятость», «предубеждение», «предрассудок», «крен», «искажение» и т.д. – отдельно от данного словосочетания определяется контекстом, при этом мы указываем на оригинальный термин.

ния (или интерпретации) вместо другого независимо от наблюдаемых данных» [36, p. 6]. Но это не должно стать критической проблемой для машинного обучения, поскольку, как подсказывает Митчелл, «при изучении обобщений для конкретной цели можно *ограничить число рассматриваемых обобщений, обратившись к знаниям из области решаемой задачи* (*task domain*) (курсив наш. – *Авт.*)» [37, p. 188].

Таким образом, если установлено, что условием эффективного обобщения является наличие у используемой программы «других источников информации» помимо «согласованности с обучающими примерами», то почему бы таким источником не стать, например, современному естествознанию?... Тогда отсутствие «универсального алгоритма обобщения» в соответствии с теоремой NFL, не является такой уж плохой новостью. Дело в том, что требование наличия как минимум двух автономных источников информации для успешного обобщения в ML (наличие данных для индуктивного вывода – «первый источник информации» и источник предвзятости в индуктивном выводе – «второй источник») позволяет использовать, пожалуй, наиболее надежный источник из всех существующих – результаты современных естественно-научных исследований. И таким образом эффективное обобщение в методах машинного обучения, по сути, не слишком зависит от сложностей, связанных с проблемами индукции[18].

Так inductive bias образуют своего рода «окна» для внедрения в модели машинного обучения результатов естествознания. Подобно тому как заменяются неисправные детали в механизмах на более

[18] Мы сознательно не рассматриваем еще более глубокий пласт этой темы, касающийся «проблемы индукции» Юма и различных вариантов ее решения. На данный момент мы полагаем, что решение проблемы теоретического обоснования методов ML в естествознании не зависит напрямую от дискуссий об основаниях индукции. Больше информации о связи теоремы NFL, проблем индукции и задачи обоснования выводов, сделанных посредством алгоритмов ML, можно найти в обстоятельной статье Т. Штеркенбурга и П. Грюнвальда [44], в которой различаются «локальное обоснование выводов», или «обоснование относительно модели» (model-relative justi?cation), и «глобальное обоснование выводов алгоритмов машинного обучения», которое остается недоступным, поскольку как раз «сводится к исходной проблеме индукции» [44, p. 10003]. Судя по всему, «локальное обоснование выводов» вполне подходит под предложенное Митчеллом использование «знания из области решаемой задачи» [37, p. 188], что, как мы покажем ниже, не является решением проблемы.

новые, в алгоритмы ML под видом inductive bias можно «имплантировать» научно обоснованные результаты, уже достигнутые в рамках естествознания.

Дж. Карниадакис и др. в своей статье [29] как раз формулируют этот подход: «Ни одна прогностическая модель не может быть построена без допущений (assumptions), и, как следствие, от моделей ML без соответствующих предубеждений/склонностей (biases) нельзя ожидать никакой эффективности в обобщении» [29, p. 424]. Соответственно, существуют три пути, чтобы «улучшить обобщение моделей машинного обучения путем включения в них [знаний из] физики», а именно: «ввести предубеждения/склонности (bias)» в выборку данных, в архитектуру модели и/или в процедуру обучения [Ibid.][19]. Этим авторам вторят Ю. Чен и Д. Жан: «Исследователи пытаются внедрить знания предметной области в процесс моделирования машинного обучения, включая следующие три этапа: предварительную обработку данных, проектирование структуры модели, а также архитектуру штрафов и вознаграждений» [15].

Например, Карниадакис с соавторами ссылаются на статью А. Кашефи и др. [31], где для обучения нейросети при решении задачи предсказания полей скоростей и давления вокруг нерегулярных геометрий тел предлагается использовать наборы данных, которые представляют собой ее численные решения. Такие данные сгенерированы для различных форм сечений цилиндра: круга, квадрата, треугольника, прямоугольника, эллипса, пятиугольника и шестиугольника. Эти формы варьируются по ориентации и масштабам, что создает, по замыслу авторов упомянутой статьи, разнообразный набор данных [31, p. 4]. Или в аналогичной задаче, связанной с моделированием турбулентного потока вблизи обтекаемого тела, для учета галилеевских преобразований предлагается внедрить в архитектуру нейросети

[19] Здесь имеется терминологический нюанс, а именно использование понятия «inductive bias» в широком и узком смыслах. Собственно, Карниадакис и др. называют «inductive bias» только те предубеждения/склонности (bias), что связаны с архитектурой модели машинного обучения, а те, что относятся к выборке данных или к процедуре обучения, называются соответственно «observational biases» и “learning biases” [29, p. 424]. Другие авторы, принимая саму классификацию, тем не менее используют «inductive bias» в качестве общего для всех предубеждений/склонностей (bias) термина (см., например: [20, p. 2]).

базис тензоров, включающий 10 изотропных тензоров и пять скалярных инвариантов, при комбинации которых гарантируется, что предсказанный тензор анизотропии напряжений Рейнольдса будет инвариантным относительно вращения системы координат [33, p. 158–160]. В задаче прогнозирования ламинарного потока жидкости за цилиндром вводится специальный параметр регуляризации, который стимулирует сохранение устойчивости моделируемого решения с целью обеспечения лучшего обучения и обобщения модели [21].

Эти примеры наглядно демонстрируют, как можно тремя различными способами воспользоваться «окнами», которые представляют собой inductive bias, для внедрения научных знаний в алгоритмы ML: на уровне данных, на уровне архитектуры модели и на уровне параметров процедуры обучения. И в результате может показаться, что это чрезвычайно удачный ход мысли – убить одним выстрелом сразу двух зайцев. Ведь наряду с устранением допущений/предположений в методах естествознания прямым следствием интеграции естествознания и машинного обучения «является *интерпретируемость* (*interpretability*), когда традиционные модели ML часто представляют собой “черный ящик”, а включение научных знаний в инструменты ML может пролить свет на физический смысл и возможные интерпретации протекающего процесса» [48]. То есть интегрирование научных знаний в структуру ML позволяет снизить критичность характерной для моделей машинного обучения проблемы «черного ящика», поскольку благодаря этому создается определенное представление о каузальных механизмах, которые сработали между входными и выходными данными.

Между парадоксом в теории и противоречием на практике: «странная война» с предвзятостью

Таким образом, стратегия интеграции ML и естествознания заключается в замене необоснованных допущений/предположений научно подтвержденными результатами. При этом предполагается, что это даст своего рода «кумулятивный эффект»: во-первых, позволит алгоритмам машинного обучения минимизировать предвзятость, связанную со спецификой индуктивных выводов (теорема NFL), а во-вторых, эти алгоритмы, в свою очередь, ликвидируют предвзятость

в математических методах естествознания («субъективные решения о корректности модели» у Брантона и Куца [13, p. 234], «необходимость делать предположения при численных решениях» у Блэксета и др. [11, p. 182] и т.д.). Однако на уровне теории здесь возникает зазор, который делает проблематичной реализацию этой стратегии.

В первую очередь следует отметить, что внедрение научных знаний в алгоритмы ML через inductive bias не равнозначно задаче полной ликвидации последних. Inductive bias используются просто как входные точки, «окна», чтобы создать эффективный симбиоз между научными методами и методами машинного обучения. Нигде не декларируется задача полного и окончательного искоренения предвзятости, ученые-компьютерщики выбирают гораздо более осторожные формулировки вроде таких, как «принуждение моделей к физической согласованности может эффективно сократить пространство поиска решения» [47, p. 11], «предварительные знания или ограничения могут привести к более интерпретируемым методам ML» [29, p. 423], или декларируется стремление «изучать зависимости, которые... имеют больше шансов представить причинно-следственные связи... также пытаются достичь лучшей обобщаемости, чем модели, основанные исключительно на данных» [30].

Поэтому в алгоритмах ML остается довольно много допущений/предположений, которые никак не связаны с достижениями естествознания. Скажем, в первом приведенном выше примере из статьи А. Кашефи и др. имеется допущение/предположение разработчиков, что ограниченного спектра обтекаемых сечений (круг, квадрат, треугольник, пятиугольник и т.д.), для которых были получены численные решения, будет достаточно при решении задачи для тела произвольной формы [31, p. 2]. И это не просто эмпирический факт – за ним, видимо, стоит фундаментальное теоретическое ограничение.

Дело в том, что, как мы говорили выше, потребность в методах машинного обучения как раз определяется фактором столкновения с объективными пределами научного познания. То есть необходимо исследовать область «скрытой динамики», где «истинные физические механизмы» и «определяющие уравнения» («физические законы») ускользают от человеческого понимания. И в этом случае ограни-

чение, по Митчеллу, «числа рассматриваемых обобщений» при создании алгоритмов ML «знаниями о предметной области» [37, p. 188] позволит добиться только того, что их результаты *не будут противоречить уже известным знаниям*. Но это в принципе не может быть тем способом, благодаря которому *становится возможным преодоление объективных пределов познания*, т.е. – открытие в физических процессах закономерностей, ускользающих от человеческого мышления.

При обзоре методов машинного обучения в естествознании это обстоятельство выражается, например, в возможности развести «внедрение знаний» (knowledge embedding) и «открытие знаний» (knowledge discovery) как два различных способа «интегрировать знания предметной области с моделями, основанными на данных» [15]. По сути, они относятся в разным фазам применения ML, где «процесс непосредственного извлечения определяющих уравнений из наблюдений и экспериментальных данных» может сопровождаться «увеличением производительности моделей машинного обучения» как раз вследствие внедрения научных знаний [Ibid.].

Это, в свою очередь, означает, что при использовании в естествознании методов ML все равно придется делать такие спорные допущения/предположения, которым не соответствуют никакие научные достижения, и, следовательно, разработчик будет вынужден их совершать на свой страх и риск. А их наличие, напомним, как раз и является поводом для обвинений ML в принадлежности к «искусству» [45, p. 99], «магии» [14, p. 7] или «алхимии» [41].

Таким образом, описанная стратегия создания «кумулятивного эффекта» за счет симбиоза методов естествознания и методов ML не соответствует озвученному нами выше критерию: найти способ взаимодействия человеческого мышления с недоступной для него областью знания. Напротив, борьба с предвзятостью как раз предполагает, что мы просто последовательно устраняем из алгоритмов необоснованные допущения/предположения, включая в них научные знания из доступных познанию областей. Если эту борьбу рассматривать как эпистемическое правило, то это, безусловно, дает практический эффект, но не сильно приближает нас к решению задачи теоретического обоснования инструментов, посредством которых такого эффекта удалось добиться. Скорее, это создает ситуацию,

в которой поставленная проблема предстает во всей своей остроте и неразрешимости.

Можно сказать, на данном этапе наших размышлений мы пришли к констатации противоречия, содержащегося в практике ML на структурном уровне. И судя по всему, оно является отражением теоретического парадокса, сформулированного нами выше. С одной стороны, стратегия интеграции машинного обучения и естествознания состоит в борьбе с предвзятостью, но с другой стороны, эту борьбу следует вести не очень активно, чтобы, не дай бог, не одержать в ней победу. Это похоже на объявление войны с дальнейшей имитацией военных действий.

Если угодно, сами инструменты ML «сопротивляются» теоретическому обоснованию их применения. Они будто бы подталкивают нас к мысли, что какая-то часть содержащихся в них допущений/предположений по умолчанию *должна* остаться необоснованной. Возможно, в этом противоречии заключается причина текущего состояния эпистемологии ML, когда приходится просто констатировать разделение науки на «предсказания» и «объяснения» и смириться с необоснованностью «прогнозно-ориентированных исследований» как с неизбежным злом [18; 35; 39; 42].

Мы остаемся при убеждении, что именно анализ способа интеграции машинного обучения и естествознания может пролить свет на интересующую нас проблему – проблему теоретического обоснования методов ML. Другое дело, что перенос методологических требований естествознания – *максимальное сокращение, минимизация субъективных допущений/предположений* – на практику использования ML в естествознании оказывается не слишком хорошей идеей, поскольку противоречит фундаментальным принципам машинного обучения. Выбор такой стратегии выглядит логично, только если рассматривать ML как *один из* математических методов, которые использовались на протяжении его (естествознания) трехсотлетней истории, когда борьба с предвзятостью действительно имела ключевое значение.

Поэтому мы считаем целесообразным в процессе дальнейших исследований расширить подход, связанный с ролью ML как способа преодоления пределов человеческого мышления. В рамках такого расширенного подхода применение ML означает гораздо более прин-

ципиальное изменение характера естественно-научного познания, чем просто использование нового математического инструмента, а борьба с предвзятостью, соответственно, теряет прежнюю значимость.

Заключение

Текущий этап исследования, связанный с поиском теоретического обоснования методов ML в естествознании, позволяет нам зафиксировать следующие результаты.

Во-первых, в настоящей статье мы показали, почему инструменты машинного обучения проникают в область естествознания. Это происходит из-за «вилки» между аналитическими и качественными/геометрическими методами исследования физической реальности, которые, видимо, достигли пределов своего применения. Если аналитические методы, представленные в основном аппаратом дифференциальных уравнений, сталкиваются с проблемой интегрируемости, то качественные методы упираются в проблему многомерности фазового пространства, в котором моделируется физическая система.

Во-вторых, мы показали, что в результате существования указанной «вилки» образуется область «скрытой динамики» (или «неизвестной физики»), для исследования которой как раз и оказываются востребованы инструменты ML. Поэтому мы предположили, что критерием успешного решения проблемы является объяснение, каким образом методы машинного обучения открывают доступ в область знания, принципиально недоступную для человеческого мышления, что, по сути, представляет собой теоретический парадокс.

В-третьих, мы описали стратегию интеграции естествознания и машинного обучения, которая состоит в борьбе с предвзятостью посредством создания своего рода «кумулятивного эффекта». Речь идет о замене inductive bias, предположений/допущений, содержащихся в конкретных алгоритмах ML, научно обоснованными результатами, и это позволяет частично решить проблему «черного ящика», характерную для машинного обучения ввиду особенностей операции обобщения (теорема «No free lunch»). И как следствие, это позволяет снизить зависимость естествознания от субъективных допущений при использовании численных/приближенных методов.

В-четвертых, мы выяснили, что эта стратегия, используемая в прикладных исследованиях ML, не может стать предпосылкой для теоретического обоснования методов машинного обучения в естествознании, поскольку невозможно полное устранение предвзятости, необоснованных допущений/предположений из применяемых алгоритмов вследствие исходного запроса – задачи открытия физических закономерностей в области «скрытой динамики» («неизвестной физики»), недоступных человеческому пониманию. А потому на текущем этапе нашей работы, по-видимому, следует скорректировать используемую стратегию – искать возможность каким-то образом изменить эпистемологический статус самих допущений/предположений.

Литература

1. *Домингос П.* Верховный алгоритм: как машинное обучение изменит наш мир / Пер. с англ. В. Горохова; науч. ред. А. Сбоев, А. Серенко. М.: Манн, Иванов и Фербер, 2016. 336 с.

2. *Майнцер К.* Сложносистемное мышление: Материя, разум, человечество / Пер. с англ. под ред. и с предисл. Г.Г. Малинецкого. М.: Либроком, 2009. 464 с.

3. *Мухин Р.Р.* Очерки по истории динамического хаоса: Исследования в СССР в 1950–1980-е годы. М.: URSS, 2018. 320 с.

4. *Пуанкаре А.* О кривых, определяемых дифференциальными уравнениями / Пер. с фр. с прим А.А. Андронова. Москва; Ленинград: Гос. изд-во тех.-теорет. лит., 1947. 388 с.

5. *Седов Л.И.* Механика сплошной среды. М.: Наука, 1970. Т. 2. 568 с.

6. *Сухно А.А., Гулин В.В.* Интегрируемость дифференциальных уравнений и «конвертация сложности» // Философия науки. 2024. № 1 (100). С. 67–95.

7. *Сухно А.А., Гулин В.В.* Контингентность сущего и дифференциальные уравнения // Логос. 2023. Т. 33, № 4. С. 95–116.

8. *Чалмерс Д.* Сознающий ум: в поисках фундаментальной теории / Пер. с англ. В.В. Васильева. М.: URSS; Либроком, 2013. 509 с.

9. *Шумский С.А.* Машинный интеллект: Очерки по теории машинного обучения и искусственного интеллекта. М.: РИОР, 2019. 340 с.

10. *Amirkhanov A., Kosiuk I., Szmolyan P., Amirkhanov A., Mistelbauer G., Gröller M., Raidou R.* ManyLands: A journey across 4D phase space of trajectories // Computer Graphics Forum. 2019. Vol. 38, No. 7. P. 191–202.

11. *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.

12. *Boge F.J.* Two dimensions of opacity and the deep learning predicament // Minds and Machines. 2022. Vol. 32. P. 43–75.

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

14. *Campolo A., Crawford K.* Enchanted determinism: power without responsibility in artificial intelligence // Engaging Science, Technology, and Society. 2020. Vol. 6. P. 1–19.

15. *Chen Y., Zhang D.* Integration of knowledge and data in machine learning. 2022. URL: https://arxiv.org/abs/2202.10337 (дата обращения: 15.06.2024).

16. *DeLanda M.* Intensive Science and Virtual Philosophy. Bloomsbury Academic; Reprint edition, 2013.

17. *Diacu F.* The solution of the n-body problem // Mathematical Intelligencer. 1996. Vol. 18, No. 3. P. 66–70.

18. *Duede E.* Deep learning opacity in scientific discovery // Philosophy of Science. 2022. Vol. 90. P. 1089–1099.

19. *Duran J.M.* Machine learning, justification, and computational reliabilism. 2023. URL: https://philpapers.org/archive/DURMLJ.pdf (дата обращения: 15.06.2024).

20. *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.

21. *Erichson B., Muehlebach M., Mahoney M.* Physics-informed autoencoders for Lyapunov-stable fluid flow prediction. 2019. URL: https://www.researchgate.net/publication/333418331_Physics-informed_Autoencoders_for_Lyapunov-stable_Fluid_Flow_Prediction (дата обращения: 15.06.2024).

22. *Fleisher W.* Understanding, idealization, and explainable AI // Episteme. 2022. Vol. 19. Spec. iss. 4: Jamaica conference special issue. P. 534–560.

23. *Goyal A., Bengio Y.* Inductive biases for deep learning of higher-level cognition // Proceedings of the Royal Society A. 2022. Vol. 478, No. 2266.

24. *Hellström T., Dignum V., Bensch S.* Bias in machine learning – What is it good for? // CEUR Workshop Proceedings. 2020. Vol. 2659. URL: http://ceur-ws.org/Vol-2659/ (дата обращения: 18.03.2025).

25. *Henderson L.* The problem of induction // The Stanford Encyclopedia of Philosophy. 2018. URL: https://plato.stanford.edu/entries/induction-problem/ (дата обращения: 15.06.2024).

26. *Hu P., Yang W., Zhu Y., Hong L.* Revealing hidden dynamics from time-series data by ODENet // Journal of Computational Physics. 2022. Vol. 461, No. 2. 111203.

27. *Humphreys P.* Computational science and its effects // Science in the Context of Application / Ed. by M. Carrier, A. Nordmann. Springer, 2011. P. 131–142.

28. *Humphreys P.* Extending Ourselves: Computational Science, Empiricism, and Scientific Method. N.Y.: Oxford University Press, 2004. 182 p.

29. *Karniadakis G.E., Kevrekidis I.G., Lu L., Perdikaris P., Wang S., Yang L.* Physics-informed machine learning // Nature Reviews Physics. 2021. Vol. 3, No. 6. P. 422–440.

30. *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.

2017. Vol. 29, No. 10. P. 2318–2331. URL: https://arxiv.org/pdf/1612.08544 (дата обращения: 15.06.2024).

31. *Kashefi A., Rempe D., Guibas L.J*. A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries // Physics of Fluids. 2021. Vol. 33, No. 2.

32. *Lauc D*. Machine learning and the philosophical problems of induction // Guide to Deep Learning Basics: Logical, Historical and Philosophical Perspectives. Springer, 2020. P. 93–106.

33. *Ling J., Kurzawski A., Templeton J*. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance // Journal of Fluid Mechanics. 2016. Vol. 807. P. 155–166.

34. *Lipton Z*. The Mythos of model interpretability // Communications of the ACM. 2018. Vol. 61, No. 10. P. 36–43.

35. *López-Rubio E., Ratti E*. Data science and molecular biology: prediction and mechanistic explanation // Synthese. 2021. Vol. 198, No. 4. P. 3131–3156.

36. *Mialon G*. On Inductive Biases for Machine Learning in Data Constrained Settings. Computer Science [cs]. Ecole Normale Supérieure (ENS), Paris, FRA., 2022. 113 p.

37. *Mitchell T*. The Need for Biases in Learning Generalizations. New Brunswick, NJ: Rutgers University, 1980.

38. *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.

39. *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.

40. *Pietsch W*. Aspects of Theory-Ladenness in Data-Intensive Science // Philosophy of Science. 2015. Vol. 82, No. 5. P. 905–916.

41. *Rahimi A., Recht B*. Reflections on Random Kitchen Sinks. 2017. URL: https://archives.argmin.net/2017/12/05/kitchen-sinks/ (дата обращения: 15.06.2024).

42. *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.

43. *Stan M*. From metaphysical principles to dynamical laws // The Cambridge History of Philosophy of the Scientific Revolution. 2020. URL: https://www.academia.edu/41468384/From_metaphysical_principles_to_dynamical_laws (дата обращения: 15.06.2024).

44. *Sterkenburg T., Grünwald P*. The no-free-lunch theorems of supervised learning // Synthese. 2021. Vol. 199. Iss.

45. *Stevens R., Taylor V., Nichols J., Maccabe A., Yelick K., Brown D*. AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science. Argonne IL: ANL, 2020.

46. *Wang Q*. The global solution of the n-body problem // Celestial Mechanics. 1991. Vol. 50. P. 73–88.

47. *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.

48. *Willard J., Jia X., Xu S., Steinbach M., Kumar V*. Integrating Physics-Based Modeling with Machine Learning: A Survey. 2020. URL: https://beiyulincs.github.io/teach/fall_2020/papers/xiaowei.pdf (дата обращения: 15.06.2024).

49. *Yazdani A., Lu L., Raissi M., Karniadakis G*. Systems biology informed deep learning for inferring parameters and hidden dynamics // PLoS Computational Biology. 2020. Vol. 16, No. 11. e1007575.

50. *Zednik C*. Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. 2019. URL: https://arxiv.org/pdf/1903.04361 (дата обращения: 15.06.2024).

References

1. *Domingos, P*. (2016). Verkhovnyy algoritm: kak mashinnoe obuchenie izmenit nash mir [The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World]. Transl. from English by V. Gorokhov, ed. by A. Sboev & A. Serenko. Moscow, Mann, Ivanov and Ferber Publ., 336. (In Russ.).

2. *Mainzer, K*. (2009). Slozhnosistemnoe myshlenie: Materiya, razum, chelovechestvo [Thinking in Complexity: The Computational Dynamics of Matter, Mind, and Mankind]. Transl. from English, ed. and with a foreword by G.G. Malinetsky. Moscow, Librokom Publ., 464. (In Russ.).

3. *Mukhin, R.R*. (2018). Ocherki po istorii dinamicheskogo khaosa: Issledovaniya v SSSR v 1950–1980-e gody [Essays on the History of Dynamic Chaos: Research in the USSR in the 1950s–1980s]. Moscow, URSS Publ., 320.

4. *Poincaré, A*. (1947). O krivykh, opredelyaemykh differentsialnymi uravneniyami [On Curves Defined by Differential Equations]. Transl. from French and notes by A.A. Andronov. Moscow & Leningrad, State Publishing House of Technical and Theoretical Literature, 388. (In Russ.).

5. *Sedov, L.I*. (1970). Mekhanika sploshnoy sredy [Continuum Mechanics], Vol. 2. Moscow, Nauka Publ., 568.

6. *Sukhno, A.A. & V.V. Gulin*. (2024). Integriruemost differentsialnykh uravneniy i “konvertatsiya slozhnosti” [Integrability of differential equations and “complexity conversion”]. Filosofiya nauki [Philosophy of Science], 1 (100), 67–95.

7. *Sukhno, A.A. & V.V. Gulin*. (2023). Kontingentnost sushchego i differentsialnye uravneniya [Contingency of beings and differential equations]. Logos, Vol. 33, No. 4, 95–116.

8. *Chalmers, D*. (2013). Soznayushchiy um: v poiskakh fundamentalnoy teorii [The Conscious Mind: In Search of a Fundamental Theory]. Transl. from English by V.V. Vasilyev. Moscow, URSS Publ., Librokom Publ., 509. (In Russ.).

9. *Shumskiy, S.A*. (2019). Mashinnyy intellect: Ocherki po teorii mashinnogo obucheniya i iskusstvennogo intellekta [Machine Intelligence: Essays on the Theory of Machine Learning and Artificial Intelligence]. Moscow, RIOR Publ., 340.

10. *Amirkhanov, A., I. Kosiuk, P. Szmolyan, A. Amirkhanov, G. Mistelbauer, M. Gröller & R. Raidou*. (2019). ManyLands: A journey across 4D phase space of trajectories. Computer Graphics Forum, Vol. 38, No. 7, 191–202.

11. *Blakseth, 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.

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

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

14. *Campolo, A. & K. Crawford*. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. Engaging Science, Technology, and Society, 6, 1–19.

15. *Chen, Y. & D. Zhang*. (2022). Integration of knowledge and data in machine learning. Available at: https://arxiv.org/abs/2202.10337 (date of access: 15.06.2024).

16. *DeLanda, M*. (2013). Intensive Science and Virtual Philosophy. Bloomsbury Academic, Reprint edition.

17. *Diacu, F*. (1996). The solution of the n-body problem. Mathematical Intelligencer, Vol. 18, No. 3, 66–70.

18. *Duede, E*. (2022). Deep learning opacity in scientific discovery. Philosophy of Science, 90, 1089–1099.

19. *Duran, J.M*. (2023). Machine learning, justification, and computational reliabilism. Available at: https://philpapers.org/archive/DURMLJ.pdf (date of access: 15.06.2024).

20. *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.

21. *Erichson, B., M. Muehlebach & M. Mahoney*. (2019). Physics-informed autoencoders for Lyapunov-stable fluid flow prediction. Available at: https://www.researchgate.net/publication/333418331_Physics-informed_Autoencoders_for_Lyapunov-stable_Fluid_Flow_Prediction (date of access: 15.06.2024).

22. *Fleisher, W*. (2022). Understanding, idealization, and explainable AI. Episteme, Vol. 19, Spec. Iss. 4: Jamaica conference special issue, 534–560.

23. *Goyal, A. & Y. Bengio*. (2022). Inductive biases for deep learning of higher-level cognition. Proceedings of the Royal Society A, Vol. 478, No. 2266.

24. *Hellström, T., V. Dignum & S. Bensch*. (2020). Bias in machine learning what is it good for? // CEUR Workshop Proceedings. 2020. Vol. 2659. URL: http://ceur-ws.org/Vol-2659/ (date of access: 18.03.2025).

25. *Henderson, L*. (2018). The problem of induction. In: The Stanford Encyclopedia of Philosophy. Available at: https://plato.stanford.edu/entries/induction-problem/ (date of access: 15.06.2024).

26. *Hu, P., W. Yang, Y. Zhu & L. Hong*. (2022). Revealing hidden dynamics from time-series data by ODENet. Journal of Computational Physics, Vol. 461, No. 2. 111203.

27. *Humphreys, P*. (2011). Computational science and its effects. In: M. Carrier & A. Nordmann (Eds.). Science in the Context of Application. Springer, 131–142.

28. *Humphreys, P*. (2004). Extending Ourselves: Computational Science, Empiricism, and Scientific Method. New York, US, Oxford University Press, 182.

29. *Karniadakis, G.E., I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang & L. Yang*. (2021). Physics-informed machine learning. Nature Reviews Physics, Vol. 3, No. 6, 422–440.

30. *Karpatne, A., G. Atluri, J.H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova & V. Kumar*. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering, Vol. 29. No, 10, 2318–2331. Available at: https://arxiv.org/pdf/1612.08544 (date of access: 15.06.2024).

31. *Kashefi, A., D. Rempe & L.J. Guibas*. (2021). A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries. Physics of Fluids, Vol. 33, No. 2.

32. *Lauc, D*. (2020). Machine learning and the philosophical problems of induction. In: Guide to Deep Learning Basics: Logical, Historical and Philosophical Perspectives. Springer, 93–106.

33. *Ling, J., A. Kurzawski & J. Templeton*. (2016). Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. Journal of Fluid Mechanics, 807, 155–166.

34. *Lipton, Z*. (2018). The mythos of model interpretability. Communications of the ACM, Vol. 61, No. 10, 36–43.

35. *López-Rubio, E. & E. Ratti*. (2021). Data science and molecular biology: prediction and mechanistic explanation. Synthese, Vol. 198, No. 4, 3131–3156.

36. *Mialon, G*. (2022). On Inductive Biases for Machine Learning in Data Constrained Settings. Computer Science [cs]. Paris, Ecole Normale Supérieure (ENS), 113.

37. *Mitchell, T*. (1980). The Need for Biases in Learning Generalizations. New Brunswick, NJ, Rutgers University.

38. *Napoletani, D., M. Panza & D. Struppa*. (2021). The agnostic structure of data science methods. Lato Sensu Revue de la Société de philosophie des sciences, Vol. 8, No. 2, 44–57.

39. *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.

40. *Pietsch, W*. (2015). Aspects of theory-ladenness in data-intensive science. Philosophy of Science, Vol. 82, No. 5, 905–916.

41. *Rahimi, A. & B. Recht*. (2017). Reflections on Random Kitchen Sinks. Available at: https://archives.argmin.net/2017/12/05/kitchen-sinks/ (date of access: 15.06.2024).

42. *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.

43. *Stan, M*. (2020). 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: 15.06.2024).

44. *Sterkenburg, T. & P. Grünwald*. (2021). The no-free-lunch theorems of supervised learning. Synthese, Vol. 199, Iss. 3.

45. *Stevens, R., V. Taylor, J. Nichols, A. Maccabe, K. Yelick & D. Brown*. (2020). AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science. United States, N. p.

46. *Wang, Q*. (1991). The global solution of the n-body problem. Celestial Mechanics, 50, 73–88.

47. *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.

48. *Willard, J., X. Jia, S. Xu, M. Steinbach & V. Kumar*. (2020). Integrating Physics-Based Modeling with Machine Learning: A Survey. Available at: https://beiyulincs.github.io/teach/fall_2020/papers/xiaowei.pdf (date of access: 15.06.2024).

49. *Yazdani, A., L. Lu, M. Raissi & G. Karniadakis*. (2020). Systems biology informed deep learning for inferring parameters and hidden dynamics. PLoS Computational Biology, Vol. 16, No. 11. e1007575.

50. *Zednik, C*. (2019). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. Available at: https://arxiv.org/pdf/1903.04361 (date of access: 15.06.2024).

Информация об авторах

Сухно Алексей Андреевич – Москва.
volyakvlasti@mail.ru

Гулин Вячеслав Владимирович – Институт механики Московского государственного университета (19192, Москва, Мичуринский проспект, 1).
kornet104@gmail.com

Information about the authors

Sukhno, Alexey Andreevich – Moscow.
volyakvlasti@mail.ru

Gulin, Vyacheslav Vladimirovich – Institute of Mechanics of Moscow State University (1, Michurinskiy Ave., Moscow, 119192, Russia).
kornet104@gmail.com

Дата поступления 20.11.2024