

УДК 004.8
ББК 32.813
Б88

Серия «Наука, идеи, ученые»

POSSIBLE MINDS
edited by John Brockman

Перевод с английского В. Желнинова

Серийное оформление и компьютерный дизайн Э. Кунтыш

Печатается с разрешения автора-составителя
и литературного агентства Brockman, Inc.

Б88 **Искусственный интеллект — надежды и опасения :**
[сборник : перевод с английского В. Желнинова] / под ред.
Джона Брокмана. — Москва : Издательство АСТ, 2020. —
384 с. — (Наука, идеи, ученые).

ISBN 978-5-17-115937-5

В далеком 1950 году американский математик, отец-основатель кибернетики и теории искусственного интеллекта Норберт Винер опубликовал работу «Человеческое применение человеческих существ» (в русском переводе – «Кибернетика и общество»), в которой выразил свои опасения, связанные с развитием искусственного интеллекта.

Сейчас, в XXI веке, проблема выглядит еще более злободневной. Наша компьютерная зависимость стала тотальной. Развлечения, покупки, работа, учеба – практически все сосредоточено в гаджетах размером с ладонь. Руководствуясь удобством и – что уж греха таить? – ленью, мы перекладываем на ИИ часть надоевших и скучных функций, а зачастую доверяем ему и принятие решений.

Пока, на волне эйфории от открывшихся перспектив, преимущества искусственного интеллекта кажутся неоспоримыми, но не получится ли так, что милые удобства, которые мы получили сейчас, в период «младенчества» искусственного интеллекта, обернутся крупными неприятностями, когда «младенец» повзрослеет и посмотрит на «родителей» критическим взглядом?

Руководствуясь формулой «кто предупрежден – тот вооружен», Джон Брокман предложил известным ученым, публицистам и философам поразмышлять о перспективах взаимодействия человека и искусственного интеллекта в свете идей, высказанных Винером, а также в свете новых реалий и последних достижений научной мысли.

УДК 004.8
ББК 32.813

ISBN 978-5-17-115937-5

© John Brockman, 2019
© Издание на русском языке AST Publishers, 2020

ОГЛАВЛЕНИЕ

Введение. О НАДЕЖДАХ И ОПАСНОСТЯХ ИИ.....	9
Глава 1. Сет Ллойд. ОШИБОЧНО, ЗАТО АКТУАЛЬНЕЕ, ЧЕМ КОГДА-ЛИБО <i>Именно в применимости кибернетической концепции к человеческим существам заключалась принципиальная ошибка Винера.....</i>	27
Глава 2. Джуда Перл. ОГРАНИЧЕНИЯ «НЕПРОЗРАЧНЫХ» ОБУЧАЕМЫХ МАШИН <i>Глубинное обучение обладает собственной динамикой, само себя выстраивает и оптимизирует и в большинстве случаев обеспечивает правильные результаты. Но, когда этого не происходит, вы не имеете ни малейшего понятия, почему это случилось и что с этим делать.....</i>	41
Глава 3. Стюарт Рассел. ЦЕЛЬ, ЗАЛОЖЕННАЯ В МАШИНУ <i>Мы можем столкнуться в будущем со сверхинтеллектуальными машинами, действия которых по определению невозможно будет предугадать, а их нечетко сформулированные цели будут конфликтовать с нашими собственными целями. И мотивацией этого сверхинтеллекта окажется стремление любыми способами обеспечить свое существование для достижения указанных целей.....</i>	49
Глава 4. Джордж Дайсон. ТРЕТИЙ ЗАКОН <i>Любая система, достаточно простая для понимания, не обретет сложности, достаточной для разумного поведения, тогда как любая система, достаточно сложная для разумного поведения, будет слишком сложной для понимания.....</i>	64

Глава 5.	Дэниел К. Деннет. ЧТО МЫ МОЖЕМ СДЕЛАТЬ? <i>Нам не нужны искусственные агенты познания. Нам нужны разумные машины.</i>	74
Глава 6.	Родни Брукс. НЕЧЕЛОВЕЧЕСКАЯ ЛОВУШКА, В КОТОРУЮ НАС ЗАВЛЕКЛИ МАШИНЫ <i>Сегодня ситуация намного сложнее, чем предвидел Винер, и меня не покидают опасения по поводу того, что она может сделаться фатальнее, чем ему когда-либо вообразалось.</i>	91
Глава 7.	Фрэнк Вильчек. ЕДИНСТВО РАЗУМА <i>Преимущества искусственного интеллекта перед естественным выглядят неоспоримыми, тогда как преимущества естественного интеллекта перед искусственным, сегодня очевидные, кажутся преходящими</i> . . .	103
Глава 8.	Макс Тегмарк. ДАВАЙТЕ ПОСТАВИМ БОЛЕЕ ВЫСОКУЮ ЦЕЛЬ, ЧЕМ СВАЛКА ИСТОРИИ <i>Необходимо проанализировать потенциальные угрозы создания ИИ, чтобы не допустить их возникновения.</i>	117
Глава 9.	Яан Таллинн. ДИССИДЕНТСКИЕ ПОСЛАНИЯ <i>Устойчивое развитие ИИ способно породить перемены космических масштабов — и этот глобальный процесс убьет все живое</i>	131
Глава 10.	Стивен Пинкер. ТЕХНОЛОГИЧЕСКИЕ ПРОРОЧЕСТВА И НЕДООЦЕНЕННАЯ КАУЗАЛЬНАЯ СИЛА ИДЕЙ <i>Нет такого закона комплексных систем, который гласил бы, что разумные агенты непременно превращаются в безжалостных мегаломаньяков.</i>	146
Глава 11.	Дэвид Дойч. НАГРАДЫ И НАКАЗАНИЯ — ЧТО ДАЛЬШЕ? <i>Ошибочные представления о человеческом мышлении и происхождении человечества порождают столь же ошибочные представления об ИИ и способах его создания</i>	163
Глава 12.	Том Гриффитс. ПРИМЕНЕНИЕ МОДЕЛЕЙ ЧЕЛОВЕЧЕСКОГО ПОВЕДЕНИЯ В ИИ <i>Автоматические интеллектуальные системы, способные на корректные высказывания о человеческих желаниях, должны обладать надежными генеративными моделями человеческого поведения</i>	178
Глава 13.	Анка Драган. ВВЕСТИ ЧЕЛОВЕКА В УРАВНЕНИЕ ИИ <i>В реальном мире ИИ должен взаимодействовать с людьми и размышлять об этом. Людям придется каким-то образом формализовывать задачи для ИИ.</i>	189

- Глава 14. Крис Андерсон. СПУСК ПО ГРАДИЕНТУ
Из того, что системы ИИ порой демонстрируют локальные минимумы возможностей, не стоит делать вывод, что они не слишком жизнеспособны. Люди — возможно, все формы жизни в целом — тоже часто остаются на локальных минимумах 200
- Глава 15. Дэвид Кайзер. «ИНФОРМАЦИЯ» ПО ВИНЕРУ, ПО ШЕННОНУ И ПО НАШИМ НЫНЕШНИМ ПРЕДСТАВЛЕНИЯМ
Многие ключевые доводы «Человеческого применения человеческих существ» относятся скорее к девятнадцатому столетию, чем к двадцать первому. Винер, похоже, не совсем разобрался в теории Шеннона об информации как состоящей из неделимых, лишенных значения битов 210
- Глава 16. Нил Гершенфельд. МАСШТАБИРОВАНИЕ
Хотя машинное производство и машинное мышление как будто не соотносятся друг с другом, в действительности развитие первого неразрывно связано с развитием второго и наоборот 221
- Глава 17. У. Дэниел (Дэнни) Хиллис. ПЕРВЫЙ МАШИННЫЙ ИНТЕЛЛЕКТ
Гибридные сверхинтеллекты — например, национальные государства и корпорации, обладают собственными насущными целями, их действия далеко не всегда служат интересам тех людей, которые их создали 234
- Глава 18. Венки Рамакришнан. СТАНУТ ЛИ КОМПЬЮТЕРЫ НАШИМИ ПОВЕЛИТЕЛЯМИ?
Наши страхи по поводу ИИ отражают веру в то, что именно разум делает нас особенными 248
- Глава 19. Алекс (Сэнди) Пентленд. ЧЕЛОВЕЧЕСКАЯ СТРАТЕГИЯ
Как можно создать экосистему успешного взаимодействия человека и машины, нечто отличное от машинного общества, киберкультуру, в которой все мы останемся людьми, культурой с «человеческим лицом»? 262
- Глава 20. Ханс-Ульрих Обрист. СДЕЛАТЬ НЕЗРИМОЕ ЗРИМЫМ: ИИ И ИСКУССТВО
Многие современные художники высказывают разнообразные сомнения относительно перспектив ИИ и напоминают нам, что не нужно приписывать термину «искусственный интеллект» исключительно позитивные характеристики ... 280
- Глава 21. Элисон Гопник. ИИ ПРОТИВ ЧЕТЫРЕХЛЕТОК
Изучение поведения детей может снабдить программистов полезными сведениями для разработки методов компьютерного обучения 296

- Глава 22. Питер Гэлисон. МЕЧТЫ «АЛГОРИСТОВ»
ОБ ОБЪЕКТИВНОСТИ
К настоящему времени все юридические, этические, формальные и экономические признаки алгоритмов являются квазиинфинитивными..... 312
- Глава 23. Джордж М. Черч. ПРАВА МАШИН
Быть может, стоит меньше переживать о «нас-против-них», зато больше задумываться о правах всех разумных существ в условиях формирующегося беспрецедентного разнообразия умов 323
- Глава 24. Кэролайн А. Джонс. ХУДОЖЕСТВЕННОЕ ПРИМЕНЕНИЕ КИБЕРНЕТИЧЕСКИХ СУЩЕСТВ
Деятельность кибернетически предрасположенных художников затрагивает формирующееся поведение, которое ускользает от ИИ в его нынешнем воплощении...... 343
- Глава 25. Стивен Вольфрам. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И БУДУЩЕЕ ЦИВИЛИЗАЦИИ
Грандиозный прорыв случится тогда, когда мы наконец обретем подлинное бессмертие. Будет ли это достигнуто биологически или кибернетически — остается неясным, но рано или поздно так наверняка случится..... 359

Искусственный интеллект — это сегодняшний день, это, так сказать, история в основе прочих историй. Это одновременно Второе пришествие и Апокалипсис, добрый ИИ против злого ИИ. Книга, которую вы держите в руках, продолжает разговор об искусственном интеллекте; в нее вошли соображения целого ряда влиятельных персон из мира ИИ и сопредельных миров о том, что такое искусственный интеллект и что за ним скрывается. Проект, получивший название «Глубокое осмысление», начался по-настоящему в сентябре 2016 года, когда некоторые авторы книги собрались, чтобы обсудить упомянутые вопросы в отеле «Мэйфлауэр Грейс» в Вашингтоне, штат Коннектикут.

Очень быстро после этой первой встречи стало ясно, что присущие широкой культуре упования и опасения, связанные с ИИ, во многом аналогичны тем надеждам и тревогам, которые окружали идеи Норберта Винера относительно кибернетики, особенно в 1960-е годы, когда творцы принялись обсуждать новые технологии в своих произведениях. Я был непосредственным свидетелем распространения этих идей; на самом деле не будет преувеличением сказать, что именно они определили мой жизненный путь. С наступлением цифровой эры в начале 1970-х годов о Винере как будто позабыли, но сегодня его кибернетические идеи успели проникнуть на-

столько глубоко, оказались усвоенными настолько полно, что им больше не требуется отдельное обозначение. Они повсюду, и потому будет логично начать с них.

Новые технологии = новые восприятия

До ИИ была кибернетика — представление об автоматическом и саморегулируемом управлении, изложенная в основополагающем тексте Норберта Винера 1948 года. Я могу датировать собственное осознанное обращение к ней 1966 годом, когда композитор Джон Кейдж пригласил меня и еще четверых или пятерых молодых людей из области искусства присоединиться к участию в цикле семинаров, посвященных средствам массовой информации, коммуникациям, искусству, музыке и философии; эти семинары отражали интерес Кейджа к идеям Винера, Клода Шеннона и Маршалла Маклюэна, каждый из которых считался авторитетом в нью-йоркских артистических кругах, где я тогда вращался. В частности, Кейдж высоко оценил утверждение Маклюэна о том, будто, изобретая электронные технологии, мы экстернализируем свою центральную нервную систему, то есть наш разум, и отныне приходится допускать, что «существует один-единственный разум, общий для всех людей».

Подобного рода идеи все чаще привлекали творцов, с которыми мне доводилось тогда сотрудничать (я был программным менеджером нью-йоркской «Режиссерской синемаатеки» и отвечал за серию мультимедийных постановок под названием «Нью Синема 1», также известную как «Большой кинофестиваль», под руководством авангардного режиссера и импресарио Йонаса Мекаса). Среди этих творцов были визуальные художники Клас Олденбург, Роберт Раушенберг, Энди Уорхол, Роберт Уитман; художники-кинематики Шарлотта Мурман и Нам Джун Пайк; мастера хэппенинга Аллан Капроу и Кэроли Шнеманн; танцовщица Триша Браун; режис-

серы Джек Смит, Стэн Вандербик, Эд Эмшуиллер и братья Кучар; авангардный драматург Кен Дьюи; поэт Герд Стерн и группа USCO¹; музыканты-минималисты Ламонт Янг и Терри Райли; а еще Уорхол заманил к нам музыкальную группу «Velvet Underground». Многие из перечисленных людей читали Винера и вели разговоры на темы кибернетики. На одном из семинаров Кейдж сунул руку в портфель, достал экземпляр «Кибернетики» и вручил мне со словами: «Это вам».

В разгар фестиваля мне неожиданно позвонил коллега Винера Артур К. Соломон, руководитель аспирантской программы по биофизике в Гарварде. Винер умер годом ранее, другие близкие коллеги Соломона и Винера в Массачусетском технологическом институте и Гарварде узнали о «Большом кинофестивале» из статей в «Нью-Йорк таймс» — и заинтересовались этими попытками интерпретировать идеи Винера. Соломон предложил мне привезти несколько творцов в Кембридж² и встретиться с ним и группой ученых, в которую входили исследователь сенсорных коммуникаций из МТИ Уолтер Розенблит, математик из Гарварда Энтони Эттингер и инженер МТИ Гарольд (Док) Эджертон, изобретатель стробоскопа.

Как и во многих других ситуациях «соприкосновения искусства с наукой», в которых я оказывался с тех пор, двухдневная встреча превратилась, как говорится, в ожидаемый провал: мы говорили на разных языках. Но я несколько не отчаивался, да и само событие имело ряд любопытных последствий, например именно тогда нам довелось впервые увидеть воочию «компьютер». Тогда компьютеры

¹ Американская арт-группа, наибольшая активность которой пришлась на середину 1960-х гг. Аббревиатура USCO расшифровывается как «Наша компания» (Us Company). Группа экспериментировала с «дополненным» кино, визуализацией музыки, инсталляциями и пр. — *Здесь и далее, кроме особо оговоренных случаев, примеч. ред.*

² Имеется в виду университетский центр в американском штате Массачусетс.

были редкостью; во всяком случае, никто из нас прежде их не видел. Благодаря любезному приглашению мы очутились в каком-то обширном пространстве посреди кампуса МТИ и узрели «холодную комнату», приподнятую над полом и заключенную в стекло; за этой стеклянной оградой техники в белых лабораторных халатах, шапках и перчатках деловито собирали перфокарты, извергавшиеся из огромного агрегата. Когда я подошел ближе, от моего дыхания стекло запотело. Я протер стекло, увидел тот «компьютер» — и влюбился с первого взгляда.

Позже, осенью 1967 года, я отправился в Менло-парк, чтобы встретиться со Стюартом Брэндом, с которым познакомился в Нью-Йорке в 1965 году, когда он числился в составе художественной группы USCO. Ныне, вместе со своей женой, математиком Лоис, он готовил к публикации первое издание «Каталога всей Земли». Пока Лоис и остальная команда упорно вносили последние правки в «Каталог», мы со Стюартом два дня подряд отсиживались в уголке, штудировав, обсуждая и комментируя тот самый экземпляр «Кибернетики» Винера в мягкой обложке, который Кейдж вручил мне год назад.

Вдохновленный совокупностью идей, я начал разрабатывать тему, своего рода мантру, которая стимулировала все мои дальнейшие усилия: «Новые технологии суть новое восприятие». Опираясь на мысли теоретика коммуникаций Маршалла Маклюэна, архитектора и дизайнера Бакминстера Фуллера, футуролога Джона Макхейла и антропологов Эдварда (Неда) Т. Холла и Эдмунда Карпентера, я начал активно начитывать работы по теории информации, кибернетике и теории систем. Маклюэн предложил мне прочесть исследование Дж.З. Янга «Сомнение и уверенность в биологической науке», где утверждалось, что мы создаем инструменты и формируем себя посредством их использования. Другим рекомендованным текстом стала статья Уоррена Уивера и Кло-

да Шеннона «Недавние дополнения к математической теории коммуникации» (1949), которая начинается так: «Слово «коммуникация» далее будет употребляться в чрезвычайно широком смысле, дабы охватить все процедуры и способы, с помощью которых один разум может влиять на другой. Разумеется, мы говорим не только об устной и письменной речи, но также о музыке, изобразительном искусстве, театре, балете — фактически обо всем человеческом поведении».

Кто мог знать, что всего два десятилетия спустя мы начнем трактовать человеческий мозг как компьютер? А в следующие два десятилетия, объединяя наши компьютеры в сеть интернет, начнем понимать, что человеческий мозг — не сверхмощный компьютер, но совокупность, сеть компьютеров? Об этом наверняка даже не догадывался Винер, специалист по аналоговым цепям обратной связи, предназначенным для управления машинами, как не догадывались и создатели машин.

***«Мы должны перестать лобызать плетку,
которой нас хлещут»***

Через два года после выхода «Кибернетики», в 1950 году, Норберт Винер опубликовал работу «Человеческое применение человеческих существ», безусловно более философскую, в которой выразил свою обеспокоенность относительно безудержной коммерческой эксплуатации и других непредвиденных последствий новых технологий управления. Я не читал «Человеческое применение человеческих существ» до весны 2016 года, когда взял в руки экземпляр ее первого издания, стоявший в моей библиотеке рядом с «Кибернетикой». Меня откровенно шокировало осознание того, насколько прозорлив оказался Винер в 1950 году по поводу наших дней. Первое издание этой работы стало бестселлером — более того, положило начало широкому общественному обсуждению

этих проблем, — но под давлением коллег Винер выпустил в 1954 году пересмотренное и «смягченное» издание, из которого исчезла заключительная глава «Голоса ригидности».

Историк науки Джордж Дайсон отмечает, что в давно забытом первом издании работы Винер предсказывал возможность «возникновения нового грозного фашизма, который будет опираться на *machine a gouverner*¹»:

Ни одна элита не избежала его критики, от марксистов и иезуитов («весь католицизм по сути представляет собой тоталитарную религию») до ФБР («наши магнаты изучили пропагандистские техники русских и обнаружили, что это хорошо и полезно») и финансистов, которые стремятся «сделать американский капитализм и пятую свободу бизнесмена высшими ценностями во всем мире». Ученые... удостоились столь же пристального внимания и сравнения с церковью: «Действительно, руководители больших лабораторий очень схожи с епископами, обладают обширными связями с могущественными людьми во всех сферах жизни и, к сожалению, одержимы смертным грехом гордыни и жадной власти.

Эта иеремиада² дорого обошлась Винеру. Как пишет Дайсон:

«Этот набат предпочли проигнорировать, и не потому, что Винер ошибался относительно цифровых вычислений, а потому, что период завершения рукописи осенью 1949 года совпал с появлением более серьезных угроз. Винер не имел ничего против цифровых вычислений, зато категорически выступал против ядерного оружия и отказался присоединиться к тем, кто создавал цифровые компьютеры ради конструирования в тысячу раз более мощной водородной бомбы».

¹ Управляющую машину (*фр.*). Все дальнейшие цитаты из этой работы Винера приводятся по изданию: *Винер Н.* Кибернетика и общество. Человеческое применение человеческих существ. М.: АСТ, 2019.

² Сетование на пороки общества и их обличение (по образу и подобию упреков, которые бросал своим современникам библейский пророк Иеремия).

Поскольку «Человеческое применение человеческих существ» уже давно не переиздавалось, мы остаемся в неведении по поводу винеровского *cri de coeur*¹, еще более актуального сегодня, чем шестьдесят восемь лет назад: «Мы должны перестать лобызать плетку, которой нас хлещут».

Разум, мышление, интеллект

Среди причин, по которым мы сегодня мало что слышим о «Кибернетике», выделяются две основные: во-первых, хотя «Человеческое применение человеческих существ» в свое время считалось важной работой, оно противоречило чаяниям многих коллег Винера, в том числе Джона фон Неймана и Клода Шеннона, которые были заинтересованы в коммерциализации новых технологий. Во-вторых, пионер компьютерных исследований Джон Маккарти не одобрял взгляды Винера и отказывался использовать винеровский термин «кибернетика». Он придумал термин «искусственный интеллект» — и стал, скажем так, отцом-основателем этой области науки.

Как объяснил мне Джуда Перл, который в 1980-е годы разработал новый подход к искусственному интеллекту под названием «байесовская сеть»:

Винер на самом деле породил воодушевляющие надежды на то, что однажды мы сможем сконструировать разумную машину. Он не был специалистом по компьютерам. Он рассуждал об обратной связи, о коммуникации, об аналоговых механизмах. Рабочей метафорой была для него схема обратной связи, область, в которой он являлся экспертом. На заре цифровой эпохи, в начале 1960-х годов, люди желали обсуждать программирование, говорить о кодах, вычислительных функциях, кратковременной памяти и долговременной памяти — таковы были тогдашние зна-

¹ Крика души (*фр.*).