Андрей Г. Кузнецов

Европейский университет в Санкт-Петербурге; Университет ИТМО, Санкт-Петербург, Россия
ORCID: 0000-0002-0249-5890

Туманности нейросетей: «черные ящики» технологий и наглядные уроки непрозрачности алгоритмов

doi: 10.22394/2074-0492-2020-2-157-182

Резюме:
В статье утверждается, что: 1) современные технологии, моделирующие искусственный интеллект на основе знания нейронаук, служат наглядным эмпирическим подтверждением теоретических аргументов, высказанных в исследованиях наук и технологий в конце 1980-х — начале 1990-х годов; 2) актуальная дискуссия о непрозрачности алгоритмов позволяет сместить перспективу на классический, но амбивалентный для STS троп раскрытия «черных ящиков». Для демонстрации этого, во-первых, дается экспозиция проблемы нейтральности и прозрачности технологий. Анализируются три типичных ответа на проблему нейтральности технологий внутри и за пределами конструктивистских STS. Утверждается, что несмотря на поверхностные различия между ними, все три ответа одинаково концептуализируют технологию как нейтральный проводник. Во-вторых, показывается связь проблемы нейтральности и прозрачности технологий с ключевым для исследований наук и технологий (STS) тропом раскрытия «черных ящиков» технологий. Обсуждаются социально-политическое и методологическое измерения метафоры «черного ящика». В-третьих, альтернатива концептуализации технологии как нейтральной ищется в акторно-сетевой теории Бруно Латура. Здесь технологии понимаются как событийная ассоциация разнородных сущностей, нередуцируемая к условиям ее возможности. Конструирование технологий понимается как процесс медиации, где создания удивляют своих создателей и наоборот. В акторно-сетевой теории Латура технологии рассматриваются

Кузнецов Андрей Геннадиевич — кандидат социологических наук, научный сотрудник Центра исследований науки и технологий Европейского университета в Санкт-Петербурге; ординарный доцент Университета ИТМО. E-mail: akuznetsov@eu.spb.ru
Исследование выполнено при поддержке гранта Российского научного фонда (проект РНФ № 20-78-10106) «Беспилотные автомобили и общество: взаимодействие технологий, социоэкономических сценариев и регулирования в радикальной инновации».

как непрозрачные и не-нейтральные сущности. Наконец, приводятся примеры из области технологий на основе нейросетей и алгоритмов глубокого машинного обучения, которые являются наглядным и эмпирическим подтверждением латуровской концепции технической медиации. Отдельное внимание уделяется непрозрачности и (не)подотчетности алгоритмов машинного обучения.

Ключевые слова: акторно-сетевая теория, ценностная нейтральность технологий, «черные ящики», нейросети, беспилотные автомобили, исследования науки и технологий, машинное обучение

Andrei G. Kuznetsov

European University at Saint-Petersburg; ITMO University, Russia
ORCID: 0000-0002-0249-5890

Neural Network Nebulae: 'Black Boxes' of Technologies and Object-Lessons from the Opacities of Algorithms

Abstract:
The paper deals with the quandary of the neutrality and transparency of technologies. First, I show how this problem is connected with the image of the opening of 'black boxes' that is pivotal to much of science and technology studies. Second, methodological and socio-political dimensions of the 'black box' metaphor are discussed. Third, I analyze three typical solutions to the problem of the neutrality of technologies outside and inside constructivist technology studies. It is demonstrated that despite their apparent differences, these solutions are similar in their logic of conceptualizing technology as a neutral intermediary. Forth, I look for an alternative to this logic in the actor-network theory of Bruno Latour. Here technologies are conceived in terms of an eventful association of heterogeneous entities irreducible to its conditions of possibility. The construction of technologies is understood as mediation, or as a 'making-do' process where creators are surprised by their creations and vice versa. In Latour's actor-network, technologies are interpreted as opaque and non-neutral entities. Finally, I turn to some object-lessons from smart technologies powered by neural networks to demonstrate that these are empirical vindications of Latour's conception of technical mediation. Particular attention is paid to the opacity and (non)interpretability of machine learning algorithms.

Andrei G. Kuznetsov — PhD in Sociology, Research Fellow, STS-Centre, European University at Saint-Petersburg; Associate Professor, ITMO University. E-mail: akuznetsov@eu.spb.ru
The research is done with the support of the Russian Science Foundation grant 'Self-Driving Cars and Society: Interaction of Technologies, Socioeconomic Scenarios and Regulation Within The Radical Innovation' (project no. 20-78-10106).

Keywords: Actor-Network Theory, value neutrality of technologies, black boxes, neural networks, machine learning, self-driving cars, science & technology studies (STS)

Введение

Современные нейронауки могут помочь иначе взглянуть не только на поведение людей, но и на поведение машин. Искусственный интеллект, используемый в современных «умных» технологиях, таких как беспилотные автомобили, моделируется не на основе строгих формально-логических алгоритмов (серии условных операторов «если — то»), а на основе нервных систем человека и животных. Алгоритмы, смоделированные таким образом, называются нейросетями, они не являются жестко запрограммированными человеком, а учатся самостоятельно распознавать паттерны в данных посредством установления связей между искусственными нейронами, подобно тому как это происходит в восприятии и опыте живых существ. Понятие синаптической пластичности — усиление/ослабление связей между нейронами по мере накопления опыта, — а также поведенческие механизмы позитивного/негативного подкрепления находятся в центре исследования нейросетей и глубокого машинного обучения [Lipson, Kurman 2016: 200]. В данной статье показано, что современные технологии, моделирующие искусственный интеллект на основе знания нейронаук, любопытным образом служат наглядным эмпирическим подтверждением теоретических аргументов, высказанных в исследованиях наук и технологий (Science & Technology Studies, STS) в конце 1980-х — начале 1990-х годов. Кроме того, современная дискуссия о непрозрачности алгоритмов позволяет сместить перспективу на классический, но амбивалентный для STS троп раскрытия «черных ящиков».

Для этого я, во-первых, проанализирую три типичных ответа на проблему нейтральности технологий внутри и за пределами конструктивистских STS. Я покажу, что, несмотря на поверхностные различия между ними, все три ответа одинаково концептуализируют технологию как нейтральный проводник. Во-вторых, я артикулирую связь проблемы нейтральности с ключевым для STS тропом раскрытия «черных ящиков» технологий. В-третьих, я укажу на то, что альтернативу трем проанализированным до этого позициям в отношении проблемы нейтральности технологий можно найти в акторно-сетевой теории Бруно Латура, где технологии рассматриваются как непрозрачные и не-нейтральные сущности. Наконец, я использую иллюстрации из области технологий на основе нейросетей как наглядные эмпирические подтверждения латуровской концепции технической медиации.

Исследования наук и технологий и проблема (ценностной) нейтральности технологий

Одним из важных достижений исследований наук и технологий был трансфер методологий из социологии науки в социологию техники и обратно [Pinch, Bijker 1984], который нашел свое выражение в том числе в термине «технонаука» [Латур 2013]. Но не следует забывать, что у конструктивистских исследований науки и исследований технологии были свои предметные особенности. Выражением этой предметной специфики стали два разных морока, от которых они стремились избавиться. Мороком исследований науки был образ предсуществующей структуры мира, которая открывается учеными и затем воспроизводится (репрезентируется) в человеческой сфере с большей или меньшей точностью. Конструктивистские исследования науки были нацелены на борьбу с различными формами платонизма и «реализма» в понимании науки.

Морок исследований технологии совсем другой. Здесь никто никогда особенно не оглядывался на платонизм, поскольку очевидно, что технологии сделаны, сконструированы [Latour 1996: 23-24]. И именно потому что техника сконструирована, возникает проблема, вращающаяся вокруг вопроса о нейтральности техники. Пронизаны ли технические устройства социальными и политическими отношениями или они ценностно нейтральны? У этой проблемы есть обертон, связанный с фигурой прозрачности технологий, который я озвучу позже.

В первом приближении можно выделить три ответа на проблему нейтральности техники[1]. Первый ответ осложнен вышеописанным «реализмом», соединенным с концепцией отношений между наукой и техникой, где первая открывает, а вторая — применяет. В этой комбинации возникает понимание техники как просто материализации и конкретизации научного знания, открывающего человеку возможности для господства над природой [Ogburn 1964]. Поэтому неуспешные или неработающие технологии объявляются противоречащими природе вещей, т. е. в принципе нереализуемыми. Второй ответ заключается в том, что технологии нейтральны и могут быть использованы как для «добра», так и для «зла» [Pitt 2014]. Третий ответ состоит в том, что технологии внутренне пронизаны ценностями, поскольку они конструируются в социально-политическом процессе взаимодействия между разными группами [Winner 1988; Pinch, Bijker 1984].

1 См. свежее обсуждение этой проблемы в [Miller 2020], правда, выдержанное в ином ключе, чем этот анализ.

Кажется, что конструктивистские исследования технологий всецело на стороне третьего ответа. Но это только кажется. На деле, как станет ясно далее, все три ответа в интересующем нас здесь отношении не сильно различаются, поскольку концептуализируют технику как нейтральную. Разница лишь в том, делают ли они акцент на дизайне (проектировании/разработке) или на использовании, и где они локализуют условия возможности дизайна технологии.

Первый ответ делает акцент на дизайне и локализует условия его возможности за пределами общества. Материальная структура технического устройства либо репрезентирует онтологическую структуру природы, либо является применением и воплощением универсальных категорий человеческого рассудка [Weber 2012]. Так или иначе, техническое устройство — лишь послушный проводник внесоциальных сил. Этот ответ включен в данное обсуждение именно потому, что он хорошо показывает, что проблема нейтральности технологий не связана исключительно с дискуссией об их связи с ценностями. Тенденция видеть в технологии нейтральный проводник внешних сил (вне зависимости от их природы) более масштабна. Второй и третий ответ отличаются от первого тем, что вращаются вокруг концептуализации отношений между интенцией конструктора/пользователя и действиями технического устройства, но они также делают выбор либо в пользу использования, либо в пользу дизайна.

Второй ответ подразумевает, что внутри самой технологии не «зашито» никаких ценностей, но устройство послушно реализует волю или интенцию пользователя. Обычно здесь используют примеры вопиюще простых «технологий». Например, ножом можно с равным успехом резать и салат, и людей. Конечные эффекты использования техники всецело зависят от ценностей ее пользователя. Технология полностью определяется ее (социальным) использованием, пронизанным интересами и ценностями.

Третий ответ делает акцент на дизайне, но теперь его условия возможности локализованы не в природе, а в обществе. Устройство и динамика технологий определяется борьбой социальных групп и их интересов. Поэтому ценности и социально-политические отношения полагаются непосредственно вписанными в содержание устройств. Например, в одной из «дисциплинарных легенд» STS Лэнгдон Уиннер описывает социально-политическую подоплеку дизайна мостов на Лонг-Айленде главным архитектором, а потом мэром Нью-Йорка Робертом Мозесом. Он спроектировал мосты так низко, чтобы под ними к пляжам Манхэттена могли проехать белые владельцы автомобилей, но не чернокожие пассажиры автобусов [Winner 1988: 22–23]. Помимо того что, как выяснилось, это

фактически не так [Woolgar, Cooper 1999], и что этому анекдоту можно легко противопоставить контристорию Дж. Скотта о Бразилии, где «благие» и сильные намерения государства не реализовались по намеченному плану [Скотт 2005: 194–209], эта легенда Уиннера проблематична *концептуально*. Так же, как и первый ответ, она исходит из того, что технологии — послушные проводники человеческих интенций, воль и через это ценностей. Разница лишь в том, что теперь эти ценности полагаются вписанными в содержание самой технологии. Технология полностью определяется ее дизайном, но в отличие от первого ответа сам этот дизайн укоренен в обществе-политике и пронизан ценностями. Поэтому технологии — это политика другими средствами, способ осуществить дискриминацию и доминирование не только материально, но и приватно, т. е. за пределами широкой публичной дискуссии, ограничив делиберацию относительно узким кругом экспертов.

Все три ответа концептуализируют технологии как послушные проводники. Второй и третий ответы создают видимость противоречия только из-за того, что в них различается локализация ценностей. Если ценности локализованы за пределами устройства, технология полагается ценностно нейтральной. Если ценности каким-то образом удалось внедрить внутрь техники, то она ценностно, политически, социально ангажирована.

Прозрачность технологий и амбивалентный троп раскрытия «черных ящиков»

Проблема нейтральности связана с проблемой прозрачности посредством центрального для STS образа «черного ящика». Под акронимом STS скрывается то, что Стив Фуллер в свое время назвал «двумя церквями»: академической Science & Technology Studies и активистской Science, Technology & Society [Fuller 1997: 181–183]. Поэтому образ «черного ящика» имеет как методологическое, так и социально-политическое измерение.

В методологическом плане «черный ящик» — это один из приемов, позволяющих следовать за акторами и настраивать фокус исследования. Методологическое значение фигуры «черного ящика» определяется как минимум тремя пунктами.

1. Проводит границу между STS и другими исследованиями науки и техники на том основании, что любой, кто принадлежит этому полю, разделяет посылку, что все, что в момент времени t представляется непроблематичным и незыблемым, является результатом комплексной истории и было проблематичным и зыбким в момент t-1. Поэтому сторонники STS непременно раскрывают «черные ящики», т. е. рассказывают сложные и запутанные истории

фактов и артефактов, а не принимают их как данность, фокусируясь, например, только на их восприятии, оценивании или эффектах в обществе [Латур 2013: 25–27].

2. Поскольку в любой момент времени стабильное состояние любых наук и технологий представляет собой комплекс «черных ящиков», связанных друг с другом или вложенных друг в друга, то невозможно раскрыть их все, нужно где-то остановиться. «Черный ящик» позволяет провести для STS-исследователя границу, за которую нет смысла заступать. Если акторы используют факт или устройство как ресурс в своей текущей деятельности, игнорируя при этом его комплексную историю, то мы имеем дело с «черным ящиком». И если все наши акторы в равной степени считают данный факт или устройство «черным ящиком», то нам нет смысла расследовать его историю, поскольку это расследование не позволит провести различия между акторами и потому не будет участвовать в объяснении [Latour 1996: 18–19].

3. «Черный ящик» отсылает к границе того, что ученые и инженеры считают проблематичным или непроблематичным. Это позволяет сосредоточиться на том, что важно и интересно для них, и не задавать нерелевантных вопросов [Латур, Вулгар 2012: 178–179;, 199–206].

В социально-политическом плане троп раскрытия «черного ящика» сопрягается с разницей между официальным или публичным образом наук и технологий и приватными практиками их осуществления. Как и в случае других профессий, проблемой здесь становится шок публики от несоответствия между публичным фасадом и приватным положением дел. Раскрытие «черных ящиков» наук и технологий здесь выступает и ресурсом, и проблемой для STS в их взаимоотношениях со своими акторами. Это ресурс, поскольку речь идет об относительно уникальном типе знания, предлагаемом STS, и который не могут дать другие социологии наук и технологий. Но это и проблема. Раскрытие «черных ящиков» подразумевает вынесение приватной «кухни» управления и эксплуатации технологий в публичное пространство. Как хорошо известно из социологии профессий, такая операция часто воспринимается практиками и широкой публикой как угроза социальной легитимности в данном случае наук и технологий. И неважно, используется ли это знание для социальной критики изнутри STS или другими акторами, или просто шокирует здравый смысл публики. Неудивительно, что попытка заглянуть в «мягкое подбрюшье» таких авторитетных профессий, как наука и инженерия, привела STS в конфронтацию с представителями естественных наук, известную как «научные войны», несмотря на то что это делалось с натуралистическими (а не социально-критическими) целями [Edge 1979: 114; Bloor

1991 (1976): 46], т. е. чтобы научно описать само производство знания и техники.

Сочетание проблемы нейтральности и амбивалентного для STS тропа раскрытия «черных ящиков» технологий позволяет озвучить *обертон прозрачности*. Выше мы видели, что все три ответа рассматривают технику как нейтральный проводник внешних сил. Первый и третий ответы видят в дизайне устройств/систем воплощение (вне)социальных сил, определяющих использование технологий. Для второго ответа внутреннее устройство техники — лишь нейтральный инструмент для реализации воль, интенций, ценностей пользователя.

Но в этих концептуализациях технологии не только *нейтральны*, но и *прозрачны*. И те, кто говорят о ценностной нейтральности технологии, и те, кто раскрывают ее «черные ящики», склонны смотреть сквозь технологии на что-то еще более важное, вместо того, чтобы описывать их. В обоих случаях более или менее явно замешана двойная (социально)-критическая операция: а) переключение внимания с содержания науки и техники на условия их возможности; б) наделение этих условий бóльшим весом в объяснении. У этой операции есть методологическое и социально-политическое измерения. В первом случае имеет место интеллектуальная критика, а во втором — социальная критика.

Рассмотрим сначала методологическое измерение критики. Первый ответ переключает внимание с технологий на науку и далее наделяет большим весом либо онтологические (реализм), либо трансцендентальные условия возможности законов природы, воплощенные в материальном дизайне устройств. Второй ответ, аргумент о ценностной нейтральности, в принципе предполагает переключение с содержания устройства на его использование и редукцию технологии к использованию, т. е. на то, что делают с ней люди. Третий ответ — конструктивистский аргумент *a la* Уиннер предлагает раскрыть «черный ящик» технологий, но только чтобы сказать, что социально-политические условия возможности манхэттенских мостов (ранние фазы планирования и разработки) доминируют над всем остальным процессом. Во всех трех случаях мы смотрим сквозь технологии, на их условия возможности, которые полагаются более реальными и к которым можно редуцировать их материалы.

В социально-политическом плане раскрытие «черных ящиков» представляется переключением внимания с аргументов и конструкций инженеров на них самих и обстоятельства их работы. Эмпирическими коррелятами социально-критической операции являются разоблачения, сенсационность, шок здравого смысла публики, оттого что на самом деле все обстоит не так, как опи-

сывается в «официальном» дискурсе о технологиях. Гомология интеллектуальной и социальной критики[1] позволяет проследить и социально-политические эффекты транспарентизации технологий. Переключение внимания с технической вещи на условия ее возможности воспринимается и публикой, и профессионалами как ее дереализация, обесценивание, подрыв социальной легитимности, которым сопутствуют возмущение и защитная контркритика.

К непрозрачности технологий: переориентация критики в акторно-сетевой теории Б. Латура

Итак, понятно, что социально-конструктивистские взломщики «черных ящиков» по преимуществу оказываются под властью морока нейтральности и прозрачности технологий. Если так, то каковы альтернативные решения этой проблемы внутри STS? Общим трендом, маркирующим признание того, что социальный конструктивизм быстро превращается в социальный редукционизм, и стремление иначе решить проблему нейтральности/прозрачности технологии, стало обращение к аргументу о со-производстве технологий и общества [Jasanoff 2004; Bijker, Law 1992].

Однако это лишь тренд, внутри которого есть разные трактовки приставки «со-» в слове «со-производство»[2]. Некоторые подходы понимают это «со-» как диалектический или реципрокный процесс, где изначально есть два типа инстанций или реальностей — технологии (материальное) и общество, — которые, подобно эшеровским рукам, рисуют друг друга.

Но в акторно-сетевой теории, или социологии перевода, «со-» трактуется совсем иначе. Изначально подразумевается тождество или неразличимость социального и технического (как типов реальности) в едином процессе перевода. Внутри него различимы только эмпирически разные инстанции активности, идентичности которых поставлены на карту [Каллон 2015]. В ходе процесса эти эмпирически (а не по типу) разные акторы меняются или вовсе элиминируются. Результатом же сложного процесса перевода является производство границы между технологией и обществом как двумя типами реальности. Но это только поздний и локальный продукт

1 Ср. с гомологией операций исследователей науки и самих ученых в [Латур, Вулгар 2012: 27, 30-37].

2 Производство тоже трактуется по-разному, и, более того, появляются подходы, указывающие на ограниченность метафоры конструкции или производства и делающие акцент на перформативности и задействовании (enactment).

стабилизации. Это значит, что в будущем в отношении данного конкретного результата эта граница может быть вновь дестабилизирована. Это также значит, что в другом месте для процессов инновации или исследования, т. е. процессов перевода, будет характерно такое же тождество или отсутствие стабильной границы между «социальным» и «техническим» акторами.

Главная посылка, отличающая акторно-сетевую теорию, состоит в том, что исходно есть не два типа реальности, две инстанции, соединяемые диалектически или иным образом, а единый эмпирический процесс перевода или медиации, результатом которого является дистилляция двух «чистых» типов реальности после серии событий. Это достаточно прозрачно выражено в названии работы Латура [Latour 1990] «Технология *есть* общество, сделанное прочным».

На развитии этого решения я хотел бы сосредоточиться далее. Первоначально попытка Латура [2013: 225–229] избежать социального конструктивизма состояла в том, чтобы бифокально следить за тем, как в процессе технической инновации параллельно изменяются социальные (социограмма) и материальные элементы (техногramма). Это решение, однако, все еще допускало диалектическое прочтение. Поэтому вскоре эта рамка описания была преобразована в материально-семиотическую модель социотехнических графов [Latour, Mauguin, Teil 1992; Latour 1990; Latour 1992]. В этой модели социотехническая динамика инноваций определяется взаимодействием программ и антипрограмм, разделенных «линией фронта» разногласий вокруг исследуемой технологии. Программа действия — это желательный с точки зрения инженеров (в широком смысле) сценарий использования их технологии. Чтобы максимизировать соблюдение предложенной ими программы действия, инженеры увязывают свое желание с цепочкой гетерогенных (т. е. человеческих и не-человеческих) акторов. Антипрограммы — это обстоятельства или активности гетерогенных акторов, которые с точки зрения инженеров мешают реализации программы действия. Чтобы нейтрализовать антипрограммы и увеличить количество пользователей на стороне желательной для них программы действия, инженеры стремятся модифицировать технологию с помощью новых гетерогенных акторов. Эти модификации визуализируются с помощью графа, регистрирующего изменение содержания программы действия и количества следующих ей пользователей через смещение «линии фронта».

Эта теоретическая, стилистическая и графическая инновация имела несколько важных следствий. Во-первых, теперь по обе стороны от «линии фронта», т. е. как внутри программы действия тех-

нологии, так и в ее антипрограмме, оказываются элементы, которые с точки зрения здравого смысла можно назвать «социальными» и «техническими»[1].

Во-вторых, Латур избавился от необходимости выбирать между дизайном и использованием. На первых подступах к изучению технологий Латур был сконцентрирован на скриптах, вписанных в артефакты [Johnson 1988], что допускало возможность технодетерминистского прочтения, т. к. позволяло вычитывание социограммы из техногаммы. Однако с моделью социотехнических графов понятие технологии существенно трансформировалось[2]. Технология — это теперь не артефакт, в который посредством скриптов вписаны инженер и пользователь, но целая цепочка гетерогенных акторов и переводов, которые обеспечивают их соединение в одной программе действия. Программа действия какого-либо устройства *есть* общество. И если это общество стабилизировано, то мы имеем дело с технологией. По ту сторону «линии фронта» внутри антипрограммы могут находиться как отдельные «паразиты» [Latour 1996: 71–72], так и контробщества, с альтернативными устройствами и способами использования. Одним словом, по обе стороны диаграммы теперь находятся комбинации дизайнов и использований. Нет нужды видеть социотехническую динамику только как взаимодействие и противоборство между дизайнами и использованиями. И нет нужды делать выбор и отвечать на вопрос, что обладает большей детерминирующей силой, дизайн или использование. Одни использования могут быть добавлены к дизайну, а другие временно противопоставлены ему, так же как и альтернативные дизайны. Итак, вот что значит утверждение «технология *есть* общество». Впоследствии понятие ассоциации окончательно заменило понятие общества, которое отсылало к сущности, которая будто бы всегда уже существует до всякого процесса перевода, а не собирается заново в каждом конкретном случае (ре)инновации.

Таким образом, понимание технологии как общества, ассоциации разнородных элементов, позволяет описывать ее динамику не как взаимодействие дизайна и использования, а как противоборство программ и антипрограмм действия. При этом и те, и другие представляют собой цепочки людей, не-человеков, знаков, практик, идей, воплощенных в дизайнах, использованиях, ремонтах, обслуживании и т. д. Это позволяет не выбирать между

1 В поисках примера см. анализ социотехнической динамики камеры Кодак в [Latour 1990].

2 См. подробный анализ этой трансформации в [Кузнецов 2015].

этими компонентами главный, являющийся условием возможности всех остальных. Нет смысла устанавливать иерархию между компонентами, как если бы только главного из них достаточно для стабилизации технологии, а без остальных можно было бы обойтись. Нет, каждый из них по-своему является sine qua non. Поэтому нет смысла переключаться на описание условий возможности (контекста) техники, полагая, что этого будет достаточно для понимания процесса и события ее стабилизации. Нужно описывать содержание программ и антипрограмм, не предполагая, что все сводимо к главному компоненту, для которого остальные — деривативы.

Описание, если оно следует за технологией достаточно долго и охватывает все ее конституенты, эквивалентно объяснению [Latour 1990: 121, 129–130]. Фокусировка только на условиях возможности ничего не объясняет. В этом смысл латуровской переориентации критики. «Ошибкой было бы верить, что мы тоже дали социальное объяснение научным фактам. Нет, хотя и правда, что поначалу мы как добротные критики, обученные в хороших школах, пробовали использовать оружие, врученное нам нашими лучшими и старшими, чтобы взломать (одно из их любимых выражений означающее разрушить) религию, власть, дискурс, гегемонию. Но к счастью (да, к счастью!), один за другим мы стали свидетелями того, что *черные ящики науки оставались закрытыми*, и что это скорее [наши] орудия лежат в пыли наших мастерских, разобранные и сломанные. Проще говоря, критика была бесполезной в отношении достаточно твердых объектов» (курсив — *АК*) [Latour 2004: 242]. Новая исследовательская установка, в отличие от старой критической, подразумевает переориентацию с условий возможности технологии на описание того, «*как много участников* собрано в *вещи*, чтобы сделать ее существующей и поддерживать ее существование» [Ibid.: 245-246]. Новая критика должна быть направлена не *от*, а *к* технологии как обществу, ассоциации, ассамбляжу, собранию, вещи (а не объекту) [Ibid.: 246].

К этой обновленной критике, которая отучивает исследователя расщеплять технологию-ассоциацию на первичные и вторичные конституенты, добавляется понимание технического артефакта как чего-то непрозрачного, преподносящего сюрпризы как события. В только что процитированной статье Латур обращается к понятию суперкритического ума А. Тьюринга, который генерирует больше идей, чем получает. Это понятие «требует, чтобы все сущности, включая компьютеры, перестали быть объектами, определяемыми просто их “входами” и “выходами”, и вновь стали вещами, опосредующими, ассемблирующими, собирающими гораздо больше складок, чем “четверица” [Хайдеггера]» [Ibid.: 248].

Эта более поздняя формулировка Латура опирается на ряд модификаций, сделанных с начала 1990-х в понятии перевода. Понятие перевода, подразумевавшее, что есть тот, кто переводит, и тот, кто подвергается переводу, было замещено понятием медиации, которое не вызывает таких коннотаций. Медиация указывает на набор операций, предшествующих отношению, но направленных на его установление. Поэтому оно сохраняет смысловой компонент переговоров, подчеркивая, что идентичности обеих сторон отношения поставлены на карту. С этим связано определение действия в работе «Об интеробъективности». «*Faire c'est faire faire*. Делать — значит делать свершившимся. Когда один действует, другие переходят к действию» [Латур, 2006: 189]. Другое возможное прочтение этого трудного для перевода афоризма: «Делать — значит *делать делающим*». Оба прочтения дают понимание, что действие — это медиация, событие, устанавливающее связь, значимую для обеих сторон. X является актором, только если его начинание подхватывается другими Y, Z, которые переходят к действию. И нет смысла выбирать (как это было выше между дизайном и использованием) между X, Y, Z в качестве primum movens действия [ср. Каллон 2015: 206], только все они вместе — конституенты действия. В таком понимании концепт медиации применяется к технологии [Latour 1994].

Чуть позже в «Надежде Пандоры» эта концепция действия получает новый обертон: «действие слегка охвачено тем, на что оно действует» [Latour 1999: 298]. Создатель, равно как и пользователь устройства, охвачен тем, что создает или использует. Если не забывать, что технология — это общество, ассоциация, целая цепочка переводов, из которой не следует вычитать ничего, поскольку все конститутивно, то становится понятно, что технология не прозрачна (даже для своих конструкторов), а преподносит сюрпризы, делает то, что не определяется ее «входными параметрами». «Нет объекта, субъекта, противоречия, снятия (Aufhebung), господства, обобщения, духа, отчуждения. Но есть события. Я никогда не действую; я всегда слегка удивлен тем, что я делаю. То, что действует через меня, также удивлено тем, что я делаю, шансом мутировать, изменяться... Нас удивляет то, что мы делаем (make), даже когда мы обладаем, даже когда мы верим, что обладаем полным господством (complete mastery). Даже разработчица программного обеспечения удивлена своим творением после написания двух тысяч строк программы» [Ibid.: 281–283].

Наконец, в книге «Пересборка социального» Латур [2014: 153] пишет: «Слово “перевод” теперь приобретает специализированное значение: это отношение, не переносящее причинно-следственную связь, а приводящее к сосуществованию двух посредников». Введенное в этой книге различение между промежуточными звеньями

(intermediaries) и посредниками (mediators)[1] указывает на ту самую новую критическую ориентацию. Понятие технологии как посредника-медиатора позволяет понять технику как суперкритическую, соединяющую дизайны и использования, людей и нечеловеков.

Отталкиваясь от новой критической установки, наставляющей не выбирать между первичными и вторичными качествами технологий и не устанавливать априорной иерархии между ее конституентами, я обращусь к нескольким наглядным урокам, тому, что англоговорящие называют object-lessons, которые могут преподать нам современные умные технологии в связи с вышеизложенной концепцией технологии в акторно-сетевой теории. Для этого я использую иллюстрации из области технологий на основе нейросетей как наглядные эмпирические подтверждения вышеизложенной концепции акторно-сетевой теории. В мире технологий, «мозгом» которых являются нестрогие алгоритмы нейросетей, понимание технологий как непрозрачных и преподносящих сюрпризы своим создателям становится общим местом, разделяемым и акторами, и исследователями.

От следования правилам к их порождению

Ключевая особенность технологий на основе нейросетей — они не столько следуют правилам, сколько порождают их. Далее я сначала покажу, что концепция осязаемых материальных технологий как порождающих правила была теоретически сформулирована в STS уже в конце 1980-х — начале 1990-х. Затем я приведу две иллюстрации того, что подобное понимание разделяется современными инженерами строгих алгоритмов программного обеспечения и фактически становится эмпирической обыденностью технологий, использующих нестрогие алгоритмы нейросетей.

В свое время Брайан Уинн [Wynne 1988] показал неадекватность доминирующего в публичной сфере образа технологий действительной практике их функционирования и управления. В публичной сфере технологии понимаются как то, что следует предустановленным правилам и протоколам. Неадекватность этого образа проявляется в том, как резко меняется риторика экспертов и официальных представителей технологических систем после серьезных аварий. До события утверждается, что все под контролем, а после вступает в дело защитная и оправдательная риторика, утверждаю-

1 Надо заметить, что уже в «Науке в действии» Латур [2013: 228] употребляет различение проводники/мультипроводники примерно в том же самом смысле.

щая нереалистичность исходного представления о сложных технических системах как о не подверженных рискам. Технологические аварии и катастрофы благодаря следующим за ними публичным расследованиям дают представление о действительных практиках обращения с технологиями. Они показывают, что рутинное управление, эксплуатация и поддержание сложных технических систем не следует правилам, а генерирует их. «Практики не следуют правилам, скорее правила следуют за развивающимися практиками» [Wynne 1988: 153]. Сам факт, что актуальная работа технологий оказывает давление на правила, зафиксирован в обновлении регулирующих их кодексов, политики, протоколов. Но в повседневной рутине операторы технологических систем сталкиваются с ситуациями, требующими локальной нормализации технологий, т. е. не просто применения правил, подразумеваемых дизайном, а изобретения новых и изменения старых правил, переопределения стандартов безопасности и штатного функционирования, вводя представление о «нормальных авариях».

Такая нормализация ведет к непредсказуемым последствиям. Одним из аспектов нормализации является контекстуализация — попытка учесть локальные требования при развертывании и эксплуатации технологии. Контекстуализация нередко бывает частичной и ведет к внутренней фрагментации технологии. В ответ на локальные требования меняется только часть системы. В момент внесения изменений трудно предвидеть, вступят ли они в дальнейшем в противоречие с другими частями и функциями системы. Это создает потенциал для непредвиденных аварий и катастроф, часто отложенных во времени после длительного периода «нормальной» работы.

Эти идеи Уинна могут создавать впечатление, что изначально консистентный дизайн подвергается фрагментации в практике использования. Но Латур [Latour 1992] показал, что эти процессы имеют место и в социотехнической динамике разработки или дизайна технологий. Стремясь максимизировать соблюдение желательного для них сценария, инженеры добавляют к исходной программе действия технологии «патчи», являющиеся ответом на разные антипрограммы, частью которых могут быть, как мы видели, и альтернативные дизайны, и контриспользования, и элементы, не укладывающиеся ни в одну из этих категорий. Поэтому уже дизайн представляет собой необязательно консистентное наслоение ответов разработчиков на принятые ими в расчет разнородные и разнонаправленные антипрограммы технологии. Каждое из таких наслоений подразумевает создание *ad hoc* правила, которое добавляется к тем, что уже были внутри программы действия. За каждым из таких добавлений стоит большой объем работы. Гетерогенным

акторам, не желающим почему-то следовать сценарию инженеров, потребовались время и усилия на изобретение антипрограмм. Инженерам потребовались время и усилия на разработку «патчей» для программы действия, нейтрализующих антипрограммы. Отсюда ясно, что нет смысла рассматривать технологии как нейтральные проводники интенций или ценностей. Ни дизайн, ни использование не предполагают наличие за ними консистентного комплекса интенций или ценностей, а, напротив, соединяют и удерживают вместе разрозненные и противоречивые требования релевантных сущностей.

Все это становится еще более очевидным там, где, казалось бы, степень прозрачности и человеческого господства над материалом технологии должна быть самой высокой, а именно в области разработки программного обеспечения. Например, инженер программного обеспечения Джоэль Сполски критикует тенденцию программистов писать код с нуля, а не использовать повторно уже имеющийся код. Это происходит потому, что «читать код труднее, чем писать его». Вот как он описывает устройство кода, который был написан давно. «Вы можете спросить почти любого программиста сегодня о коде, с которым они работают. “Это большой волосатый беспорядок”, — скажут они... Почему это беспорядок? Они ответят: “Ну, посмотрите на эту функцию. Она длиной в две страницы! Ничего из этого не имеет отношения к ней! Я не знаю, зачем нужна половина из этих API-вызовов!”... Это просто функция отображения окна, но она нарастила на себе немного волос и хлама, и никто не знает почему. Ну, я скажу вам почему: это исправления ошибок. [...]. Прежде чем найти каждую из этих ошибок, потребовались недели использования в реальном мире. Программисту, возможно, потребовалось провести пару дней, чтобы воспроизвести ошибку в лаборатории и исправить ее [...]. [Даже] если ошибок много, исправлением может быть одна строка кода или даже пара символов, но много работы и времени было вложено в эти два символа. Когда вы выбрасываете код и начинаете с нуля, вы выкидываете все это знание. Все накопленные исправления ошибок. Годы программирования» [Spolsky 2000].

Выше речь шла о строгих алгоритмах. В случае нестрогих алгоритмов нейросетей вышеописанные свойства технологий проступают еще более рельефно и постепенно становятся частью публичного дискурса. В отличие от традиционного программного обеспечения (строгих алгоритмов) алгоритмы глубокого машинного обучения не программируются посредством предзаданных формальных правил, а обучаются, т. е. разучивают самостоятельно сгенерированные правила, чтобы решать поставленные перед ними задачи [Stilgoe 2017: 29]. Почему? Рассмотрим случай беспилотных автомобилей.

Для решения проблемы вождения машина должна воспринимать и классифицировать окружающую среду, чтобы затем принимать вероятностные решения при наступлении определенных условий. Но количество и сложность ситуаций в практике вождения не позволяет запрограммировать алгоритмы беспилотных автомобилей посредством формальных правил. Поэтому в качестве решения используются глубокие (многослойные) нейросети, обучающиеся на обширных дата-сетах. «Google не учат свои компьютеры водить. Они собирают данные — их машины проехали в сумме 200 000 миль, записывая все, что они видят, — и позволяют своим алгоритмам самостоятельно вычислять правила» [Vanderbilt 2012].

Таким образом, аргумент STS, что технологии (т. е. стабилизированные ассоциации гетерогенных акторов) — это не столько нейтральные проводники, просто следующие заложенным в дизайне правилам или командам пользователей, сколько медиаторы, обладающие собственной активностью, порождающей правила, этот теоретический аргумент становится эмпирической обыденностью там, где используются нестрогие алгоритмы нейросетей.

Возвращение «черных ящиков» и непрозрачности нейросетей

Успехи в разработке алгоритмов, управляющих беспилотными автомобилями, порождают оптимизм в отношении машинного генерирования правил вообще. Себастьян Трун в бытность тимлидом проекта беспилотных автомобилей Google, утверждал, что «данные могут создать лучшие правила» [Ibid.]. Но это не только риторика. Помимо транспорта нейросети все больше определяют решения в важных сферах социальной жизни. Способность программного обеспечения на основе глубокого машинного обучения генерировать свои правила вкупе с его распространением в значимых сферах жизни людей порождают дискуссию о политике алгоритмов, а точнее, об их подотчетности и прозрачности для публики и государства. Внутри этой дискуссии алгоритмы на основе глубоко машинного обучения и, в частности, нейросети представляют особый интерес, поскольку в последнее время они широко применяются для задач классификации, на основе которых делаются предсказания о поведении людей и принимаются решения, имеющие социальные последствия [Burrell 2016: 3]. Например, нейросети рекомендуют, стоит ли выдавать кредит конкретному человеку? Другие сферы применения таких алгоритмов касаются отслеживания индивидуального поведения потребителей, поиска информации, фильтрации сообщений, медицинской диагностики, образования и т. п.

В центре дискуссии о подотчетности алгоритмов вновь оказывается образ «черного ящика». Но теперь он получает иную окраску. Классический троп STS имел хождение преимущественно в академическом мире, указывая на проблематичную тенденцию социальных ученых выносить в анализе материальное содержание за скобки и фокусироваться на социально-политическом контексте и эффектах технологий. В этом ключе ряд авторов призывает к раскрытию «черных ящиков» нейросетей как средству аудита и установления прозрачности их работы [Pasquale 2015; Diakopoulos 2013].

Но эти призывы в целом проблематичны, хотя и уместны в отдельных случаях. Они упускают из виду, что в случае нейросетей образ «черного ящика» является не только исследовательским, но и акторским. Как мы увидим далее, нейросети непрозрачны или неинтерпретируемы не только для исследователей или аудиторов, но и для самих их создателей. Поэтому призывы к раскрытию «черных ящиков» алгоритмов встречают как минимум два возражения. С одной стороны, ставится под сомнение возможность раскрытия этих «черных ящиков» социологическими и историческими средствами [Stilgoe 2017: 29]. С другой — ставится под вопрос сам идеал прозрачности, до сих пор определявший дискуссию о публичной подотчетности алгоритмов. Даже если заглянуть внутрь «черных ящиков» возможно и необходимо, этого может быть недостаточно, чтобы сделать системы на основе глубокого машинного обучения подотчетными. Рассмотрим сначала первое, а затем второе возражение против устойчивого в STS тропа раскрытия «черных ящиков».

Чтобы понять первое возражение, нужно разобраться в различных типах непрозрачности технологий на основе нейросетей. Дженна Баррелл [Burrell 2016: 3–7] выделяет три типа непрозрачности. Во-первых, сами алгоритмы и дата-сеты, на которых их обучают, как правило, являются интеллектуальной собственностью разработавших их компаний или государств. Поэтому они объявляются секретом, а доступ к ним ограничивается. В случае алгоритмо-пилотируемых мобилей мы получаем ситуацию, в которой эти умные устройства принимают решения, имеющие масштабные эффекты в публичной сфере, но доступ к принципам и логике этих решений охраняется частной собственностью. Разумеется, критики такой политики алгоритмов видят в ней лишь прикрытие для уклонения от подотчетности перед регуляторными органами и широкой публикой с целью сокрытия паттернов дискриминации и манипуляции потребителями. Бороться с такой формой непрозрачности алгоритмов можно и нужно за счет развития движения за открытый код и за счет подотчетности кода различными формам

алгоритмического аудита. Но это не единственная форма непрозрачности алгоритмов.

Во-вторых, даже если мы имеем доступ к коду, способность его читать и писать является высококвалифицированным навыком, доступным очень немногим. Более того, это еще и высокоспециализированное умение. Поэтому даже программистам-инсайдерам может быть трудно разобраться в сегментах кода, к написанию которых они не имеют непосредственного отношения. Ставший ученым программист Кевин Слэвин утверждает: «Мы пишем вещи [код], которые мы больше не понимаем» [цит. по: Neyland 2015: 51]. Если для создателей алгоритмов они не являются полностью прозрачными, то для людей вне этого процесса — подавно.

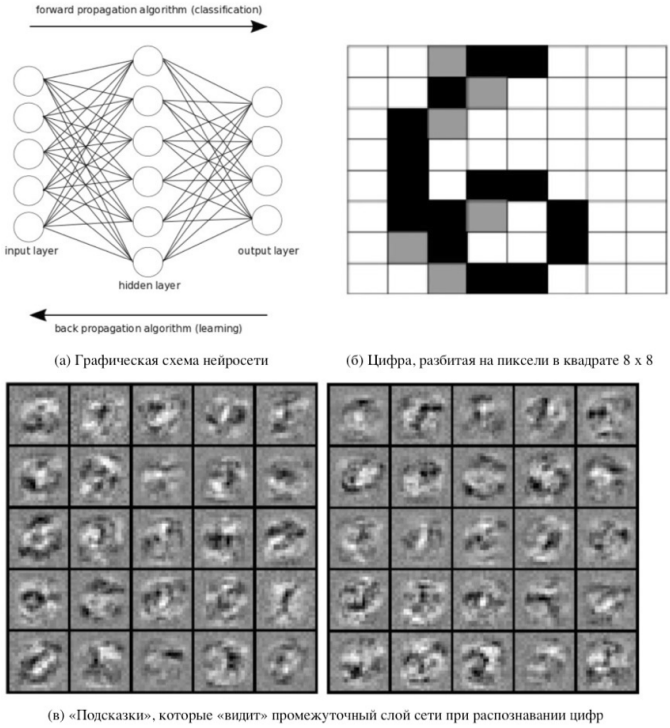

Рис. 1. Визуализация обучения нейросети при распознавании цифр [Burrell 2016].
Fig. 1. Visualization of neural network applied to handwritten digits recognition.

Наконец, третий тип непрозрачности алгоритмов обусловлен различием машинного и человеческого способов обучения. Дж. Баррелл рассматривает распространенный способ обучения не-

строгих алгоритмов, а именно контролируемое глубокое обучение нейросети, или обучение с учителем (рис. 1). Этот способ машинного обучения включает в себя два алгоритма: «классификатор» и «обучающийся». Сначала «обучающийся» проходит обучение на тестовых данных, результатом которого является «матрица весов», используемая в дальнейшем «классификатором». Тренировочные данные для «обучающегося» предварительно отбираются и размечаются человеком. «Классификатор» берет входные данные (набор характеристик) и производит выходные данные (категории), например, в виде диагноза болезни или квалификации сообщения как (не)спама. При этом входные и выходные данные образуют два отдельных слоя, связанных между собой промежуточным скрытым слоем (рис. 1а).

Как эта общая схема работает в хрестоматийном случае распознавания изображений, а именно рукописных цифр от 0 до 9? Чтобы упростить задачу для алгоритма, данные специально подготавливаются: каждая цифра записывается в четко очерченный квадрат, который затем разбивается на пиксели (рис. 1б). Каждая цифра представляется в виде комбинации пикселей (набора черт), каждому из которых соответствует тон на градиенте оттенков серого (матрица весов). Пиксели составляют входные данные для «классификатора», который производит выходные данные в виде распознанных цифр. На выходе мы имеем цифры от 0 до 9. Но промежуточный скрытый слой нейросети «видит» набор пикселей, характеризующий «входные» цифры, но только для машины (рис. 1в). Человек не угадывает в них паттерна, характеризующего цифры, но нейросеть успешно распознает изображения цифр на основе содержащихся здесь «подсказок». Поразительная визуальная трансформация происходит после того, как «обучающийся» выучил оптимальные значения для матрицы весов, под которыми понимаются точные классификации входных данных (отдельные пиксели разного тона) в выходные (цифры от 0 до 9).

Это значит, что, решая понятную человеку задачу, алгоритмы машинного обучения делают это не путем разбиения задачи на столь же понятные человеку подзадачи (например, распознавание вертикальной палочки, закругления или окружности как частей цифр). Они решают эту задачу своим собственным способом, не имеющим прямого эквивалента в человеческом мышлении. Дж. Баррелл заключает: «Первичной целью этого примера было дать быстрое визуальное понимание того, как “думает” машина. Последнее изображение должно казаться неинтуитивным, произвольным и дезорганизованным. Однако именно распознавание письма не является задачей, [решаемой путем] “сознательного” рассуждения и у человека. Люди распознают визуальные элементы непосредственно и подсо-

знательно (поэтому и в человеческом процессе распознавания символов также определенно есть своего рода непрозрачность)» [Ibid.: 7].

Второе возражение против тропа раскрытия «черных ящиков» связано с критикой идеала прозрачности как гарантии подотчетности и управляемости технологий на основе нейросетей. М. Ананни и К. Кроуфорд критикуют эпистемологические посылки идеала прозрачности, укорененные в корреспондентской теории истины, и выделяют десять его ограничений как руководящей нити политики алгоритмических систем. Для данной дискуссии важны три из них. Во-первых, неадекватность идеала прозрачности задаче аудита нейросетей проявляется в редукции «знать» к «видеть». «Смотреть внутрь системы не обязательно значит понимать ее поведение или истоки [...]. Узнавать о сложных системах значит не просто быть способным заглянуть внутрь систем или разобрать их на части. Скорее, это значит динамически взаимодействовать с ними, чтобы понять, как они ведут себя в отношении своих окружений» [Ananny, Crawford 2016: 980–981].

Во-вторых, алгоритмические системы и, в частности, нейросети не являются полностью прозрачными и подотчетными для своих собственных создателей. Известен казус с алгоритмами Google Photos, которые по непонятным для их разработчиков причинам распознавали чернокожих людей как горилл [Dougherty 2015]. Подобные случаи имели место и с алгоритмами HP и Nikon [Wade 2010]. «Доступ к коду может быть необходимым, чтобы привлечь систему к ответственности, но видеть код недостаточно. Сами системные разработчики часто не способны объяснить, как работает сложная система, какие части существенны для их работы, или как эфемерная природа вычислительных репрезентаций совместима с законами прозрачности» [Ananny, Crawford 2016: 982].

В-третьих, как уже понятно из предшествующей дискуссии, технологии обладают динамикой. «Социальное» и «техническое» в традиционном смысле в этой динамике переплетаются. Это справедливо как для осязаемых технологий и программного обеспечения на основе строгих алгоритмов, так и для нестрогих алгоритмов типа нейросетей. Поэтому даже обеспечения доступа к исходному коду полным обучающим и тестировочным дата-сетам может быть недостаточно для понимания функционирования системы. В расчет необходимо принять итеративные и интерактивные операции, связанные с работой и обновлением системы. Это особенно важно в случае нейросетей, обучающихся и адаптирующихся по мере поступления новых данных. «Нет “одной” системы, чтобы посмотреть внутрь нее, когда сама система распределена между и встроена в окружения, которые определяют ее работу» [Ibid.: 982].

Учитывая все это, М. Ананни и К. Кроуфорд предлагают не заглядывать внутрь «черных ящиков» алгоритмов, а рассматривать их по всей их протяженности (look across) и видеть в них «социотехнические системы, которые не *содержат* сложность, но *задействуют* (enact) сложность, соединяясь и переплетаясь с ассамбляжами людей и не-людей» [Ibid.: 974]. Вторя латуровской переориентации критики, они заключают: «Если истина — это не позитивистское открытие, а реляционное достижение между соединенными в сеть человеческими и не-человеческими агентами, то мишень прозрачности должна сместиться. [Е]сли систему до́лжно увидеть, чтобы понять и сделать подотчетной, то тип «ви́дения», требуемый акторно-сетевой теорией истины, подразумевает смотрение не *внутрь* чего-либо, но *по всей протяженности* системы» [Ibid.: 983–984].

Заключение

Подведем итог дискуссии, которую можно и нужно продолжить. С одной стороны, задачей этой работы было показать, что нестрогие алгоритмы в форме искусственных нейросетей, проходящих глубокое обучение, наглядно подтверждают то, что STS пытались доказать на примере других, менее замысловатых технологий. С середины 1980-х они стремились убедить специалистов и широкую публику, что машины, хотя и являются полностью сделанными человеком, содержат в себе непрозрачность, неопределенность и могут удивлять своих создателей. Сегодня же то, что еще каких-то 10 лет назад было смелым и неочевидным теоретическим тезисом, проговаривается официальными лицами Netflix, которые удивляются поведению своих алгоритмов генерирования микрожанров кино [Madrigal, 2014]. Самообучающиеся нейросети, служащие мозгом многих современных умных устройств, делают образ «призрака в машине» обыденным.

С другой стороны, эмпирические реалии и научные дискуссии вокруг технологий на базе нейросетей, в частности, и алгоритмов машинного обучения вообще показывают, что центральный для STS троп раскрытия «черных ящиков» изменил свое значение. Изначально он был связан с двойной критической операцией: переключением внимания с наблюдаемых эффектов технологии на условия ее возможности и приоритизацией этих условий возможности перед остальными конституентами технологии. Понятый в русле старой критики, этот троп имел амбивалентные последствия как в социально-политическом, так и в методологическом плане. В частности, сравнительно быстро стало понятно, что социально-конструктивистское раскрытие «черных ящиков»

склонно представлять технологию как нейтральную и прозрачную, как и другие неконструктивистские решения. Попытки схватить «плотность» и непрозрачность технологий нашли выражения в различных концепциях со-производства технологии и общества.

Вероятно, дальше всех в этом направлении продвинулась акторно-сетевая теория и Бруно Латур, предложивший переориентацию критики. Эта переориентация включает в себя два компонента: 1) понимание технологии как общества, ассоциации, собрания, ассамбляжа, вещи, где ни один из конституентов не приоритизируется; 2) понимание технологии как непрозрачной и преподносящей сюрпризы даже своим создателям. В результате троп раскрытия «черных ящиков» теперь указывает на социально-политические и методологические ограничения старой критической установки в понимании и управлении новыми технологиями.

Библиография / References

Каллон М. (2015) Некоторые элементы социологии перевода: одомашнивание морских гребешков и рыбаков залива Сен-Бриё. *Социология власти*, 27 (1): 196–231.

— Callon M. (2015) Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *Sociology of Power*, 27 (1): 196–231. — in Russ.

Кузнецов А.Г. (2015) Латур и его «технолог»: вещи, объекты и технологии в акторно-сетевой теории. *Социология власти*, 27 (1): 55–89.

— Kuznetsov A.G. (2015) Latour and His “Technologist”: Things, Objects, and Technologies in Actor-Network Theory. *Sociology of Power*, 27 (1): 55–89. — in Russ.

Латур Б., Вулгар С. (2012) Лабораторная жизнь. Конструирование научных фактов. Глава 2. Антрополог посещает лабораторию. *Социология власти,* 6–7: 178–234.

— Latour B. (2012) Laboratory Life. Chapter 2. An Anthropologist Visits the Laboratory. *Sociology of Power*, 6–7: 178–234. — in Russ.

Латур Б. (2006) Об интеробъективности. В.С. Вахштайн (ред.) *Социология вещей*, М.: Изд. дом «Территория будущего»: 169–199.

— Latour B. (2006) On Interobjectivity. V.S. Vakhshtayn (ed.) *Sociology of Things*, M.: Territory of the Future: 169–199. — in Russ.

Латур Б. (2013) *Наука в действии: следуя за учеными и инженерами внутри общества*, СПб.: Изд-во Европейского ун-та в С.-Петербурге.

— Latour B. (2013) *Science in Action: How to Follow Scientists and Engineers Through Society*, SPb.: EUSP. — in Russ.

Латур Б. (2014) *Пересборка социального: введение в акторно-сетевую теорию*, М.: ИД ВШЭ.

— Latour B. (2014) *Reassembling the Social: An Introduction to Actor-Network-Theory,* M.: HSE. — in Russ.

Скотт Дж. (2005) *Благими намерениями государства*, М.: Университетская книга.

— Scott J. (2005) *Seeing Like a State,* M.: University book. — in Russ.

Ananny M., Crawford K. (2016) Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society,* 20 (3): 980–981.

Bijker W., Law J. (eds) (1992) *Shaping Technology / Building Society: Studies in Sociotechnical Change*, Cambridge, Mass.; London: MIT Press.

Bloor D. (1991 [1976]) *Knowledge and Social Imagery*, 2nd ed., Chicago; London: The University of Chicago Press.

Burrell J. (2016) How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society,* 3 (1): 1–12.

Diakopoulos N. (2013) *Algorithmic Accountability Reporting: On the Investigation of Black Boxes.* Report, Tow Center for Digital Journalism, Columbia University. (https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2)

Dougherty C. (2015) Google photos mistakenly labels black people "gorillas". *The New York Times*, 1 July. (http://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas)

Edge D. (1979) Quantitative Measures of Communication in Science: A Critical Review. *History of Science*, 17 (2): 102–114.

Fuller S. (1997) Constructing the High Church-Low Church Distinction in STS Textbooks. *Bulletin of Science, Technology & Society,* 17 (4): 181–183.

Jasanoff S. (ed.) (2004) *States of Knowledge: The Co-Production of Science and the Social Order*, London; New York: Routledge.

Johnson J. (1988) Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer. *Social Problems,* 35 (3): 298–310.

Latour B., Mauguin Ph., Teil G. (1992) A Note on Socio-Technical Graphs. *Social Studies of Science,* 22 (1): 33–57.

Latour B., Woolgar S. (1986) *Laboratory Life: The Construction of Scientific Facts*, Princeton, New Jersey: Princeton University Press.

Latour B. (1990) Technology Is Society Made Durable. *The Sociological Review,* 38 (1_suppl): 103–131.

Latour B. (1992) Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. W. Bijker, J. Law (eds) *Shaping Technology / Building Society: Studies in Sociotechnical Change*, Cambridge, Mass.; London: MIT Press: 225–259.

Latour B. (1994) On Technical Mediation: Philosophy, Sociology, Genealogy. *Common Knowledge,* 3 (2): 29–64.

Latour B. (1996) *Aramis, or The Love of Technology*, Cambridge, Mass.; London: Harvard University Press.

Latour B. (1999) *Pandora's Hope: Essays on the Reality of Science Studies*, Cambridge, Mass.: Harvard University Press.

Latour B. (2004) Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry*, 30 (2): 225–248.

Lipson H., Kurman M. (2016) *Driverless: Intelligent Cars and the Road Ahead*, Cambridge, Mass.: MIT Press.

Madrigal A. (2014) How Netflix Reverse-Engineered Hollywood? *The Atlantic*, January 2. (https://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/)

Miller B. (2020) Is Technology Value-Neutral? *Science, Technology, & Human Values*, First Published Online January 22. (https://journals.sagepub.com/doi/10.1177/0162243919900965)

Neyland D. (2015) Bearing Account-Able Witness to the Ethical Algorithmic System. *Science, Technology, & Human Values*, 41 (1): 50–76.

Ogburn W.F. (1964) *On Culture and Social Change: Selected Papers*, Chicago: University of Chicago Press.

Pasquale F. (2015) *The Black Box Society*, Cambridge, Mass.; London: Harvard University Press.

Pinch T., Bijker W. (1984) The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, 14 (3): 399–441.

Pitt J.C. (2014) Guns Don't Kill, People Kill"; Values in and/or around Technologies. P. Kroes, P.-P. Verbeek (eds) *The Moral Status of Technical Artifacts*, Dordrecht, the Netherlands: Springer: 89-101.

Spolsky J. (2000) *Things You Should Never Do*, Part I. (https://www.joelonsoftware.com/2000/04/06/things-you-should-never-do-part-i/)

Stilgoe J. (2017) Machine Learning, Social Learning and the Governance of Self-Driving Cars. *Social Studies of Science*, 48 (1): 25–56.

Vanderbilt T. (2012) Let the robot drive: The autonomous car of the future is here. *Wired*, January 20. (https://www.wired.com/2012/01/ff_autonomouscars/2/)

Wade L. (2010) HP software doesn't see black people. *Sociological Images*, 5 January. (https://thesocietypages.org/socimages/2010/01/05/hp-software-doesnt-see-black-people/)

Winner L. (1988) *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, Chicago: University of Chicago Press.

Woolgar S., Cooper G. (1999) Do Artefacts Have Ambivalence? Moses' Bridges, Winner's Bridges and Other Urban Legends in S&TS. *Social Studies of Science*, 29 (3): 433–449.

Weber A. (2012 [1920]) Fundamentals of Cultural Sociology: Social Process, Civilizational Process and Cultural Movement. C. Loader (ed.). *Alfred Weber and the Crisis of Culture, 1890-1933*, New York: Palgrave Macmillan, US: 165-205.

Wynne B. (1988) Unruly Technology: Practical Rules, Impractical Discourses and Public Understanding. *Social Studies of Science*, 18 (1): 147–167.

Рекомендация для цитирования:

Кузнецов А.Г. (2020) Туманности нейросетей: «черные ящики» технологий и наглядные уроки непрозрачности алгоритмов. *Социология власти*, 32 (2): 157-182.

For citations:

Kuznetsov A.G. (2020) Neural Network Nebulae: 'Black Boxes' of Technologies and Object-Lessons From Opacities of Algorithms. *Sociology of Power*, 32 (2): 157-182.

Поступила в редакцию: 12.06.2020; принята в печать: 23.06.2020

Received: 12.06.2020; Accepted for publication: 23.06.2020