

Стратегия цифровой экономики  
Digital economy strategy

УДК 167/168, 004.8

DOI: 10.55959/MSU2070-1381-107-2024-194-205

Коллизии методологии и эпистемологии в науке о данных

Петрунин Юрий Юрьевич

Доктор философских наук, профессор, SPIN РИНЦ: [2206-8155](https://orcid.org/2206-8155), ORCID: [0000-0003-4218-2255](https://orcid.org/0000-0003-4218-2255), [petrunn@spa.msu.ru](mailto:petrunn@spa.msu.ru)

Факультет государственного управления, МГУ имени М.В. Ломоносова, Москва, РФ.

Аннотация

Возникшая относительно недавно наука о данных (Data Science) заняла достойное место в структуре наук. Применение науки о данных показало выдающиеся возможности решения многих сложных задач в различных сферах деятельности. Основой ее успеха стала новая методология познания, включающая в себя концепции и методы больших данных (Big Data), искусственного интеллекта (Artificial Intelligence), междисциплинарного подхода (информатики, статистики, математики, социальных и гуманитарных наук). Новая научная парадигма Data Science кардинально трансформирует научную методологию и поэтому нуждается в обосновании. Для решения поставленной задачи используются наукометрический метод, методы case-study, сравнительный анализ, методологический и эпистемологический анализ. В статье рассматриваются случаи методологических и эпистемологических коллизий, препятствующих эффективности применения науки о данных, их причины и следствия. Конкретно анализируются примеры совершенствования поисковых систем в интернете, оптимизации управления научными исследованиями, работы автомобильных навигаторов в мегаполисах. В результате проведенного исследования выделяются две группы противоречий между методологией и эпистемологией науки о данных. Первая группа связана с субъективными причинами дилемм, вторая — с объективными. В первой группе преобладают методологические причины возникающих конфликтов, во второй — эпистемологические причины возникающих противоречий. На взгляд автора, объективные парадоксы являются более сложными. Они затрагивают глубокие вопросы философии науки. В любом случае выделенные противоречия ведут к снижению потенциала науки о данных, приводят к ошибочным решениям и ложным прогнозам, и они должны быть устранены.

Ключевые слова

Наука о данных, парадигма науки о данных, методология науки о данных, эпистемология науки о данных, отраслевая наука о данных, ранжирование данных, принятие решений.

Для цитирования

Петрунин Ю.Ю. Коллизии методологии и эпистемологии в науке о данных // Государственное управление. Электронный вестник. 2024. № 107. С. 194–205. DOI: 10.55959/MSU2070-1381-107-2024-194-205

Collisions of Methodology and Epistemology in Data Science

Yuriy Yu. Petrunin

DSc (Philosophy), Professor, ORCID: [0000-0003-4218-2255](https://orcid.org/0000-0003-4218-2255), [petrunin@spa.msu.ru](mailto:petrunin@spa.msu.ru)

School of Public Administration, Lomonosov Moscow State University, Moscow, Russian Federation.

Abstract

Data Science, which emerged relatively recently, has taken its rightful place in the structure of sciences. The application of data science has shown outstanding possibilities for solving many complex problems in various fields of activity. The basis of its success was a new methodology of cognition, including the concepts and methods of Big Data, Artificial Intelligence, an interdisciplinary approach (computer science, statistics, mathematics, social and humanitarian sciences). The new scientific paradigm of Data Science radically transforms scientific methodology and therefore needs to be substantiated. To solve the problem, the scientometric method, case-study methods, comparative analysis, methodological and epistemological analysis are used. The article considers cases of methodological and epistemological collisions that hinder the effectiveness of data science, their causes and consequences. Specifically, examples of improving search engines on the Internet, optimizing the management of scientific research, and the operation of car navigators in megacities are analyzed. As a result of the conducted research, two groups of contradictions between the methodology and epistemology of data science are distinguished. The first group is associated with subjective causes of dilemmas, the second — with objective ones. In the first group, methodological reasons for the emerging conflicts prevail, while in the second group — epistemological reasons for the emerging contradictions. In the author's opinion, objective paradoxes are more complex. They touch upon deep questions of the philosophy of science. In any case, the identified contradictions lead to a decrease in the potential of data science, lead to erroneous decisions and erroneous forecasts, and they must be eliminated.

Keywords

Data science, data science methodology, data science epistemology, industry data science, data ranking, decision making.

For citation

Petrunin Yu.Yu. (2024) Collisions of Methodology and Epistemology in Data Science. *Gosudarstvennoye upravleniye. Elektronnyy vestnik*. No. 107. P. 194–205. DOI: 10.55959/MSU2070-1381-107-2024-194-205

Дата поступления/Received: 18.10.2024

## Введение

Термин Data Science (наука о данных; наука, основанная на данных/выводимая из данных) был предложен в 1974 году датским ученым в области компьютерных наук лауреатом премии Алана Тьюринга П. Науром [Naur 1974]. (Первоначально Наур называл новую науку datalogy [Naur 1966].) Речь шла о соединении математики, информатики, статистики и новых алгоритмов анализа эмпирических данных с целью получения глубоких знаний. Как точно высказался один из главных редакторов международного журнала Data Science Journal М. Майерник, «наука о данных возникла как набор методов, инструментов и навыков, которые можно применять для извлечения экономической или интеллектуальной ценности из растущих массивов данных» [Mayernik 2023].

Начиная со второго десятилетия XXI столетия интерес к науке о данных стремительно растет как в академическом сообществе, так и в обществе в целом (Рисунок 1 и Рисунок 2). Практика науки о данных в социально-экономической сфере показала отличные результаты в таких направлениях, как ритейл, медицина, финансовая и банковская сфера, строительство, транспорт, коммуникации, медиа и развлечения, образование, производство и природные ресурсы, управление, энергетика и коммунальные услуги, индустрия аутсорсинга<sup>1</sup>.

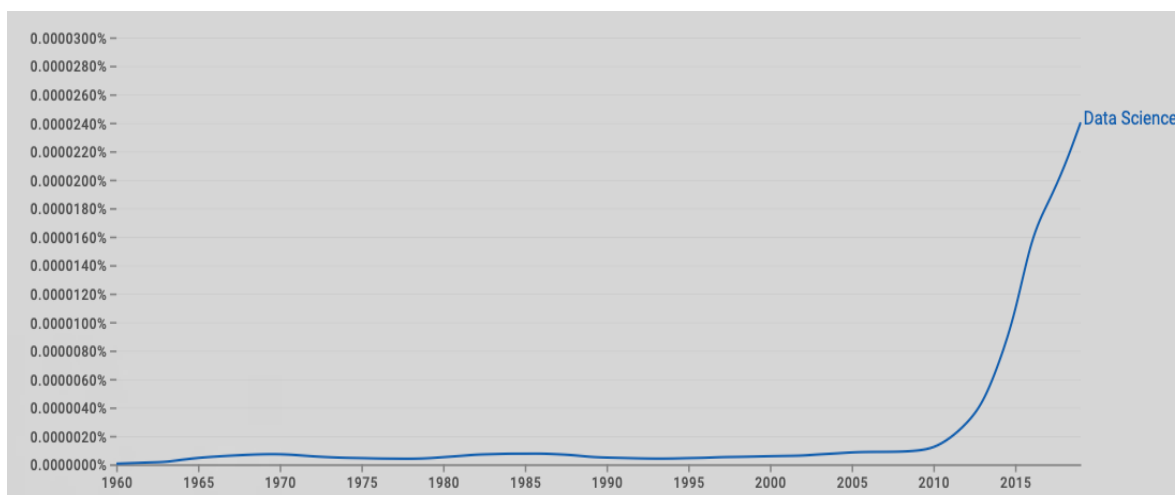


Рисунок 1. Динамика использования словосочетания Data Science в Google Ngram Viewer по годам в процентах от всех слов<sup>2</sup>

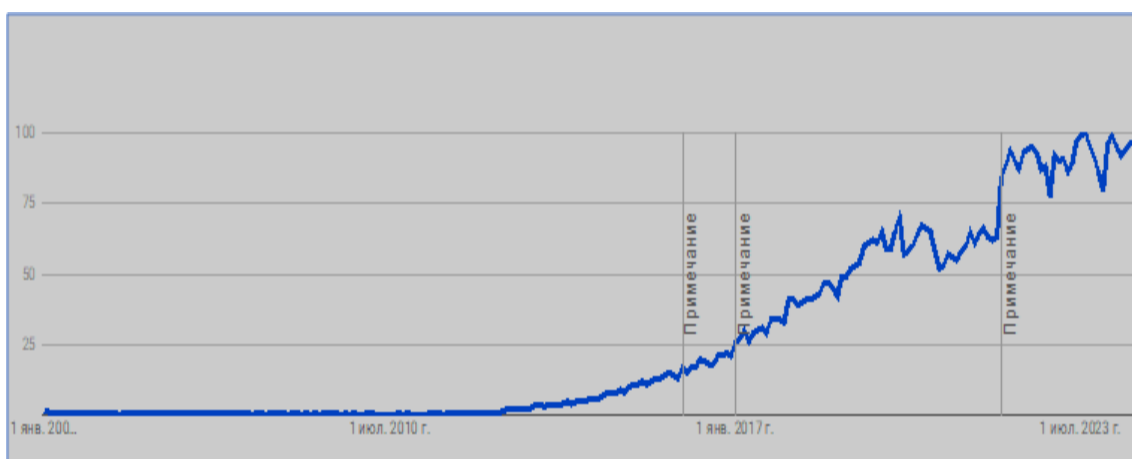


Рисунок 2. Популярность запроса Data Science за последние 20 лет по месяцам<sup>3</sup>

<sup>1</sup> 11 Industries That Benefits the Most from Data Science // Geeks for Geeks [Электронный ресурс]. URL: <https://www.geeksforgeeks.org/11-industries-that-benefits-the-most-from-data-science/> (дата обращения: 01.09.2024).

<sup>2</sup> График построен по запросу автора 14.07.2024.

<sup>3</sup> Построено Google Trends по запросу автора 14.07.2024.

Прогрессу науки о данных способствовало появление таких новых концепций сбора и обработки данных, как большие данные (Big Data, BD) и интернет вещей (Internet of Things, IoT), а также всеобщая цифровизация и прорывные алгоритмы искусственного интеллекта (Artificial Intelligence, AI): искусственные нейронные сети (Artificial Neural Networks, ANN), машинное обучение (Machine Learning, ML), глубинное обучение (Deep Learning, DL), большие лингвистические модели (Large Language Models, LLM) и др.

Как ответ на современные глобальные вызовы модернизируются учебные программы вузов, появляются новые курсы, например «Введение в Data Science», «Цифровая экономика», «Информационная безопасность и цифровой суверенитет в государственном и муниципальном управлении» и т. д. [Петрунин и др. 2024].

Важнейшей частью науки о данных стала новая методология познания действительности. Это неизбежно следует из междисциплинарного происхождения науки о данных и открытия ранее неизвестных инструментов изучения социальной и природной реальности. Наука о данных представляет собой не только эффективную прикладную науку, но и фундаментальную науку, активно осмысливающую свое место среди других академических дисциплин. За последние годы было опубликовано множество статей, посвященных методологическим [Петрунин 2016; Lebedev 2019], эпистемологическим [Desai et al. 2022; Pietsch 2022; Campagnolo 2021; Symons, Alvarado 2022; Lowrie 2017] и онтологическим [McQuillan 2018] аспектам и проблемам науки о данных. В научных публикациях и дискуссиях были отмечены не только позитивные моменты развития новой науки, но и сделаны критические замечания, особенно специалистами из области социально-гуманитарных наук, в отношении методологического и эпистемологического фундамента данной отрасли знаний, его определенной противоречивости и недостаточной обоснованности [Китчин 2017; Кочедыков 2023]. Таким образом, задача «укрепления» данного основания, его уточнения и выделения перспектив развития остается весьма актуальной.

В данной статье речь пойдет о некоторых возникающих коллизиях при использовании науки о данных в социальной сфере, связанных с нестыковкой методологических и эпистемологических подходов, снижающих полезность и эффективность новой науки.

### **Парадигма науки о данных**

Творцы новой академической дисциплины выводят парадигму науки о данных из эволюции научного познания в целом [The Fourth Paradigm 2009; Zhang 2023]. Первый исторический этап человеческого познания определяется ими как эмпирический, связанный с описанием феноменов/явлений реальности (в том числе и математическое описание, например, в астрономии). Этот этап начался в античности и продлился до эпохи Ренессанса.

Сущность второго этапа выражается в теоретическом подходе и связанном с ним построении объяснительных концепций изучаемых объектов. Если первая парадигма может быть охарактеризована как в целом индуктивная, то сменившая ее в результате научной революции XVII века парадигма основывается на гипотетико-дедуктивной методологии. Этот период длился до появления вычислительной техники в XX веке. Вычислительная наука становится парадигмой третьего этапа. Компьютеры позволяют строить модели сложных явлений действительности (в том числе и социальных, цифровых двойников реальных объектов или процессов), верифицировать параметры модели, интерпретировать результаты моделирования и прогнозировать течение/поведение изучаемых процессов.

С возникновением концепции больших данных начинается четвертый этап эволюции научного познания. Происходят важные методологические и эпистемологические изменения.

Поскольку в очень больших объемах данных стирается различие между выборкой и генеральной совокупностью, то отпадает необходимость в априорных теориях/гипотезах. «Применение методов, свободных от исходных предположений, позволяет данным самим “говорить” о себе, не завися от предубеждений и моделей мышления человека» [Китчин 2017, 122]. «Ученым больше нет необходимости делать обоснованные предположения, формулировать гипотезы и конструировать модели, а затем проверять их на основе экспериментальных данных или примеров» [Prensky 2010].

Новая парадигма позволяет исключить человеческий элемент из анализа данных, а вместе с ним и сопутствующую необъективность. Система не ждет, пока ей зададут вопрос или нацелят ее на анализ конкретных взаимосвязей в данных, а вместо этого самостоятельно выявит закономерности. Более того, «добыча данных (data mining) обнаруживает такие закономерности и взаимосвязи, которые мы и не подумали бы искать»<sup>4</sup>. Используемые методы искусственного интеллекта уже подтверждают такой поворот событий в математике, физике, экономике.

Аналитические методы, используемые в новой парадигме, — искусственные нейронные сети, машинное обучение и др. — в значительной степени отходят от классической рациональности научного познания: прежде всего воспроизводимости (полной или частичной) полученных результатов и прозрачности выводов, что не исключает получение истинных утверждений [Петрунин 2024].

Последнюю, четвертую парадигму принято называть «парадигмой, исследующей данные» (data exploration (eScience)) [The Fourth Paradigm 2009]: от датчиков с адронных коллайдеров или датчиков расходов воды в квартире, от активности социальных сетей или навигаторов в айфоне до обычных, но с каждым годом увеличивающихся объемов статистических данных. Фактически для вышедшей на первый план научной дисциплины весь мир — данные. Американский философ, логик и математик Уиллард Ван Орман Куайн писал, что «существовать — значит быть значением переменной» [Quine 1948]. Неплохо подходит для слогана науки о данных.

### **Новые задачи науки о данных**

Взрывное развитие науки о данных уверенно решает не только поставленные ранее задачи. Некоторые традиционные задачи, которые считались второстепенными для науки, вышли на первый план. Например, в классической науке главная цель — разработка каузальной теории/модели для объяснения закономерностей окружающего мира. Из верифицированной теории можно сделать прогноз динамических процессов этого мира. В науке о данных статус прогнозирования существенно изменяется: во-первых, существуют корреляционные зависимости между переменными, которые не являются каузальными. Во-вторых, открытие некаузальных связей между переменными не означает ненужности полученных результатов. Обнаружение связей между переменными само по себе является фундаментом принятия решений. Например, обнаруженная связь между переменными «доверие/недоверие к начальству» и «вероятность/прогнозирование в ближайшие несколько лет увольнения сотрудника» вполне полезна для управления персоналом, хотя не обоснована никакой теорией [Петрунин, Силюнова 2018]. Повысить зарплату, готовить смену, искать нового сотрудника — выбор из палитры действий руководства организации зависит от анализа данных. Экономические или демографические тренды Росстата также построены на данных временных рядов, которые используют математические модели и, как у астрономов античности вычисление движение планет, основаны на длительных наблюдениях, а не на открытии причинных моделей изучаемых процессов. Эти прогнозы важны не только для предвидения будущего, но и для оценки принимаемых государственных решений/программ. Таким образом, прогнозирование в науке о данных оказывается необязательно связанным с каузальными моделями.

<sup>4</sup> Big data 'Eureka!' don't just happen // Harvard Business Review [Электронный ресурс]. URL: [http://blogs.hbr.org/cs/2012/11/eureka\\_doesnt\\_just\\_happen.html](http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_happen.html) (дата обращения 01.09.2024).

Но кроме известных ранее появились и задачи, которые прежде вообще не считались научными. В первую очередь это относится к ранжированию, или рейтингованию. Ранжирование есть оценивание совокупности объектов/субъектов на основе данных и их упорядочивание на основе этих оценок. Ранжирование лучших книг или фильмов за последние сто лет, игроков в бейсболе или футболе, стран с самым высоким индексом счастья или индексом депрессии кажется лишь популярным приложением к серьезной науке. Однако, например, для выбора того, какому студенту повысить стипендию или как организовать оптимизацию поиска в интернете, нужно оценить (ранжировать) учащихся на основе их академических оценок, количества выступлений на научных конференциях, спортивных достижений, а сайты — по релевантности к запросу.

Задача ранжирования/рейтингования является основой для другой новой задачи, особенно в социально-гуманитарных науках, — задачи принятия решений. Рассматривая возникшие отраслевые разделы науки о данных, такие как доказательная медицина, доказательная политика, экономика данных, можно увидеть, что ранжирование стало фундаментом принятия управленческих решений в здравоохранении, в менеджменте, финансовой стратегии, в государственной политике. Рейтингование/ранжирование становится важнейшим механизмом оптимизации таких решений.

Сложность задачи ранжирования связана с неоднозначностью методик оценивания, их многомерностью и динамичностью из-за постоянно меняющихся данных<sup>5</sup>. В связи с этим актуальными проблемами становятся вопросы методологии и эпистемологии науки о данных. Попробуем уточнить эти важные понятия, их различие и взаимосвязь.

Методология разрабатывает и применяет инструменты/методы исследования для познания мира или его отдельных фрагментов/процессов. Система таких методов образует методологические концепции. Самыми универсальными в истории науки стали гипотетико-дедуктивная и индуктивная методологии. Наука о данных считается неоиндуктивизмом, поскольку базируется на эмпирических данных.

Эпистемология занимается вопросами истинности добытого знания. Она существенно связана с методологическими аспектами любой науки. Например, статистические органы меняют методы расчета определенных показателей, чтобы более точно отразить объект познания. Это методология. Однако, что именно отражают эти статистические показатели, насколько истинно получаемое используемой методологией знание, является вопросом эпистемологии. Например, насколько разные статистики (бедности, уровня образованности, экономического неравенства) отражают реальную бедность, образованность, экономическое неравенство населения той или иной страны. На таких агрегированных данных строятся графики динамики показателей, которые позволяют определить правильность принимаемых решений.

Чем, например, измерять уровень бедности страны? Средним уровнем дохода граждан страны? Однако во многих странах со средним высоким доходом граждан соседствует высокий уровень бедности. Почему это происходит? Известно, что средняя арифметическая адекватно отражает выборку только при условии нормального (гауссовского) распределения значений переменной. Получается, что при значительном разбросе доходов граждан средняя арифметическая не может быть объективным показателем типичного дохода. Гораздо лучше для этого подходит медиана или иногда мода. Поэтому базы статистических данных перешли от средней арифметической доходов к медиане и/или моде доходов, которые более точно и объективно характеризуют доходы населения и, соответственно, долю бедных в государстве. Данную задачу можно решить и другими способами, изменив методологию расчета показателей.

<sup>5</sup> Например, самый яркий пример перманентного пересмотра показателя ВВП изложен в [Астафьева, Турунцева 2021].

### **Методология и эпистемология данных**

Рассмотрим несколько примеров методологически-эпистемологических коллизий, возникающих в задаче ранжирования, и, как следствие, необоснованности и неэффективности принятия решений в разных отраслях науки о данных:

- 1) поисковых системах;
- 2) управлении наукой;
- 3) использовании систем навигации.

**Поисковые системы в интернете.** Обратим внимание на Рисунок 1 и Рисунок 2 в начале статьи. Если первый график более-менее точно отражает объективные процессы в языке, то на втором графике четко видны три перпендикулярных отрезка с названием «Примечание» и указанными датами по оси X. Они означают изменения методики расчета запросов Google Trends по термину Data Science. Остановимся на последнем перерасчете. В 2023 году произошли очередные обновления в поисковиках и сервисах «Яндекса» и Google. Полная картина апдейта поисковых алгоритмов неизвестна широкой публике. Компании делают объявления лишь о существенных изменившихся принципах и уточненных приоритетах. Это делается потому, что, зная методологию расчета ранжирования, можно использовать ее для достижения более высокого рейтинга сайта, что фактически приводит к искажению объективности данных, манипуляции ими. Это очень важный момент/аспект задачи ранжирования/рейтингования: знание методологии ранжирования может ухудшать ее эпистемологическую значимость, то есть полезность для принятия решений.

Разумеется, дополнительная информация все-таки раскрывается пользователями и становится известной миру. Простой пример — алгоритм ранжирования компании «Яндекс»<sup>6</sup>. С некоторыми инновациями большинство может согласиться как с разумными и важными: например, повышение рейтинга сайта, если после его посещения пользователь не переходит на другие адреса, — это означает, что он нашел ответ на интересующий его вопрос. Или другие примеры: повторное посещение сайта поднимает позиции последнего; добавление адреса сайта в закладки поднимает позиции сайта, уменьшает позиции в рейтинге малополезного контента (в котором присутствуют заимствованные материалы и компиляции) и т. д. Обратим внимание: использование чужого контекста не только приводит учащихся к неудовлетворительным оценкам, граждан к штрафам за плагиат, но и всех пользователей интернета к напрасной потере времени, а менеджеров к ошибочному принятию решений, основанных на «мусорных» данных.

Но есть в стратегии поисковых алгоритмов «Яндекса» и сомнительные инновации. Например, ссылки на «Википедию» повышают рейтинг. «Википедия», конечно, имеет свои достоинства: оперативные обновления, полнота описания, широкий кругозор, качественные иллюстрации и др. Но, с другой стороны, идеологическая ангажированность, грубые ошибки во многих актуальных темах, анонимность авторов, ведущая к безответственности, известны многим профессионалам, критикующим раскрытие конкретных тем за их ненаучность и неквалифицированность.

Что касается Google, то там тоже много похожего. Приведем более животрепещущие примеры, на которых Google сфокусировался. Речь идет о генеративном искусственном интеллекте (ГИИ). Как относится к сгенерированным текстам? Позиция компании исходит из того, что «системы ранжирования Google необходимы для того, чтобы преимущество в результатах поиска получал уникальный контент высокого качества, который соответствует стандартам E-E-A-T (опыт, компетентность, авторитетность и достоверность) ... Качество контента для нас гораздо

<sup>6</sup> Утечка кода Яндекса: стали известны факторы ранжирования // Promo Pult [Электронный ресурс]. URL: <https://blog.promopult.ru/seo/utechka-koda-yandeksa-stali-izvestny-factory-ranzhirovaniya.html> (дата обращения 01.09.2024).

важнее того, каким способом он создан»<sup>7</sup>. Из позиции компании следует, что ГИИ может открыто использоваться, если соответствует стандартам, указанным выше. Если сгенерированный текст не соответствует декларируемым требованиям, то это нужно рассматривать как спам.

Обратим внимание также на то, что AI-компании планируют выдавать результаты поиска на запрос пользователя не в виде списка адресов интернета, а в готовом текстовом или графическом виде. В настоящее время так работает ГИИ. При этом очевидно, что готовый «ответ» не имеет ссылок на первичные источники вообще либо указывает расплывчатые ссылки: например, на фамилию автора, название книги или статьи, научной школы и т. п. Мы уже убедились в том, что ChatGPT «не дружит» со ссылками. Но ссылки на источники, особенно в социально-гуманитарных науках, — фундамент любого исследования. Элиминация ссылок или хотя бы умаление их значимости приводят к невозпроизводимости и нефальсифицируемости результатов, что является отказом от классической эпистемической рациональности. Правда, OpenAI обещает уделить должное внимание ссылкам на источники информации в [SearchGPT Prototype](#). Посмотрим, сбудутся ли эти обещания.

Непрозрачность ранжирования материалов интернета ИИ-поисковиком также вызывает сомнения у ведущих экспертов. «Ладно, если сервис будет что-то не то рекомендовать, но главное — не сможет находить нужные страницы или статьи... В таком случае некорректные материалы поднимаются вверх»<sup>8</sup>.

**Управление наукой.** Государственное (и не только государственное) управление научными исследованиями нуждается в правильной оценке отдельных ученых и научных организаций (коллективов), чтобы различными способами (через гранты, премии, наградную политику и т. п.) мотивировать наиболее важные и перспективные направления науки. В частности, для объективного оценивания уровня исследователя, коллектива, журнала используются системы (платформы) научного цитирования, которые агрегируют различные статистики публикационной активности, формируют на основе этой статистики рейтинги журналов, академических институтов, отдельных ученых. К таким базам относятся WoS, Scopus, РИНЦ и др. Остановимся подробнее на отечественной платформе РИНЦ и рейтинговании научных журналов.

В течение последних 15 лет методики для рейтинга менялись: появлялись новые наукометрические показатели/статистики (как количественные, так и качественные: импакт-фактор, индекс общественного признания и др.) и методологии расчета интегрального показателя Science Index. На основе этого формировалась группа ведущих высокорейтинговых журналов (так называемое ядро РИНЦ<sup>9</sup>). В марте 2023 года была проведена кардинальная реформа показателя Science Index, существенно увеличившая влияние журналов из ядра РИНЦ. Конечно, совершенствование статистик и методологии расчета интегрального индекса необходимы для более объективной картины научного ландшафта и исключения влияния вненаучных факторов на рейтингование. Однако введение инноваций в 2023 году привело к тому, что некоторые журналы опустились или поднялись в рейтинге на несколько порядков. Конкретный пример: в августе 2022 года Российский индекс научного цитирования РИНЦ обновил информацию за предыдущий год, и журнал «Государственное управление. Электронный вестник» занял 24-е место в общем рейтинге SCIENCE INDEX и 2-е место в рейтинге SCIENCE INDEX по тематике «Организация и управление». В марте 2023 года (без нового обновления статистик РИНЦ) он оказался на 1310-м месте.

<sup>7</sup> Правила Google Поиска в отношении контента, созданного искусственным интеллектом // Центр Google [Электронный ресурс]. URL: <https://developers.google.com/search/blog/2023/02/google-search-and-ai-content?hl=ru> (дата обращения 01.09.2024).

<sup>8</sup> OpenAI займется поиском. Каковы ожидания от SearchGPT // Коммерсантъ [Электронный ресурс]. URL: <https://www.kommersant.ru/doc/6862192?ysclid=m0ait8rkp1699417636> (дата обращения 01.09.2024).

<sup>9</sup> Формировались и другие группы ведущих журналов.

Похожая ситуация произошла с журналом «Вопросы философии». Журнал считался очень авторитетным и в международном рейтинге входил в категорию Q2. Последние 10–15 лет этот журнал занимал высокие позиции и в отечественном рейтинге РИНЦ: от первого места до десятого. С приснопамятного марта 2023 года он сместился на 124 место (на август 2024 г.), то есть на порядок ниже. Объяснения для этого могут быть только два: либо старые методы оказались *чрезвычайно* неадекватными (то есть были сформированы непрофессионалами), либо новые методы были пролоббированы вошедшими в ядро РИНЦ журналами<sup>10</sup>. Иначе объяснить мегаизменения рейтингов многих журналов невозможно.

Но инновация не только привела к новым методам расчета рейтинга журналов, но и пересчитала все предшествующие показатели журналов за прошедшие 10 лет, то есть переписала всю архивную историю прошлого. Можно ли переписывать историю? Можно, если новое описание ближе к истине. Например, современные методы анализа данных могут быть более точными в оценке потери населения США во время Великой депрессии с 1929 по 1933 гг. Но репутация научного журнала или отдельной статьи является изначально субъективной, контекстуально и культурно-исторически зависимой, а ее радикальный пересмотр является искажением прошлого («переписывание истории»).

В эпистемологии эта проблема обозначается как дилемма антикваризма и презентизма. Презентизм представляет собой методологию, описывающую прошлые события современным терминами и представлениями/знаниями. Антикваризм ориентирован на понимание прошлого в терминах и представлениях/знаниях изучаемого периода времени. Классическая иллюстрация этой дилеммы — вопрос «Открыл ли Колумб Америку?». С современной точки зрения правильный ответ — да. С точки зрения XV века правильный ответ — нет, потому что командор и его команда не знали никакой Америки, не стремились открывать Новый Свет. Их задача была достигнуть Индии, опираясь на представления о шарообразности Земли. Именно это было главной целью путешествия Колумба<sup>11</sup>.

Возвращаясь к рейтингованию научных журналов, повторим, что научная репутация в конкретный промежуток времени определяется (относительно) независимыми общественными институтами, в данном случае РИНЦ. Пересмотр прошлого новыми инструментами и методологиями не соответствует реальной истории, а скорее является подгонкой временных рядов, скрывающих странные пертурбации динамики авторитета научных публикаций.

В области управления наукой и образованием есть и другие подобные примеры. Так, при прохождении преподавателем конкурса или вхождении в члены диссертационного совета ВАК также в 2023 г. были введены повышенные требования по количеству публикаций — за 5 лет 8 статей, но вводились они таким образом, что это условие начинало действовать с момента публикации новых требований. То есть введенные правила распространились на прошлое.

Распространение новой методологии расчета статистических показателей на прошлое — типичный пример *когерентной теории истины*, идеалом которой является согласованность/непротиворечивость всех утверждений о некотором фрагменте реальности.

Рассмотренный пример можно обозначить как *проблему эпистемического антикваризма-презентизма в науке о данных*.

<sup>10</sup> Эффект Матфея — «богатые становятся еще богаче, бедные — еще беднее».

<sup>11</sup> Впрочем, есть и другие версии цели Колумба. «Его идеи о форме Земли, — писал академик В.И. Вернадский, — были очень странными. Не отличаясь, подобно другим великим мореплавателям того времени, достаточным астрономическим и математическим образованием и не будучи в состоянии ориентироваться в громоздком и неудобном аппарате того времени, Колумб думал сделать из своих наблюдений вывод о том, что Земля не имеет форму шара, а форму груши, и на узком ее конце находится возвышение, которое Колумб считал местом входа в рай. Он развивал, следовательно, теорию, которую проповедовали многие церковные писатели того времени... — о литосфере, плавающей в гидросфере, с несовпадающими центрами, т. е. придерживался того воззрения, которое всецело разрушалось его великим открытием» [Вернадский 1981, 160].



Перенос новых методов обработки первичных данных на предшествующую статистику могут приводить к серьезным искажениям динамики социальных процессов и неверной оценке прошлого. Одним словом, данные действия относятся к эпистемологии.

Особенно это касается рефлексивных систем, или систем с рефлексией. Системы с рефлексией относятся к саморегулируемым (самоорганизованным). В естественных науках данные достаточно четко отделены от методологии их познания. Астроном, наблюдающий за движением звезд, геолог, собирающий горный материал, биолог, коллекционирующий экземпляры растений или живых существ, не аффилированы с предметом исследования. Любое рейтингование в социальных системах неизбежно связано с часто даже не осознаваемой предвзятостью анализа отобранного эмпирического материала (данных) или в худшем случае манипулированием данными в своих интересах.

**Системы автомобильной навигации в мегаполисе.** Другая методологически-эпистемологическая коллизия в науке о данных проявляется при управлении мегаполисом на основе концепции «умный город» и связана с оперативной обработкой данных и ее целью. Яркой иллюстрацией этой дилеммы является работа автомобильного навигатора. Водитель едет на машине, навигатор указывает ему оптимальный маршрут по времени на основе собранной информации<sup>12</sup>. Программа показывает водителю несколько вариантов решений, указывая вероятное время достижения цели движения в минутах (типичная задача ранжирования и прогнозирования). Причем современные системы навигации постоянно отслеживают ситуацию на улицах города и могут в любой момент предложить водителю выбранный ранее маршрут заменить на лучший. Однако часто бывают случаи, когда водитель согласен с выбором нового маршрута и готов сменить направление движения, но в последний момент перед изменением маршрута вдруг на мониторе появляется надпись «то же время» (как и по ранее выбранной траектории движения). Причем только что навигатор уверял, что изменение маршрута укоротит его на 10–15 минут. Как это возможно?

Очевидно, можно выделить несколько причин парадокса.

Первое предположение (гипотеза) состоит в том, что вычисление стохастических транспортных потоков миллиона машин (плюс светофоры, ДТП и др. обстоятельства) представляет собой очень сложную и не очень предсказуемую задачу. Однако мгновенные изменения в предлагаемых оптимальных путях (различающихся более чем на 15–20% времени в дороге) — весьма маловероятные события.

Второе предположение (гипотеза) состоит в том, что подсказки навигатора служат не обнаружению лучшего маршрута для конкретного водителя, а служат оптимизации транспортных потоков в целом. Но что хорошо для всех, не обязательно хорошо для конкретного водителя. Он честно оплачивает работу навигатора с целью получить оптимальный маршрут для себя, а не для города, а городское управление транспортными потоками осуществляется за его счет. Данная гипотеза подтверждается личным опытом автора статьи. Когда мы едем на двух машинах с супругой на дачу, то часто бывают случаи, когда навигатор одной и той же компании «Яндекс» показывает нам разные маршруты. При этом расстояние между нашими машинами колеблется от 50 до 100 метров. Обращаю внимание, что высчитанная компьютером разница указывает уменьшение времени достижения цели на 15–20 минут.

Другой, неальтернативной гипотезой может быть следующее объяснение. Почему в последний момент перед изменением маршрута транспортного средства выводится сообщение «то же время»? Простой ответ, основанный на водительском опыте, говорит о том, что когда большинство участников автомобильного движения ориентируется на навигаторы, то они свернут туда, где маршрут короче. Но тогда будет разгружаться первоначальный маршрут, и часто по нему можно

<sup>12</sup> В мире и России существуют несколько баз оперативных данных о загруженности дорог.

добрать до цели быстрее. Парадокс оптимизации транспортных потоков индивидуальных субъектов/ агентов (даже если их решения основываются на оперативных данных о загрузке дорог) состоит в том, что их знание о наименее загруженных дорогах влияет на изменение этой загруженности. Получается, что знание о еще не наступившем будущем влияет на это будущее. Можно назвать такую *проблему в науке о данных эпистемологически-футурологической*.

### **Заключение**

Теория и практика науки о данных имеет не только проблемы в обосновании, но и некоторые трудно разрешимые противоречия, особенно в социально-гуманитарных дисциплинах. Сформировавшиеся ключевые задачи науки о данных — ранжирование, обоснование принятия решений, прогнозирование — выявили конфликты между методологией и эпистемологией данной науки. При этом можно выделить две группы коллизий. Первая связана с субъективными причинами, вторая — с объективными. В качестве типичных примеров первого типа можно привести изменения ключевых показателей для принятия решений, детерминированных самими решениями соответствующих органов управления. Под эти решения модернизируются и задачи рейтингования/ ранжирования, и «ожидаемое» прогнозирование.

Во вторую группу можно отнести случаи, связанные с объективными дилеммами, такими как, например, влияние знания о будущем на реальные будущие события, что приводит к логическому парадоксу. Или случай выбора презентистского подхода к архивным данным, приводящего к сослагательному варианту истории (переписыванию истории) либо к антикваристскому подходу, ведущему к социальному мультиверсу (несоизмеримости прошлого и настоящего). Ясно, что оба результата противоречат рациональному научному методу и подрывают потенциальные возможности Data Science.

Как говорится в известном высказывании, есть три вида лжи: просто ложь, чудовищная ложь и статистика. Мы должны сделать все, чтобы наука о данных не превзошла свою «прабабушку» в отрицательном рейтинге.

### **Список литературы:**

- Астафьева Е.В., Турунцева М.Ю. Пересмотры ВВП: данные и оценка статистических свойств // Экономический журнал ВШЭ. 2021. Т. 25. № 1. С. 65–101. DOI: [10.17323/1813-8691-2021-25-1-65-101](https://doi.org/10.17323/1813-8691-2021-25-1-65-101)
- Вернадский В.И. Избранные труды по истории науки. М.: «Наука», 1981.
- Китчин Р. Большие данные, новые эпистемологии и смена парадигм // Социология: методология, методы, математическое моделирование. 2017. № 44. С. 111–152.
- Кочедыков И.Е. Об опыте применения больших данных в политической науке // Политическая наука. 2023. № 4. С. 226–251. DOI: [10.31249/poln/2023.04.09](https://doi.org/10.31249/poln/2023.04.09)
- Петрунин Ю.Ю. Искусственный интеллект и методологические вопросы управления знаниями // Философские науки. 2016. № 8. С. 67–74.
- Петрунин Ю.Ю. Искусственные нейронные сети в экономике: математический инструмент, модель или методология? // Вестник Московского университета. Серия 6. Экономика. 2024. № 4. С. 92–113. DOI: [10.55959/MSU0130-0105-6-59-4-5](https://doi.org/10.55959/MSU0130-0105-6-59-4-5)
- Петрунин Ю.Ю., Агаян Г.М., Бухарин В.В., Григорян А.А., Шевцова И.В., Шикина Г.Е. Интеграция математических методов и цифровых технологий как основа создания комплекса фундаментальных курсов в подготовке современных управленческих кадров // Вестник Московского университета. Серия 21. Управление (государство и общество). 2024. Т. 21. № 1. С. 139–167. DOI: [10.55959/MSU2073-2643-21-2024-1-139-167](https://doi.org/10.55959/MSU2073-2643-21-2024-1-139-167)

Петрунин Ю.Ю., Силюянова Ю.А. Статистические и нейросетевые методы в исследовании управленческих проблем в организации // Нейрокомпьютеры: разработка, применение. 2018. № 10. С. 39–47.

Campagnolo G.M. Participative Epistemology in Social Data Science: Combining Ethnography with Computational and Statistical Approaches // *International Journal of Social Research Methodology*. 2021. Vol. 25. Is. 3. P. 391–403. DOI: [10.1080/13645579.2021.1892379](https://doi.org/10.1080/13645579.2021.1892379)

Desai J., Watson D., Wang V., Taddeo M., Floridi L. The Epistemological Foundations of Data Science: A Critical Review // *Synthese*. 2022. Vol. 200. DOI: [10.1007/s11229-022-03933-2](https://doi.org/10.1007/s11229-022-03933-2)

Lebedev S. Methodology of Neo-inductivism: Critical Analysis // *Proceedings of the 4th International Conference on Contemporary Education, Social Sciences and Humanities (ICCESSH 2019)*. 2019. DOI: [10.2991/iccessh-19.2019.2](https://doi.org/10.2991/iccessh-19.2019.2)

Lowrie I. Algorithmic Rationality: Epistemology and Efficiency in the Data Sciences // *Big Data & Society*. 2017. Vol. 4. Is. 1. DOI: [10.1177/2053951717700925](https://doi.org/10.1177/2053951717700925)

Mayernik M.S. Data Science as an Interdiscipline: Historical Parallels from Information Science // *Data Science Journal*. 2023. Vol. 22. DOI: [10.5334/dsj-2023-016](https://doi.org/10.5334/dsj-2023-016)

McQuillan D. Data Science as Machinic Neoplatonism // *Philosophy and Technology*. 2018. Vol. 31. P. 253–272. DOI: [10.1007/s13347-017-0273-3](https://doi.org/10.1007/s13347-017-0273-3)

Naur P. *Concise Survey of Computer Methods*. Lund: Studentlitteratur, 1974.

Naur P. The Science of Datalogy // *Communications of the ACM*. 1966. Vol. 9. Is. 7. P. 485. DOI: [10.1145/365719.366510](https://doi.org/10.1145/365719.366510)

Pietsch W. *On the Epistemology of Data Science. Conceptual Tools for a New Inductivism*. Cham: Springer, 2022. DOI: [10.1007/978-3-030-86442-2](https://doi.org/10.1007/978-3-030-86442-2)

Premsky M.H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom // *Italian Journal of Educational Technology*. 2010. Vol. 18. Is. 2. DOI: [10.17471/2499-4324/277](https://doi.org/10.17471/2499-4324/277)

Quine W. *On What There Is* // *Review of Metaphysics*. 1948. Vol. 2. Is 5. P. 21–38.

Symons J., Alvarado R. Epistemic Injustice and Data Science Technologies // *Syntheses*. 2022. Vol. 200. DOI: [10.1007/s11229-022-03631-z](https://doi.org/10.1007/s11229-022-03631-z)

*The Fourth Paradigm: Data-Intensive Scientific Discovery* / Ed. by T. Hey, S. Tansley, K. Tolle. Redmond: Microsoft Research, 2009.

Zhang L. Looking Back to the Future: A Glimpse at Twenty Years of Data Science // *Data Science Journal*. 2023. Vol. 22. Is. 7. DOI: [10.5334/dsj-2023-007](https://doi.org/10.5334/dsj-2023-007)

### References:

Astafieva E.V., Turuntseva M.Yu. (2021) Revisions of GDP: Data and Assessment of Statistical Properties. *Ekonomicheskij zhurnal VShE*. Vol. 25. No. 1. P. 65–101. DOI: [10.17323/1813-8691-2021-25-1-65-101](https://doi.org/10.17323/1813-8691-2021-25-1-65-101)

Campagnolo G.M. (2021) Participative Epistemology in Social Data Science: Combining Ethnography with Computational and Statistical Approaches. *International Journal of Social Research Methodology*. Vol. 25. Is. 3. P. 391–403. DOI: [10.1080/13645579.2021.1892379](https://doi.org/10.1080/13645579.2021.1892379)

Desai J., Watson D., Wang V., Taddeo M., Floridi L. (2022) The Epistemological Foundations of Data Science: A Critical Review. *Synthese*. Vol. 200. DOI: [10.1007/s11229-022-03933-2](https://doi.org/10.1007/s11229-022-03933-2)

Hey T., Tansley S., Tolle K. (eds.) (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond Microsoft Research.

Kitchin R. (2014) Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*. No. 44. P. 111–152.

- Kochedykov I.E. (2023) On Big Data Application for Political Science Research. *Politicheskaya nauka*. No. 4. P. 226–251. DOI: [10.31249/poln/2023.04.09](https://doi.org/10.31249/poln/2023.04.09)
- Lebedev S. (2019) Methodology of Neo-inductivism: Critical Analysis. *Proceedings of the 4th International Conference on Contemporary Education, Social Sciences and Humanities (ICCESSH 2019)*. DOI: [10.2991/iccessh-19.2019.2](https://doi.org/10.2991/iccessh-19.2019.2)
- Lowrie I. (2017) Algorithmic Rationality: Epistemology and Efficiency in the Data Sciences. *Big Data & Society*. Vol. 4. Is. 1. DOI: [10.1177/2053951717700925](https://doi.org/10.1177/2053951717700925)
- Mayernik M.S. (2023) Data Science as an Interdiscipline: Historical Parallels from Information Science. *Data Science Journal*. Vol. 22. DOI: [10.5334/dsj-2023-016](https://doi.org/10.5334/dsj-2023-016)
- McQuillan D. (2018) Data Science as Machinic Neoplatonism. *Philosophy and Technology*. Vol. 31. P. 253–272. DOI: [10.1007/s13347-017-0273-3](https://doi.org/10.1007/s13347-017-0273-3)
- Naur P. (1974) *Concise Survey of Computer Methods*. Lund: Studentlitteratur.
- Naur P. (1966) The Science of Datalogy. *Communications of the ACM*. Vol. 9. Is. 7. P. 485. DOI: [10.1145/365719.366510](https://doi.org/10.1145/365719.366510).
- Petrinin Yu.Yu. (2016) Artificial Intelligence and Knowledge Management Methodological Questions. *Filosofskiye nauki*, No. 8. P. 67–74.
- Petrinin Yu.Yu. (2024) Artificial Neural Networks in Economics: Mathematical Tool, Model or Methodology? *Vestnik Moskovskogo universiteta. Seriya 6. Ekonomika*. No. 4. C. 92–113. DOI: [10.55959/MSU0130-0105-6-59-4-5](https://doi.org/10.55959/MSU0130-0105-6-59-4-5)
- Petrinin Yu.Yu., Agayan G.M., Bukharin V.V., Grigoryan A.A., Shevtsova I.V., Shikina G.E. (2024) Integration of Mathematical Methods and Digital Technologies as the Basis for the Creation of a Set of Fundamental Courses in the Training of Modern Management Personnel. *Vestnik Moskovskogo universiteta. Seriya 21. Upravleniye (gosudarstvo i obshchestvo)*. Vol. 21. No. 1. P. 139–167. DOI: [10.55959/MSU2073-2643-21-2024-1-139-167](https://doi.org/10.55959/MSU2073-2643-21-2024-1-139-167)
- Petrinin Yu.Yu., Siluyanova Y.A. (2018) Statistical and Neural Network Methods for Analyzing Indicators of Healthy Administration. *Neyrokomp'yutery: razrabotka, primeneniye*. No. 10. P. 39–47.
- Pietsch W. (2022) On the Epistemology of Data Science. Conceptual Tools for a New Inductivism. Cham: Springer. DOI: [10.1007/978-3-030-86442-2](https://doi.org/10.1007/978-3-030-86442-2)
- Premsky M.H. (2010) Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom. *Italian Journal of Educational Technology*. Vol. 18. Is. 2. DOI: [10.17471/2499-4324/277](https://doi.org/10.17471/2499-4324/277)
- Quine W. (1948) On What There Is. *Review of Metaphysics*. Vol. 2. Is 5. P. 21–38.
- Symons J., Alvarado R. (2022) Epistemic Injustice and Data Science Technologies. *Syntheses*. Vol. 200. DOI: [10.1007/s11229-022-03631-z](https://doi.org/10.1007/s11229-022-03631-z)
- Vernadskiy V. I. (1981) *Izbrannyye trudy po istorii nauki* [Selected works on the history of science]. Moscow: "Nauka".
- Zhang L. (2023) Looking Back to the Future: A Glimpse at Twenty Years of Data Science. *Data Science Journal*. Vol. 22. Is. 7. DOI: [10.5334/dsj-2023-007](https://doi.org/10.5334/dsj-2023-007)