

DOI: 10.22363/2313-2272-2019-19-3-581-586

## Искусственный интеллект как проблема современной социальной аналитики\*

**Рецензия на книги: Boden M.A. AI: Its Nature and Future.**  
Oxford: Oxford University Press, 2016. 156 p.;  
**Collins H. Artificional Intelligence: Against Humanity's Surrender to Computers.** Cambridge: Polity Press, 2018. 232 p.

Способен ли искусственный интеллект превзойти человеческий разум? Произойдет ли это в ближайшее время? Каковы последствия развития искусственного интеллекта для человечества? Сегодня это принципиальные вопросы не только для научной фантастики или журналистики, но и для ученых, причем в равной степени для представителей естественных наук и исследователей социальных проблем. Что не удивительно, если принять во внимание бесспорные достижения в разработке и включении искусственного интеллекта в повседневную жизнь современного общества. Проникновение в жизнь людей и ускоренная «социализация» превращают искусственный интеллект в том числе в социологическую проблему. Однако профессиональные социологи и социальные аналитики гораздо реже, чем философы или психологи, обращаются к проблематике, очерченной тремя вопросами, вынесенными в начало абзаца.

Мы попытаемся кратко проанализировать и сравнить две монографии, посвященные поискам ответов на проблемы развития искусственного интеллекта [4; 5]. Рецензируемые книги написаны признанными классиками в своих исследовательских областях (Маргарет Боден — в когнитивных исследованиях, Гарри Коллинз — в социологии науки), карьеры которых посвящены в том числе изучению проблем искусственного интеллекта (безусловно, авторы знают друг о друге, но идеи друг друга не обсуждают: Боден не упоминает Коллинза; Коллинз ссылается на другую работу Боден по истории искусственного интеллекта, только чтобы отметить, что она упомянула его исследования по той же проблематике, тем самым легитимировав себя как эксперта в этой области). Обе книги подводят промежуточные итоги развития искусственного интеллекта, и их авторы пытаются найти ответы на разные вопросы — в этом, вольно или невольно, проявляется сущность дисциплинарных и теоретических разделений в области исследований искусственного интеллекта. Тем не менее, на наш взгляд, эти работы интересно читать и анализировать именно вместе.

---

\* © Резаев А.В., Трегубова Н.Д., 2019.

Статья подготовлена при поддержке Российского научного фонда. Проект № 18-18-00097.

Статья поступила 06.05.2019 г. Статья принята к публикации 07.06.2019 г.

Сравнивая монографии, следует обратить внимание на три аспекта. Первый — соотношение между разными «искусственными интеллектами»: между наукой и технологией, технологией и идеологией, частным и общим искусственным интеллектом, символическим и коннекционалистским. Второй — границы и разделения внутри самой исследовательской области. Третий — пространство дебатов, которое создается различными подходами и дисциплинарными перспективами в исследованиях искусственного интеллекта. В целом мы попытаемся ответить на вопрос: нужно ли социологам заходить на данное проблемное поле, и если да — то с каких сторон?

Прежде всего, обе книги представляют введение в проблематику искусственного интеллекта для тех читателей, которые знакомы с ней лишь поверхностно. Здесь, однако, следует сделать оговорку: рецензируемые книги — научные монографии, а не научно-популярные бестселлеры, учебники или энциклопедии. Каждая из них выстроена вокруг изложения авторской позиции: это очевидно у Коллинза, менее очевидно, хотя и различимо у Боден. Специфика исследовательского поля такова, что в вопросах об искусственном интеллекте не существует консенсуса, поэтому даже подбор сюжетов и имен выражает позицию автора. Еще одна необходимая оговорка заключается в том, что оба автора не являются экспертами по технологиям искусственного интеллекта: это интеллектуалы, теоретики и отчасти эмпирические исследователи, которые анализируют работу искусственного интеллекта, созданного не ими.

Итак, интерес авторов к искусственному интеллекту различается, что связано как с их дисциплинарной принадлежностью, так и с личными предпочтениями: было бы неверно утверждать, что Боден представляет позицию всех когнитивистов, а Коллинз — всех социологов. Интерес Боден состоит в моделировании/воспроизведении искусственным интеллектом человеческого разума (*mind*) и жизни вообще. Название монографии можно перевести как: «Искусственный интеллект: его природа и будущее». «Природа» (*nature*) появляется в нем не случайно: Боден придерживается позиции, согласно которой нет принципиального различия между разумом человека и искусственным интеллектом, поскольку оба являются совокупностью «виртуальных машин». Однако, согласно Боден, есть существенная разница между теоретической возможностью создания интеллекта, подобного человеческому, и практической реализацией этой цели. В ходе изложения автор постоянно обращается к различию между технологическими и научными целями в разработках искусственного интеллекта: если технологические связаны с решением конкретных задач, то научные определяются тем, как искусственный интеллект может помочь человечеству в развитии знаний о мозге и эволюции живой материи. Монография балансирует между описанием существующих технологий искусственного интеллекта и характеристикой более или менее фантастических прогнозов о будущем его развитии, однако основная тема книги — возможность общего искусственного интеллекта (*artificial general intelligence, AGI*). Основной тезис автора состоит в том, что создание *AGI* возможно, и основное препятствие на этом пути — недостаток знаний о том, как работает человеческий мозг.

В первой главе Боден вводит читателя в историю проекта искусственного интеллекта, его наддисциплинарного духа и междисциплинарных делений, и представляет проблематику «виртуальных машин». Вторая глава начинается с поста-

новки проблемы общего искусственного интеллекта и далее распадается на описание проектов, которые пытались реализовать амбиции AGI, однако сталкивались с разного рода препятствиями (в основном, это проекты в традициях символического искусственного интеллекта). Третья глава обсуждает проблемы в воспроизведении языка, креативности и эмоций — тех аспектов человеческого существования, которые часто игнорируются разработчиками искусственного интеллекта, однако, согласно Боден, необходимы для создания общего искусственного интеллекта. Четвертая и пятая главы представляют экскурс в подходы к созданию искусственного интеллекта, связанные с искусственными нейронными сетями (коннекционализмом), робототехникой и «искусственной жизнью». Шестая глава вводит читателя в философские дебаты о том, является ли интеллектом то, что создается под рубрикой «искусственного интеллекта». Здесь Боден подчеркивает сложность дискуссий и глубину теоретических различий в проблемах, связанных с философским осмыслением искусственного интеллекта, и мягко обозначает свою позицию. В седьмой главе автор рассматривает идею сингулярности, обсуждая вероятные последствия развития новых технологий для жизни людей, включая социальные и этические проблемы.

Монография Боден в большей степени, нежели книга Коллинза, может читаться как справочник по искусственному интеллекту, в том числе и потому, что позиция автора укладывается в «мейнстрим» исследовательского поля, представляя собой одну из вариаций гипотезы о мозге как о вычислительной машине. Большинство разработчиков конкретных технологий, вероятно, не задумываются о философских основаниях проекта искусственного интеллекта, однако его основатели и идеологи вдохновлялись — и продолжают вдохновляться — идеей о создании разума, подобного или превосходящего человеческий. Чтобы подобные амбиции могли существовать, необходимо принимать допущение, что между мозгом и компьютером нет принципиальных различий. Это утверждение подвергалось критике с разных сторон, прежде всего, со стороны философов (и не только [3]). Социологическая критика искусственного интеллекта менее известна, но не менее интересна, и именно к ней и обращается Коллинз.

Название монографии Коллинза мы рискнем перевести как «Измысленный интеллект: Против капитуляции человечества перед компьютерами». Несмотря на название, тон книги скорее нейтральный, с последовательными и взвешенными аргументами. Тем не менее, в отличие от Боден, Коллинз начинает с критики в адрес тех, кто полагает, что современные технологии искусственного интеллекта вот-вот превзойдут человека. Его интересует не сама теория искусственного интеллекта (Коллинз не обсуждает гипотезу о мозге как компьютере), а то, как работают конкретные технологии в ситуациях социального взаимодействия. Основной тезис книги автор формулирует следующим образом: при существующих технологиях невозможно создать искусственный интеллект, укорененный в человеческом обществе и, как следствие, способный участвовать в разговорах наравне с людьми. Поэтому капитуляция перед впечатляющими достижениями машин является преждевременной. Для Коллинза именно владение языком (операционализируется как прохождение теста Тьюринга) является показателем общего искусственного интеллекта, функционально (если не содержательно) равного человеческому.

В первой главе Коллинз формулирует основную проблему книги и предлагает обзор существующих технологий искусственного интеллекта, которые подразделяет на основанные на правилах (символический искусственный интеллект) и основанные на паттернах и прецедентах (искусственные нейронные сети). Во второй главе автор проясняет методологические основания собственной аргументации, отвечая на два вопроса: что значит, что создание искусственного интеллекта с такими-то характеристиками «невозможно», и кто является экспертом в поле исследований искусственного интеллекта? Третья глава анализирует «ремонт» (repairs) в разговоре: в повседневной жизни люди постоянно — и незаметно для себя — осуществляют «ремонт», и, чтобы доказать владение языком, искусственный интеллект должен быть способен осуществлять подобный «ремонт». Четвертая глава рассматривает проблему социального контекста, а также вопрос о телесном воплощении интеллекта. В пятой главе обсуждаются шесть уровней искусственного интеллекта — от простых технических устройств до «искусственных обществ»; теоретический анализ сопровождается разбором конкретных достижений в области искусственного интеллекта и их отнесением к соответствующим уровням. Здесь аргументация Коллинза, в отличие от сторонников идеи сингулярности, позволяет провести различия между разными вариантами квази- и сверхчеловеческого искусственного интеллекта. Шестая и седьмая главы рассматривают принципы работы искусственных нейронных сетей — наиболее активно развивающегося направления искусственного интеллекта. Особое внимание уделяется соотношению распознавания паттернов «снизу вверх» (из данных) и «сверху вниз» (определяемых теорией или социальным контекстом) в алгоритмах искусственного интеллекта в сравнении с человеческим мозгом. Девятая глава обсуждает то, как люди учатся новому и какие человеческие свойства не могут быть воспроизведены современными компьютерами. Девятая и десятая главы суммируют аргументацию книги и приводят примеры теста Тьюринга, улучшенного в соответствии с логикой монографии; эти примеры читатель может реализовать на собственных гаджетах.

По сравнению с Боден Коллинз задает больше специфических вопросов, приводит больше конкретных примеров использования технологий искусственного интеллекта в социальном взаимодействии. Его аргументация более открыта, проблемы поставлены более остро и неожиданно, обратной стороной чего является несколько запутанная структура изложения, пестрота сюжетов и линий рассуждения. Чью аргументацию и стилистику предпочесть — выбор читателя: обе монографии, безусловно, достойны прочтения. Вместе с тем после прочтения обеих книг, несмотря на глубину и стройность изложения, остается чувство неудовлетворенности, возможно, потому, что для каждого автора рассматриваемая монография — не первая работа по искусственному интеллекту и основные аргументы (и контраргументы) формулируются не в первый и не во второй раз. Складывается ощущение, что Боден и Коллинз «заперты» внутри своих теоретических позиций и с трудом могут посмотреть на них со стороны.

Как соотносятся позиции авторов? Было бы неверно утверждать, что они противоречат друг другу: успехи и неудачи в развитии искусственного интеллекта трактуются авторами в общем сходно, хотя и с разными акцентами. Оба, по-видимому, полагают, что достижение AGI — если оно возможно в принципе — требует

концептуального и технологического прорыва. Так, Боден могла бы утверждать, что монография Коллинза обсуждает и уточняет одну проблему в разработках искусственного интеллекта — социального контекста/релевантности, для решения которой мы пока не знаем подходящих «виртуальных машин». Вместе с тем в аргументации Коллинза нет ничего, что поддерживало бы позицию Боден.

Следует ли говорить, что их позиции — и шире, социологический и философско-когнитивистский взгляд на искусственный интеллект — несоизмеримы, находятся на разных уровнях абстракции? На наш взгляд, это было бы упрощением по отношению к проблемному полю, к которому относятся монографии. Данное поле можно обозначить как «теория/критика искусственного интеллекта». Боден предлагает теорию искусственного интеллекта, Коллинз — его критику с позиций социологии. Боден в своих рассуждениях следует напрямую от мозга (*mind-brain*) к компьютеру, по касательной упоминая телесное воплощение и социальный контекст. Для Коллинза миновать социальный контекст — значит не видеть сути проблемы: мышление человека социально, оно существует и развивается в общении. Следовательно, искусственный интеллект, чтобы превратиться в подлинный интеллект, должен стать социальным. Это социологическая аргументация в чистом виде, однако то, как ее разворачивает Коллинз, связано с его пониманием социального взаимодействия: взаимодействие — это то, что происходит в языке, а ситуация взаимодействия является лишь контекстом разговора. Но в рамках социологической теории такое понимание не является единственным. Достаточно вспомнить другого Коллинза — Рэндалла, автора теории ритуалов взаимодействия, включающей «невербальное» общение (совместное курение, сексуальное взаимодействие), который упоминает и проблему искусственного интеллекта [2; 6].

Следующий шаг состоит в том, чтобы переключить внимание с языка на ситуацию взаимодействия и заметить, что люди обладают телом (попутно можно заметить, что млекопитающие и птицы со сложной социальной организацией также обладают интеллектом, во многом сходном с человеческим, однако не говорят). Проблема телесности подводит нас к третьей стороне дебатов, с которой полемизируют и Коллинз, и Боден, — к критике искусственного интеллекта с позиций экзистенциальной и феноменологической философии, представленной, главным образом, фигурой Хьюберта Дрейфуса (Коллинз упоминает Дрейфуса, обсуждает и приводит ссылки на их дебаты; Боден ограничивается упоминанием феноменологической традиции критики искусственного интеллекта). С позиции Дрейфуса мозг не сводится к вычислительной машине, а телесное воплощение является необходимым условием разума, подобного человеческому, в том числе условием социальной организации [1; 7]. Третья позиция важна потому, что, прежде всего, характеристика поля была бы без нее искаженной — Боден и Коллинз не дискутируют между собой, однако спорят с Дрейфусом, что свидетельствует о связности аргументов и дебатов, а не о двухуровневой модели «философия—социология». Кроме того, данная теоретическая позиция может выступать важным ресурсом для некоторых — но не всех — направлений социологии, анализирующих проблематику искусственного интеллекта, подобно тому, как позиция Боден может быть принята теми, кто развивает отдельные направления социологии коммуникации.

Таким образом, рецензируемые монографии заслуживают внимания читателя по нескольким причинам. Во-первых, это солидные научные работы, принадлежащие выдающимся специалистам в своих областях, подводящие итог их работе в области искусственного интеллекта (как и целого поколения исследователей). Во-вторых, сама область искусственного интеллекта становится все более актуальной для социальных аналитиков, поскольку его технологии входят в повседневную жизнь. Сегодня простая грамотность требует знакомства с тем, что такое искусственный интеллект и как можно его исследовать. Наконец, рассматриваемые монографии интересны не только в отдельности — прочитанные вместе, они намекают пространство дебатов об искусственном интеллекте, которое содержит разные маршруты «вхождения», и некоторые из них, вероятно, пока не проложены. А это значит, что именно там может возникнуть «передний край» социальной аналитики, и, возможно, именно социологам суждено внести свежесть и новизну в старые дебаты.

### Библиографический список / References

- [1] Бергер П., Бергер Б., Коллинз Р. Личностно-ориентированная социология. М.: Академический проект, 2004 / Berger P., Berger B., Collins R. *Lichnostno-orientirovannaya sotsiologiya* [Person-Oriented Sociology]. Moscow: Akademichesky projekt; 2004 (In Russ.).
- [2] Дрейфус Х. Чего не могут вычислительные машины: Критика искусственного разума. М.: Изд-во «Прогресс», 1978 / Dreyfus H. *Chego ne mogut vychislitelnye mashiny: Kritika iskusstvennogo razuma* [What Computers Can't Do: A Critique of Artificial Reason]. Moscow: Izd-vo "Progress"; 1978 (In Russ.).
- [3] Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики. М.: Эдиториал УРСС, 2003 / Penrose R. *Novy um korolya: O kompyuterakh, myshlenii i zakonakh fiziki* [The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics]. Moscow: Editorial URSS; 2003 (In Russ.).
- [4] Boden M.A. *AI: Its Nature and Future*. Oxford: Oxford University Press; 2016.
- [5] Collins H. *Artificial Intelligence: Against Humanity's Surrender to Computers*. Cambridge: Polity Press; 2018.
- [6] Collins R. *Interaction Ritual Chains*. Princeton: Princeton University Press; 2004.
- [7] Dreyfus H. Response to my critics. *Artificial Intelligence*. 1996; 80 (1).

DOI: 10.22363/2313-2272-2019-19-3-581-586

## Artificial intelligence as a challenge for the contemporary social analysis\*

**Books review: Boden M.A. *AI: Its Nature and Future*.**

Oxford: Oxford University Press, 2016. 156 p.;

**Collins H. *Artificial Intelligence:  
Against Humanity's Surrender to Computers*.**

Cambridge: Polity Press, 2018. 232 p.

---

\* © A.V. Rezaev, N.D. Tregubova, 2019.

The research was supported by the Russian Science Foundation. Project No. 18-18-00097.

The article was submitted on 06.05.2019. The article was accepted on 07.06.2019.