

Сергей Шумский

# ВОСПИТАНИЕ

новая история разума

# МАШИНА

будущая  
человеко-  
машинная  
цивилизация

социальные  
проблемы  
цифровой  
революции

эволюция  
цифровых  
технологий

**АНО**  
альпина нон-фикшн



# Предисловие научного редактора

## Неизбежность сосуществования человеческого и машинного интеллекта

За несколько последних лет тема искусственного интеллекта (ИИ) приобрела особую популярность. О внедрении ИИ в практику бизнеса сообщают Amazon, Сбербанк, Tesla, РЖД и буквально сотни других компаний, крупных и небольших. Однако, как это нередко бывает, быстрый рост популярности какого-то понятия часто приводит к возникновению недоразумений, связанных с неточным (а иногда и ошибочным) его пониманием. Именно это и происходит с концепцией искусственного интеллекта. Сложные аппаратно-программные комплексы многим представляются примерами использования ИИ как новой технологии, хотя чаще всего они являются просто более сложными вариантами привычных для нас гаджетов.

Такая ситуация может породить ощущение, что ИИ — всего лишь очередной хайп, интерес к которому со временем пройдет. Но так ли это? К сожалению, многие недавно вышедшие работы об искусственном интеллекте не дают ясного ответа на этот важный вопрос. Вы найдете на него ответ в книге, которую держите в руках.

Ее автор Сергей Шумский — российский эксперт, много лет занимающийся проблемой машинного обучения и ИИ, — поставил перед собой сложную задачу. На основе анализа прогресса технологий в разные эпохи он решил показать неизбежность использования машинного интеллекта в наши дни. С поставленной задачей автор справился блестяще. Простые, но убедительные количественные оценки показывают причины перехода от одного экономического уклада к другому. Несложные расчеты, относящиеся к операциям с информацией, поясняют, почему именно сейчас исследования в области искусственного интеллекта стремительно прогрессируют.

Как специалист, занимающийся практическими разработками в сфере ИИ, автор хорошо понимает не только масштабы уже достигнутых успехов, но и сложность задач, стоящих перед учеными и разработчиками на пути к созданию так называемого сильного искусственного интеллекта. В книге

анализируются подходы к решению этой амбициозной задачи, и перспективы, которые описывает автор, вызывают определенный оптимизм.

В научной и научно-популярной литературе широко обсуждаются социальные последствия распространения технологий ИИ, в особенности тех, которые могут возникнуть после создания сильного искусственного интеллекта. Многие авторы рисуют почти апокалиптическую картину порабощения людей роботами, оснащенными ИИ. В этой книге такие проблемы обсуждаются, но в гораздо более спокойной манере. По мнению автора, при правильном (проактивном, или упреждающем) подходе человечеству вполне по силам справиться с этими проблемами. Для этого, правда, людям придется напрячь собственный интеллект и сформулировать этические принципы, которыми следует руководствоваться при разработке и использовании ИИ.

Особый интерес представляют заключительные главы книги, в которых дан набросок программы создания сильного искусственного интеллекта, а также приводятся оценки сроков реализации этой программы. По мнению автора, 40 лет — реальный временной горизонт создания сверхчеловеческого машинного интеллекта, способного коренным образом изменить жизнь людей. Иными словами, у многих потенциальных читателей этой книги есть реальный шанс стать свидетелями и участниками интенсивного взаимодействия человеческого и машинного интеллекта. Чтение этой книги поможет подготовиться, психологически и ментально, к новым реалиям.

*Доктор физико-математических наук,  
профессор С. Р. Филонович*

## От автора

В 2019 году вышла в свет моя монография «Машинный интеллект» для специалистов в области искусственного интеллекта (ИИ). Отлично понимая, что каждая формула в книге снижает количество ее читателей вдвое, я заранее смирился с тем, что одолеть ее сможет лишь ограниченный круг профессионалов.

Между тем вопросы, связанные с пугающими темпами развития ИИ и перспективами робототехники, волнуют не только специалистов. Так или иначе это коснется каждого из нас, причем в не столь уж и отдаленном будущем. Естественно, мне хотелось поделиться современными знаниями и гипотезами о зарождающейся цивилизации машинного интеллекта и с широкой аудиторией.

К счастью, весной 2019 года я получил предложение подготовить мини-курс по технологиям цифровой экономики для Высшей школы менеджмента ВШЭ. В процессе подготовки такого мини-курса родилась и эта книга о социальных последствиях использования искусственного интеллекта. В этой связи я выражаю искреннюю признательность руководителям ВШМ ВШЭ Сергею Ростиславовичу Филоновичу и Льву Владимировичу Татарченко за моральную и материальную поддержку при подготовке рукописи.

Надеюсь, книга поможет читателям лучше разобраться в сути происходящих в современном мире перемен и наглядно представить себе, как изменится жизнь людей в ближайшие десятилетия.

# Введение

*Нам предстоит разговор о различных мыслимых аспектах цивилизации, аспектах, которые можно вывести из предпосылок, известных уже сегодня.*

СТАНИСЛАВ ЛЕМ

*Чем дольше вы смотрите назад, тем дальше видите вперед.*

УИНСТОН ЧЕРЧИЛЛЬ

Главное отличие этой книги от многих других — попытка рассмотреть искусственный интеллект в глубоком историческом контексте, не ограничиваясь историей последних десятилетий. Следуя Спинозе, мы пытаемся вписать современный момент в ту цепь событий, внутри которой он возникает с необходимостью, а не случайно. Нет лучшего способа заглянуть в будущее, чем как можно основательнее разобраться в прошлом.

Вот и мы в этой книге постараемся выяснить, каковы движущие силы развития технологий и как они определяют ход технологических революций, одну из которых мы как раз и переживаем в настоящий момент. Книга рассчитана на всех, кто интересуется будущим. При этом от читателя не требуется никаких специальных знаний. Объем книги намеренно ограничен, с тем чтобы ее можно было прочесть за пару-тройку вечеров, не теряя логическую нить рассуждений автора.

Основная мысль этой книги в том, что машинный интеллект, когда он появится, неизбежно станет частью общей человеко-машинной цивилизации и унаследует от нее наши человеческие ценности. Так что даже в будущей цивилизации машинного интеллекта, когда интеллект машин далеко превзойдет человеческий уровень, они по-прежнему смогут действовать в наших интересах, а не преследовать чуждые нам цели, как какой-то чужеродный инопланетный разум. Правда, за такой благоприятный для людей сценарий нам придется побороться — создать не просто искусственный интеллект, а дружественный ИИ, и тому, как этого добиться, в книге будет уделено особое внимание.

Но прежде мы должны убедить скептиков в том, что нам вообще следует

заниматься этой опасной субстанцией — искусственным разумом. Не проще ли отложить на время исследования и разработки и подождать, пока человечество «созреет» для таких вопросов? Наш ответ на этот аргумент из серии «Не давайте спички детям» состоит в том, что без создания искусственного разума в ближайшие годы человечество просто не сможет обойтись. Индустриальная цивилизация достигла пределов своего развития. Обеспечить приличное существование всему населению Земли (а не только «золотому миллиарду») имеющиеся технологии не в состоянии: биосфера просто не выдержит такой нагрузки. Назревающие экономические проблемы способны разрешить только новые цифровые технологии, главным образом искусственный интеллект. Мы приведем конкретные аргументы, за счет чего и как именно ИИ сможет поднять эффективность мировой экономики.

Но так ли уж драматична на самом деле сегодняшняя экономическая ситуация? Может быть, нынешний глобальный кризис — не структурный, а циклический и все само собой перемелется? Чтобы убедить читателя в том, что мир действительно находится сегодня в одном из переломных исторических моментов, мы начнем с краткого экскурса в историю человеческой цивилизации, что может показаться странным для книги, посвященной машинному интеллекту. Однако автор считает крайне поучительным представить нынешний кризис в одном ряду с великими кризисами прошлого, чтобы убедить читателя, что любой технологический уклад в какой-то момент достигает своих пределов и соответствующий кризис можно разрешить, только перейдя на новый пакет технологий.

Вот почему в главе 1 мы начинаем издали — с краткого введения в историю технологических укладов: от цивилизации охотников-собирателей через аграрную и промышленную революции к современному обществу, одной ногой уже шагнувшему в новый цифровой уклад. Для каждого уклада мы выведем «на кончике пера» предельное число людей, которое способен прокормить его технологический пакет, убедившись заодно и в том, что предел населения для индустриального уклада уже практически достигнут.

Технологический пакет представляет собой сумму накопленных цивилизацией знаний. Причем, как мы покажем, они увеличиваются гораздо быстрее численности населения (как квадрат количества людей). Поскольку суммарный объем человеческой памяти растет линейно с числом людей, неизбежен момент, когда сложность цивилизации превысит ограниченные возможности человеческого мозга. По мысли автора, именно такой момент мы сегодня и наблюдаем. Появление компьютеров, а затем и искусственного интеллекта есть закономерное следствие этого простого факта.

Таким образом, в главе 1 мы проследим в общих чертах логику развития человеческой цивилизации от древних людей до наших дней, обозначив базовые параметры цивилизации на каждой ступени ее развития. Мы покажем, за счет чего революция позднего палеолита увеличила в конечном итоге население Земли в 100 раз, а последующая аграрная революция — еще в 100 раз. Наконец, почему индустриальное общество не способно прокормить более 10 млрд человек, так как достигло своего *предела сложности*. Логика дальнейшего развития тоже прояснится: при новом цифровом укладе нам предстоит преодолеть барьер сложности современной цивилизации за счет создания и «приручения» искусственного интеллекта. Только таким образом она сможет развиваться дальше.

В главе 2 мы наконец доберемся до ИИ как закономерного этапа эволюции цифровой техники, созревания нового уклада в недрах предыдущего. Мы проследим, как формировался современный стек цифровых технологий и почему ИИ объективно не мог появиться вплоть до недавнего времени.

Ключом к пониманию цифровой революции является обобщенный закон Мура — рост компьютерных мощностей в 100 раз каждые 10 лет. До 2010-х годов суммарные компьютерные мощности были еще пренебрежимо малы по сравнению с вычислительными возможностями человеческого мышления. Этим объясняется известный парадокс, что до сих пор экспоненциальное развитие вычислительной техники практически не сказывалось на темпах роста производительности труда в большинстве отраслей экономики.

Однако за последние 10 лет ситуация качественно изменилась. Мы покажем, что сегодня совокупная вычислительная мощность компьютеров уже превосходит таковую у людей и представляет собой тот единственный обильный и дешевый ресурс, за счет которого можно повысить эффективность мировой экономики. Но для этого необходимо научить компьютеры делать человеческую работу лучше и дешевле людей: ни много ни мало создать искусственный интеллект. Таким образом, ИИ является *замыкающей технологией*, открывающей возможность перехода к новому технологическому укладу.

Так называемая революция глубокого обучения 2010-х является ответом на этот вызов. Сегодня мы наблюдаем появление первых несовершенных пока версий *слабого ИИ*, которые можно уподобить несовершенным поначалу моделям паровых двигателей времен первой промышленной революции. Однако, как мы помним, именно благодаря паровым двигателям аграрная цивилизация стала замещаться индустриальной. Аналогично современный слабый ИИ знаменует собой начало перехода к

новому цифровому укладу. Однако закончиться он может лишь с появлением *сильного ИИ*, повсеместно доступного «интеллекта из розетки», коллективного сетевого мышления искусственных личностей, аналогичного замене паровых двигателей на общедоступную «энергию из розетки» в ходе второй промышленной революции.

В главе 3 мы обратимся к экономике цифрового уклада. За счет чего она будет в разы эффективнее индустриальной? Почему в ней станет производиться больше благ с меньшими затратами? Какие формы организации бизнеса придут на смену транснациональным компаниям индустриального уклада?

Ключевым понятием здесь будет экономика «длинного хвоста» — существенное расширение рынка товаров и услуг при переходе от массового к персональному потреблению и производству. Торжество такой экономики мы наблюдаем воочию, по мере того как YouTube отбирает аудиторию у телевидения, а Amazon — покупателей у традиционных магазинов одновременно с ростом рынка мобильных приложений на платформах Apple, Microsoft и Google. Эти цифровые платформы и являются новыми ячейками будущей цифровой экономики. Именно они предъявляют сегодня основной спрос на технологии искусственного интеллекта и служат драйверами его развития — от слабого к сильному ИИ.

Но прежде чем обратиться к перспективам развития ИИ, в главе 4 мы обсудим социальные проблемы и риски, порожденные резкой сменой технологического уклада, ведь скорость его развития определяют не только те, кто за него, но и те, кто против. Анализ рисков позволит нам сформулировать «техническое задание» на создание безопасного ИИ, необходимого для выживания нашей цивилизации. Мы последовательно рассмотрим проблемы, связанные с грядущей массовой безработицей, нарастанием социальной напряженности и опасностью большой войны по инициативе искусственных стратегов, а также вопрос о том, останемся ли мы хозяевами своей судьбы, когда искусственный разум наконец превзойдет человеческий. Все эти риски вполне реальны, и проблемы, которые нам предстоит разрешить, нетривиальные. Они не сводятся к чисто техническим решениям, а подразумевают активное участие граждан в создании согласованной концепции общественной безопасности, разработку и внедрение национальных платформ цифровой демократии.

С технической точки зрения дружелюбность искусственного интеллекта должна быть встроена в операционную систему роботов, как были встроены в них «три закона робототехники» Айзека Азимова. А в дополнение к правильным «социальным инстинктам» будущей искусственной психики должны быть созданы и механизмы взаимного



контроля всех членов социума, включая агентов и роботов, подобно тому как технология блокчейн автоматически обеспечивает взаимный контроль транзакций в сети. Цивилизация машинного интеллекта обязана иметь систему встроенной коллективной безопасности по аналогии с новейшими ядерными реакторами, в которые безопасность встроена на физическом уровне.

После того как мы разобрали, кому и зачем нужен сильный ИИ, а также каким требованиям он должен удовлетворять, чтобы улучшить, а не ухудшить нашу жизнь, переходим к конкретным планам по его разработке. А именно: в главе 5 мы представим замысел создания искусственной психики роботов «по образу и подобию» человеческой, с тем чтобы их интеллект был по своей архитектуре как можно ближе к нашему. Мы определим методологию разработки сильного ИИ — обратную инженерию архитектуры мозга — и сформулируем ее цель: создание искусственной психики в качестве будущей операционной системы роботов. В довершение покажем, как может выглядеть цифровая платформа массовой робототехники со встроенной системой коллективной безопасности.

Наконец, в главе 6 мы сможем заглянуть в будущее с разной степенью уверенности на разных временных горизонтах, начиная с ближайших 5–10 до 20–40 лет — предполагаемого времени появления ИИ сверхчеловеческого уровня. Самое удивительное в этой картине — как разительно будет отличаться от современности мир наших детей и внуков. Но именно такое ускорение исторического времени вызывает разворачивающаяся на наших глазах революция в области искусственного интеллекта.

# ГЛАВА 1

## Технологические уклады и технологические революции

*Мы стоим на грани перемен, сравнимой с возникновением на Земле человека.*

ВЕРНОР ВИНДЖ

*Может быть, высокоразвитая цивилизация — это вовсе не огромная энергия, а наилучшее регулирование?*

СТАНИСЛАВ ЛЕМ

### Понятие технологических укладов

Наше изложение мы начнем с определения *технологического уклада*, задающего базовые параметры цивилизации на разных стадиях ее развития. Нас будет интересовать, как набор технологий, освоенных цивилизацией, определял на разных этапах два ее важнейших параметра — число людей, которых она может прокормить, и ее сложность, то есть количество накопленных знаний. Многие будут зависеть от того, каким образом связаны между собой эти два ключевых параметра.

Технологические уклады стабильны и способны воспроизводить себя в течение длительного времени, ведь технологии зависят друг от друга, образуя в каждый момент целостный *пакет* взаимодополняющих друг друга технологий. Этот *технологический пакет* регулируется отрицательными обратными связями в обществе, поддерживающими его устойчивость, и любые новые технологии, нарушающие существующее равновесие, встречают естественное сопротивление со стороны общества.

Смена технологических укладов происходит скачками, потому что новые технологии созревают в недрах старого уклада «тайком», в относительно небольших нишах, в которых они поодиночке не способны реализовать весь свой потенциал. Такая возможность у них появится лишь тогда, когда они смогут образовать новый целостный технологический пакет, более

жизнеспособный, чем существующий. Такая ситуация возникает лишь с появлением *закрывающей технологии* — последней, необходимой для формирования нового целостного пакета. Возникновение такой технологии нарушает устойчивость существующей системы, какие-то из ее обратных связей становятся положительными, и система скачком переходит в новое устойчивое состояние. После такой *технологической революции* снова наступает эпоха плавного эволюционного развития.

Технологические революции, как правило, сопровождаются социальными, так как смена материальных основ жизни общества не может не сказаться на его социальной надстройке — образе жизни людей и регулирующих человеческие отношения общественных институтах (общепринятых нормах поведения, убеждениях и организациях)<sup>1</sup>.

Технологии играют определяющую роль в жизни людей потому, что в совокупности позволяют осваивать необходимые для нее ресурсы, например энергетические. Какой-то из них обычно становится *базовым ресурсом*, обеспечивающим долговременное развитие. Обычно новый технологический пакет открывает перед цивилизацией новые перспективы за счет появления возможности эксплуатации нового базового ресурса, более дешевого и обильного, чем прежний.

Итак, резюмируем сказанное. Технологический уклад складывается из трех компонент:

- дешевого и обильного базового ресурса, обеспечивающего долгосрочное развитие;
- технологического пакета, позволяющего эффективно эксплуатировать базовый ресурс;
- организации жизни людей, способной поддерживать данный технологический пакет.

Далее в этой главе мы напомним читателям краткую историю смены технологических укладов на протяжении всей истории человечества. При этом мы сосредоточимся именно на эволюции технологий — материальной основы нашей цивилизации, чтобы вписать современную цифровую революцию в этот исторический процесс.

Экономисты обычно ассоциируют уклады с длинными циклами Кондратьева<sup>2</sup>, начиная их отсчет с первой промышленной революции<sup>3</sup>. Мы же здесь стремимся охватить весь период человеческой цивилизации, поскольку считаем вслед за Тоффлером<sup>4</sup>, что нынешняя революция по масштабу сопоставима с наиболее крупными переменами в истории человечества.

## Энергетические потребности человека

Для количественных оценок энерговооруженности цивилизации, от которой зависит наше благосостояние, полезно знать опорные цифры.

Среднему человеку, ведущему активный образ жизни, требуется около 3000 ккал в день [1]. Мы потребляем их с пищей в виде белков, жиров и углеводов с различной энергетической ценностью. Наиболее богаты энергией жиры (9500 ккал/кг), затем следуют углеводы (3500 ккал/кг) и белки (2400 ккал/кг). Без последних тем не менее не обойтись, так как именно они являются основным строительным материалом наших тел и именно белковые молекулярные машины обеспечивают извлечение из поступающего топлива полезной свободной энергии, питающей силу наших мускулов. Усредненный КПД нашей биологической «тепловой машины» — 10–15%, остальные 90–85% выделяются в виде тепла (см. рис. 1).

Примерно те же соотношения справедливы и для животных, которые на протяжении большей части нашей истории поставляли нам энергию для производства и транспорта (быки и лошади). В частности, лошадь примерно на порядок мощнее человека (735 Вт против 60 Вт механической энергии). По традиции мощность машин до сих пор измеряется в лошадиных силах [2]. Так, мощность типового современного автомобиля среднего класса составляет около 120 л. с. В США на человека приходится 0,8 машины, или порядка 100 лошадей, при том что на протяжении веков люди довольствовались 0,2 лошади на человека<sup>5</sup> — настолько различны энергетические мощности современной и аграрной цивилизаций.

### БАЗОВЫЕ ФАКТЫ

- Потребление энергии
  - Человек: 3000 ккал/день (145 Вт)
  - Лошадь: 45 000 ккал/день (2,2 кВт)
- Еда
  - Мясо: 2400 ккал (10 МДж)/кг
  - Зерно: 3500 ккал (15 МДж)/кг
  - Жиры: 9500 ккал (40 МДж)/кг
- Топливо
  - Дрова: 3500 ккал (15 МДж)/кг
  - Уголь: 7000 ккал (30 МДж)/кг
  - Нефть: 9500 ккал (40 МДж)/кг



Рис. 1. Потребности человека в энергии и ее источники

Этим мы обязаны переходу на ископаемое органическое топливо, энергетические характеристики которого сравнимы с компонентами нашей пищи<sup>[3]</sup>, но объемы их потребления на порядки больше: более 10 т условного топлива против 0,3 т условной пищи на человека в год (в США).

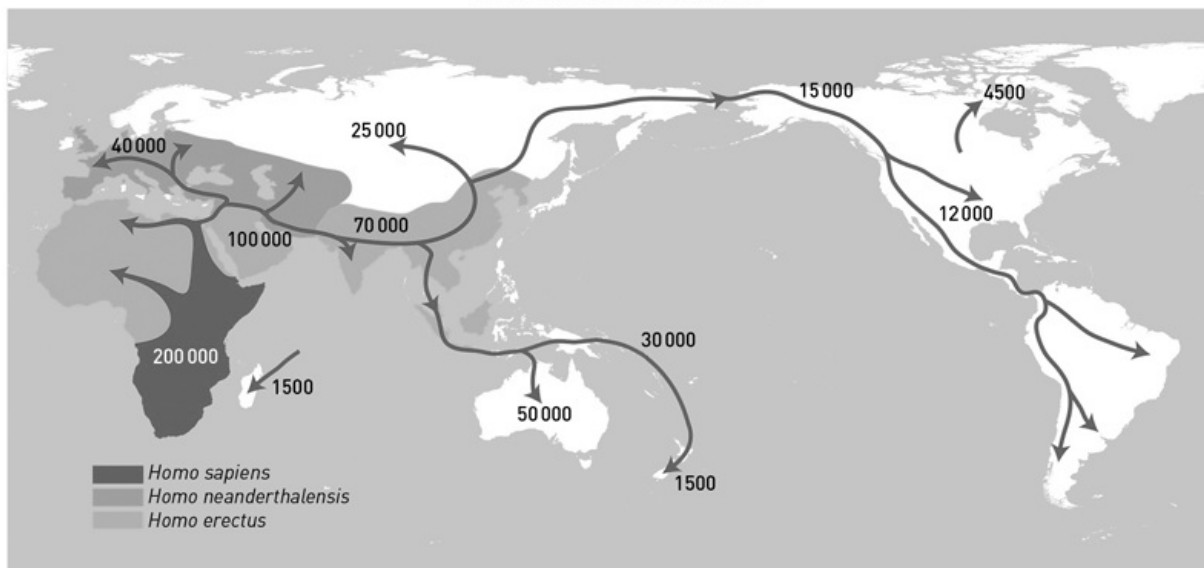
Однако раннюю историю человечества отделяют от современности несколько крупных технологических революций, к рассмотрению которых мы и переходим.

## Охота и собирательство

- Базовый ресурс: дикая природа.
- Технологический пакет: огонь, каменные, деревянные и костяные орудия, язык.
- Организация жизни: охота, собирательство, родоплеменной строй.

Древние люди доисторической эпохи использовали ресурсы дикой природы — жили охотой и собирательством. К какому-то моменту (около 70 000 лет назад) вид *Homo sapiens* накопил критическую массу изобретений, составивших первый технологический пакет человеческой цивилизации. Этот пакет (огонь, топоры, копья, дротики, лук со стрелами и т.д.) позволил людям занять нишу абсолютного хищника, довольно быстро покинуть Африку и расселиться по всему земному шару, уничтожив по пути другие, менее технологически развитые ветви рода *Homo* (см. рис. 2)<sup>6</sup>.

## ЭКСПАНСИЯ *HOMO SAPIENS*



**Рис. 2.** Экспансия *Homo sapiens* по миру с вытеснением более ранних выходцев из Африки — *Homo erectus* и *Homo neanderthalensis*. Первая волна *Homo sapiens* вышла из Африки более 100 000 лет назад, но была остановлена неандертальцами на севере Аравийского полуострова. Вторая волна, пересекшая Красное море в его узкой южной части (на схеме не показана) около 70 000 лет назад, довольно быстро завоевала всю Евразию и Австралию, а впоследствии, после перехода по замерзшему Беринговому проливу, и обе Америки. Когнитивная революция, обеспечившая успех последней экспансии, произошла где-то между этими двумя исходами<sup>6</sup>. На рисунке показана примерная датировка событий в годах от настоящего времени. ([https://commons.wikimedia.org/wiki/File:Spreading\\_homo\\_sapiens\\_la.svg](https://commons.wikimedia.org/wiki/File:Spreading_homo_sapiens_la.svg))

Какая технология играла роль замыкающей, чье появление дало старт безудержной экспансии *Homo sapiens*? Это наверняка был не огонь, так как следы костров находят на стоянках древнего человека с незапамятных времен (более полумиллиона лет назад). Покорение огня, безусловно, сыграло выдающуюся роль в эволюции человека, так как изменило саму его физиологию. Переход на лучше усваиваемую вареную пищу сопровождался сокращением желудочно-кишечного тракта, что, в свою очередь, позволило людям нарастить относительный объем головного мозга<sup>[4]</sup>. У горилл, питающихся сырой растительной пищей, мозг вдвое меньше человеческого при сопоставимых размерах тела.

По всей видимости, такой замыкающей технологией стало появление языка — одного из последствий увеличения объема мозга. Дело в том, что язык занимает определенное место в нашем мозгу, в частности языковые зоны Брока и Вернике в неокортексе. В меньшем по объему мозге других приматов этого места может просто не найтись, хотя наш мозг и имеет единую для всех приматов архитектуру<sup>7</sup>. Это именно тот случай, когда размер имеет критическое значение.

Размер мозга приматов ограничивает размер их стай, поскольку приматы, как социальные животные, должны запоминать отношения всех членов стаи между собой, а объем этих знаний растет квадратично с увеличением ее размера. Так что чем больше мозг, тем больше и мощнее стая<sup>8</sup>. Древние люди, охотясь коллективно, выигрывали от роста размера

своих сообществ, что оказывало эволюционное давление на увеличение их мозга. В какой-то момент, по-видимому примерно 70 000 лет назад, размер мозга *Homo sapiens* превысил некий критический порог, что сопровождалось появлением символического мышления и языка с развитым синтаксисом, которые напоминают современные<sup>9</sup>.

Язык и тесно связанное с ним мышление позволили людям организовываться в гораздо более крупные и сплоченные общности, скрепленные общей мифологией, противостоять которым более примитивные ветви рода *Homo* были не в состоянии. Язык, безусловно, является важнейшим изобретением человечества, необходимым условием для передачи и накопления знаний из поколения в поколение. Все остальные изобретения, от дротиков и лука до торговли и разделения труда, являются следствием появления языка и коллективного разума человечества — культуры. Поэтому революцию верхнего палеолита автор нашумевшей книги «Sapiens. Краткая история человечества» Юваль Ной Харари назвал *когнитивной революцией*<sup>10</sup>.

Главным наблюдаемым последствием этой технологической революции стало существенное увеличение количества людей по сравнению с типичной численностью популяций других крупных гоминид (порядка  $10^5$  особей). Мощность новорожденной человеческой цивилизации довольно быстро (за несколько десятков тысяч лет) выросла на порядки, при том что количество колонистов — выходцев из Африки, согласно оценкам генетиков, не превышало нескольких сотен человек.

Мы можем оценить предельную численность цивилизации охотников-собирателей следующим образом. Хищнику нашего размера нужна для пропитания площадь порядка  $10 \text{ км}^2$ . Следовательно, предельная плотность населения охотников-собирателей не превышает  $0,1 \text{ чел./км}^2$ . Умножая эту плотность на общую площадь доступной суши, получим предельное количество людей на планете к моменту ее окончательного заселения (около 15 000 лет назад, после заселения обеих Америк через Берингию) — 10 млн человек (см. рис. 3).

## ПАЛЕОЛИТ

Питание: охота/собираительство

**Человек — абсолютный хищник:** 0,1 чел./км<sup>2</sup>

**Площадь доступной суши:** 100 млн км<sup>2</sup>

**Лимит населения Земли:** 10 млн чел.

**Рис. 3.** Оценка численности цивилизации охотников-собирателей (оценка плотности популяций хищников согласно Carbone C., Gittleman J. L. (2002) A Common Rule for the Scaling of Carnivore Density. *Science*. Mar 22, 295(5563): 2273)

---

Это на два порядка превышает численность африканской популяции древних людей, прародителей современного человечества. Окончательное заселение всей поверхности Земли благодаря верхнепалеолитическим охотничьим технологиям привело к первому в человеческой истории экологическому кризису — вымиранию большинства крупных животных вследствие нашествия *Homo sapiens*<sup>11</sup>.

После освоения человеком всех более или менее пригодных для жизни уголков планеты ресурсы такого экстенсивного развития были исчерпаны. Возникли предпосылки для освоения нового базового ресурса на основе нового технологического пакета.

## Аграрный уклад

- Базовый ресурс: плодородная почва речных пойм и побережья рек.
- Технологический пакет: одомашнивание растений и животных, ремесла, судоходство, письменность.
- Организация жизни: сельские поселения и города, государство и право.

Выход из вышеупомянутого кризиса был найден около 10 000–12 000 лет назад в ходе так называемой *неолитической революции* и состоял в переходе к оседлому образу жизни, основанному на новом — аграрном — технологическом пакете.

Как всегда, его технологии появлялись в недрах старого уклада постепенно. Загоны, служившие для охоты, использовались в том числе для временного содержания пойманных животных, а затем и для их разведения. Собираительство дополнялось посевами диких культур вблизи от стойбищ и постепенным увеличением их урожайности вследствие искусственного отбора. Зависимость урожаев от плодородия почв притягивала людей в



поймы больших рек, где посадки в илистые отложения после разливов не требовали больших трудозатрат и снижали риски неурожаев. В таких местах и зародились первые великие цивилизации древности — в Египте, Междуречье, Индии, Китае, Центральной и Южной Америке.

То, что таких мест было несколько, лишний раз свидетельствует о глобальности палеолитического кризиса перенаселения. Сельское хозяйство в поймах и по берегам больших рек могло прокормить существенно больше людей, чем охота и собирательство по всей Земле. В этом легко убедиться, сделав соответствующие оценки (рис. 4).

Животноводство способно приносить порядка 100 кг белковой пищи в год с одного гектара. При полностью мясной диете оно может прокормить 0,2 чел./га — в 200 раз больше, чем охота. Молочная диета еще выгоднее и способна прокормить впятеро больше людей — 1 чел./га, поэтому буренок в деревнях берегут как зеницу ока.

Земледелие способно обеспечить еще большую плотность населения; даже при скромной урожайности зерновых 10 ц/га это составляет 2 чел./га. Несколько неожиданно, но масличные культуры (например, оливковые деревья) могут прокормить примерно столько же человек, если считать только по калориям.

Поскольку людям нужно сбалансированное по жирам, белкам и углеводам питание, специализация разных регионов на «своих» сельскохозяйственных культурах сделала торговлю необходимым элементом неолитического технологического пакета, что, в свою очередь, привело к развитию письменности, мореплавания и связанных с ними ремесел, включая металлургию.

## АГРАРНАЯ ЦИВИЛИЗАЦИЯ

### Животноводство

**Мясо:** 100 кг белка/га → 0,2 чел./га = 20 чел./км<sup>2</sup>

**Молоко:** 500 кг белка/га → 1 чел./га

### Земледелие

**Зерновые:** 1 т/га → 2 чел./га = 200 чел./км<sup>2</sup>

**Оливки:** 200 кг масла/га → 2 чел./га

### Освоение бассейнов рек

**Побережье рек:** 7 × 10<sup>5</sup> км × 8 км – 5 млн км<sup>2</sup>

**Лимит населения:** 1 млрд чел.

**Рис. 4.** Оценка аграрного населения Земли в поймах рек

Морская и речная торговля на порядки выгоднее сухопутной по энергозатратам и до появления двигателей на ископаемом топливе была единственным способом обмена пищевыми продуктами массового спроса. Действительно, гужевой транспорт сам потребляет зерновые в таком количестве, которое исключает его транспортировку на далекие расстояния. Так, лошадь способна за сутки перевезти 200 кг на 40 км<sup>12</sup>. При этом она сама потребляет более 10 кг фуража в день. Значит, за 10 дней (400 км) лошадь съест более половины груза, а с учетом обратной дороги — весь свой груз. Таким образом, рентабельные сухопутные перевозки пищевых продуктов, транспортная составляющая которых не превышает 10%, ограничены одним суточным переходом, то есть 40 км. Действительно, среднее расстояние между европейскими городами в середине XIX века составляло сопоставимые 20–30 км<sup>13</sup>.

Давайте оценим предельную мощность такой аграрной цивилизации, в которой основным источником свободной энергии является мускульная сила животных. Вспомним, что аграрные поселения привязаны к поймам судоходных рек, общая длина которых равна примерно  $7 \times 10^5$  км<sup>14</sup>. Логично ограничить ширину заселения примерно 8 км (по 1 часу ходьбы на работу и с работы в обе стороны от жилья). Это даст нам предельную площадь обрабатываемых земель:  $5 \times 10^6$  км<sup>2</sup> =  $5 \times 10^8$  га. Плюс еще примерно вдвое больше лугов для скота (он может проходить и по 12 км туда и обратно). В итоге получим предельную численность населения с аграрным технологическим пакетом: 2 чел./га  $\times$   $5 \times 10^8$  га = 1 млрд чел.

Энергетическая мощность сельскохозяйственной цивилизации определяется поголовьем тяглового скота. При средней урожайности лугов 1 т сена/га один гектар может прокормить 0,2 лошади<sup>15</sup>. Следовательно, при вдвое большей площади под луга, чем под пашню, то есть имея  $10^9$  га, люди смогут заготовить  $10^9$  т сена для 200 млн лошадей. Получаем исторически обоснованное соотношение 0,2 лошади на человека, то есть энерговооруженность такой цивилизации равна:  $0,2 \times 735$  Вт = 150 Вт/чел. (в 2,5 раза больше собственных 60 Вт/чел.).

Можно добавить сюда еще и энергию ветра — основу всей морской торговли. Однако в общем энергетическом балансе ее роль, за редкими исключениями, была невелика. Так, если принять, что установленная парусная мощность в лошадиных силах примерно равна водоизмещению в тоннах<sup>15</sup>, то голландский флот второй половины XVII века общим водоизмещением около 3 млн т был сравним по мощности с 3 млн лошадей<sup>16</sup>, притом что население Нидерландов того времени составляло 2 млн человек. Таким образом, именно энергия ветра была самым важным источником экономической мощи Соединенных провинций. Впрочем, эту

страну в разгар развернувшейся там «нулевой» промышленной революции уже нельзя считать аграрной. Скажем, объединенный флот Франции и Англии того же времени был вдвое меньшим, а суммарное население — на порядок больше (23 млн чел.), чем у Нидерландов. Так что «установленная мощность» ветровой энергии для них была эквивалентна 1,5 млн лошадей, или  $1/15$  лошадей/чел., — втрое меньше, чем «сухопутные»  $1/5$  лошадей/чел.

Таким образом, аграрная цивилизация может прокормить в 100 раз больше людей, чем палеолитическая, и обеспечить более чем втрое бóльшую энерговооруженность каждого жителя[6].

### АГРАРНАЯ ЦИВИЛИЗАЦИЯ ЭНЕРГОПОТРЕБЛЕНИЕ / ЧЕЛ.

Механическая ~ 600 Вт топлива

**Люди:** 145 Вт пищи

**Лошади:**  $0,2 \times 0,2$  кВт = 440 Вт корма

Тепловая ~ 400 Вт

**Дрова:** 400 Вт (1 т/чел. в год )

Рис. 5. Энергопотребление в аграрной цивилизации

Кроме еды человеку нужен еще и огонь для готовки и обогрева, особенно в умеренных и северных широтах: около  $1 \text{ м}^3$  дров/чел. в год[7]. Поскольку 1 га соснового леса дает прирост  $2 \text{ м}^3$  в год<sup>17</sup>, он может обогреть двух человек, то есть площадь доступных для вырубki лесов должна быть не меньше площади пашни. Доступный лес — это территория, расположенная не далее половины дневного перехода гужевого транспорта, то есть не далее 20 км. Площадь вокруг города с таким радиусом вполне достаточна для пропитания и обогрева небольшого города с населением до 60 000 человек[8]. Действительно, средневековые города обычно насчитывали не более нескольких десятков тысяч жителей, а такие крупные города, как Венеция или Флоренция, с населением порядка 100 000 человек, были скорее исключением[9].

Таким образом, мы «рассчитали» образ жизни людей (способ расселения плюс энерговооруженность) в аграрной цивилизации — деревни и небольшие города по берегам судоходных рек (особенно вблизи побережья морей и океанов, по которым осуществляется морская торговля) с

использованием тягловой силы животных — по 0,2 л. с./чел. Предел численности такой цивилизации — 1 млрд человек — был в реальности достигнут к концу XVIII века. Дальнейший рост был возможен только в рамках следующего — индустриального — технологического уклада. И действительно, приблизительно в это время в Европе началась (первая) промышленная революция.

## Индустриальный уклад

- Базовый ресурс: ископаемое топливо.
- Технологический пакет: машиностроение, двигатели, глобальные транспортные сети.
- Организация жизни: города, глобализация.

Переход от аграрного к индустриальному укладу принято делить на несколько этапов, называемых промышленными революциями.

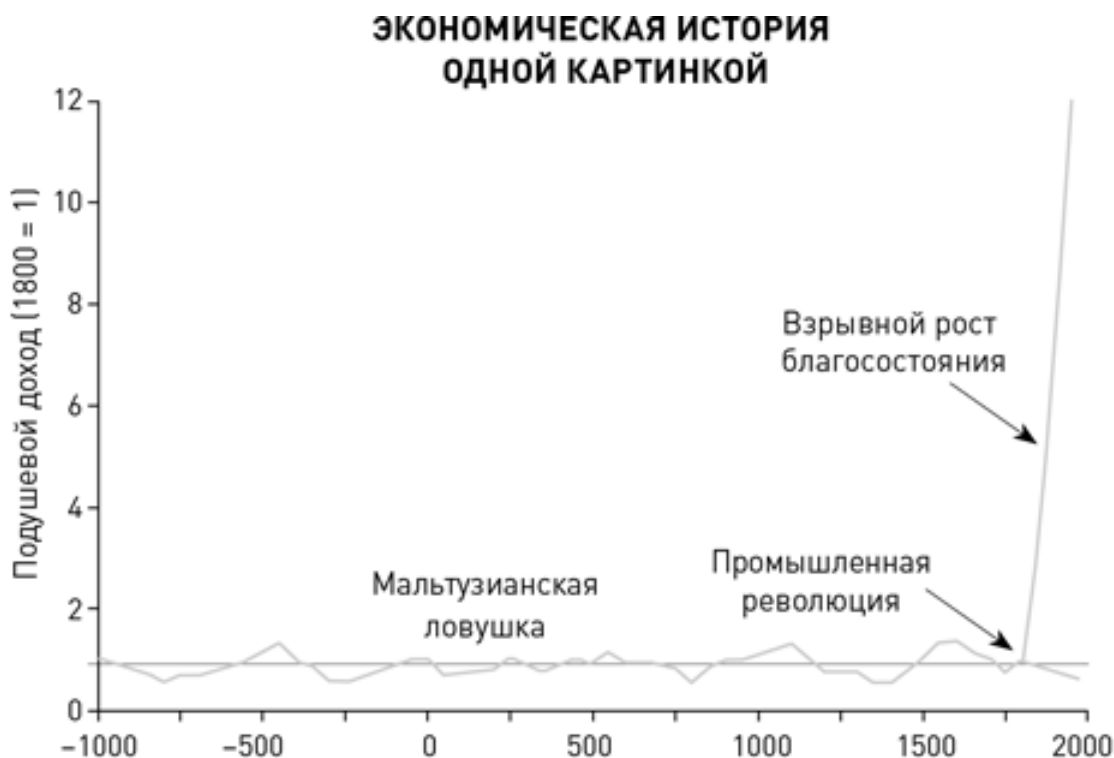
В первую очередь это первая промышленная революция XVIII — первой половины XIX века, превратившая Англию в главную «мировую фабрику». Ее энергетической основой был каменный уголь, заменивший лес, почти полностью сведенный на нет вследствие синхронного роста населения, кораблестроения и производства продуктов черной металлургии[10], а технологический пакет был основан на паровом двигателе[11].

Со второй половины XIX до середины XX века первая промышленная революция перерастает во вторую во главе с США. Список энергоносителей пополняется сначала нефтью для транспорта, а затем и природным газом для отопления. Технологический пакет массового индустриального производства основан на двигателях внутреннего сгорания, электродвигателях и турбинах, дополненных машиностроением, нефтехимией, производством стали и цветных металлов.

Поскольку базовым ресурсом остается ископаемое топливо, а технологический пакет по-прежнему основан на двигателях и машиностроении, мы будем считать эти две промышленные революции разными стадиями перехода от аграрного уклада к индустриальному. По сути, эти две революции представляют собой длинные волны циклического развития экономики, так называемые циклы Кондратьева.

Индустриальный уклад качественно изменил жизнь людей, позволив им вырваться из мальтузианской ловушки. Ранее на протяжении всей истории человечества любой прирост продуктивности мгновенно съедался ростом населения. Его численность всегда ограничивалась болезнями и голодом, большинство людей жили впроголодь и вообще редко доживали до

биологической старости. При индустриальном укладе впервые в истории темпы роста производства превысили естественные темпы роста народонаселения. В результате начиная с XIX века в мире наблюдается значительный и постоянный рост доходов на душу населения. Так что, несколько утрируя, всю экономическую историю человечества можно охарактеризовать одной картинкой (рис. 6).



**Рис. 6.** Выход из мальтузианской ловушки в результате промышленной революции: доход на душу населения начал расти только после того, как темпы роста ВВП превысили темпы рождаемости (Clark G. (2007) Farewell to Alms: A Brief Economic History of the World. Princeton, N. J.: Princeton University Press)

С появлением сети железных и автомобильных дорог исчезла жесткая привязка пашни к руслам рек[12]. Были освоены великие равнины Северной Америки, целинные земли Казахстана и другие ранее недоступные участки суши. Сегодня общая площадь пашни (15 млн км<sup>2</sup>) втрое больше, а урожайность зерновых (3 т/га) втрое выше, чем до промышленной революции, благодаря достижениям генетики, инсектицидам и удобрениям. Иными словами, новая транспортная инфраструктура и агрохимия позволяют индустриальной цивилизации прокормить на порядок больше людей, чем аграрная: до 10 млрд человек за счет использования ископаемого топлива. Промышленная революция

увеличила энерговооруженность человека в 2,5 раза: с 0,6 кВт/чел. до 1,4 кВт/чел. (в среднем по миру), а в развитых странах — более чем на порядок (рис. 5 и 7).

Изменился и стиль жизни — из сельского он стал городским. При современном оснащении тракторами, комбайнами и прочей техникой для производства сельхозпродукции на имеющихся площадях достаточно нескольких процентов населения. Поэтому переход к индустриальному укладу вызвал масштабную миграцию людей в города, размер которых больше не сдерживался транспортными ограничениями аграрного периода, связанными, как мы помним, с проблемами их снабжения внешними ресурсами. Размер современных городов ограничивается уже возможностями городского транспорта по организации суточных перетоков жителей от места проживания к месту работы и обратно. Если тратить на дорогу не более часа в одну сторону, то диаметр современного города ограничен несколькими десятками километров. Чтобы разместить постоянно растущее население на этой ограниченной площади, города стали расти сначала вверх (начиная с центральных деловых кварталов), а затем, с началом массовой автомобилизации, обрастать одноэтажными пригородами — своего рода «аэродромами подскока» на том же характерном расстоянии часовой доступности.

<b>Мир начала XXI века</b> <b>ЭНЕРГОПОТРЕБЛЕНИЕ / ЧЕЛ.</b>	<b>США начала XXI века</b> <b>ЭНЕРГОПОТРЕБЛЕНИЕ / ЧЕЛ.</b>
Механическая энергия ~ 1,4 кВт топлива <b>Транспорт: 0,6 кВт топлива</b> <b>Электричество: 0,8 кВт топлива</b>	Механическая энергия ~ 7 кВт топлива <b>Транспорт: 3 кВт топлива</b> <b>Электричество: 4 кВт топлива</b>
Тепловая энергия ~ 0,6 кВт <b>Дома и офисы: 400 Вт</b> <b>Промышленность: 200 Вт</b>	Тепловая энергия ~ 4 кВт <b>Дома и офисы: 1,3 кВт</b> <b>Промышленность: 2,6 кВт</b>
<b>Рис. 7а. Энергопотребление в среднем по миру</b>	<b>Рис. 7б. Энергопотребление в США</b>

(<https://www.iea.org/topics/global-energy-review>)

Старые города, типичное расстояние между которыми составляло, как мы помним, те же десятки километров, стали сливаться друг с другом, образуя гигантские мегаполисы и городские агломерации с миллионами, а зачастую и с десятками миллионов жителей. В скором времени

прогнозируется появление первых городских агломераций с населением, превышающим 100 млн, бóльшим, чем население иных стран. Поскольку и богатство в основном концентрируется в крупных городах, их роль в современной экономике возрастает.

Концентрация богатства в городах связана с тем, что в них обычно сосредоточено производство наиболее сложных и дорогих продуктов и услуг, создаваемых при участии представителей множества профессий. Ведь чем крупнее город, тем большее разнообразие специалистов можно собрать в одном месте, тем сложнее и дороже будет их продукция в пересчете на одного работника. Поэтому уровень жизни в мегаполисах с более развитой системой разделения труда обычно существенно выше среднего, впрочем, как и уровень цен.

Несмотря на все свои достижения, индустриальная цивилизация, основанная на ископаемом топливе, к концу XX века, по-видимому, уже достигла предела своего развития. Исследования Римского клуба [13] свидетельствуют о том, что несущая способность Земли была превышена уже к концу 1970-х<sup>18</sup>. Это означает, что современный пакет технологий не способен переработать все отходы цивилизации и планета забивается мусором (вспомним о гигантских островах из пластика в океане), резко повышается уровень CO<sub>2</sub> в атмосфере.

Свободной земли, пригодной для промышленного сельского хозяйства, на планете уже не осталось. Под производство пищи занято 50% всей доступной суши, не занятой ледниками, горами и пустынями (рис. 8). Борьба идет за сохранение имеющихся площадей, которые зачастую не успевают рекультивировать.

Замедление роста населения Земли, наблюдающееся со второй трети XX века, является важным индикатором современного глобального кризиса. Точка перегиба кривой роста народонаселения была пройдена примерно в 1970 году, когда на Земле наблюдался пик скорости роста населения (2% в год) [14]. Прирост идет только за счет сельской местности, тогда как промышленные города не способны воспроизводить свое население. При этом в начале XXI века число жителей городов превысило количество сельских жителей, и их доля продолжает расти. В результате демографы прогнозируют стабилизацию населения Земли в XXI веке на уровне около 10 млрд человек, притом что на протяжении всей известной истории оно росло все возрастающими темпами (кроме, естественно, кризисных периодов).

## ИСПОЛЬЗОВАНИЕ ПОВЕРХНОСТИ ЗЕМЛИ

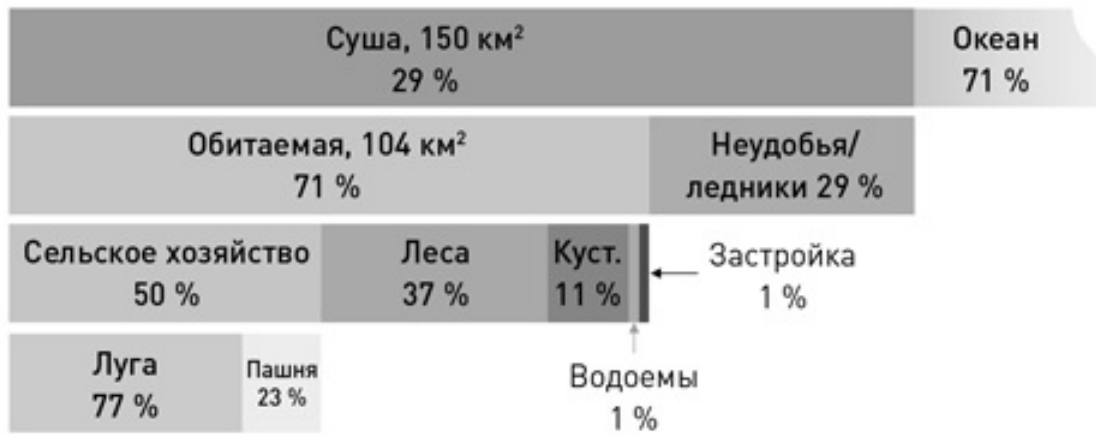


Рис. В. Мировое сельское хозяйство использует 50% всех доступных земель

(<http://OurWorldinData.org>)

Ответом на этот новый глобальный кризис является то, что обычно называют третьей промышленной революцией, связанной с цифровой обработкой информации, и четвертой — связанной с киберфизическими системами<sup>19</sup>. Мы будем считать их разными фазами большой *цифровой революции*, переходом к новому *цифровому* технологическому укладу, ключевым ресурсом которого станет уже не энергия, а информация.

Поскольку наша основная задача — осмыслить эту разворачивающуюся на наших глазах революцию, имеет смысл остановиться на ее причинах и особенностях подробнее. В следующем разделе мы покажем, что эти причины носят фундаментальный характер, являясь следствием закона гиперболического роста населения Земли. Вернее, предстоящий переход к новому укладу вызван теми же причинами, что и закон гиперболического роста народонаселения.

### Закон гиперболического роста народонаселения

Если мы наложим все известные факты о населении Земли на временную ось, окажется, что лучше всего этот график описывается гиперболой:

$$N = N_0 / (t - t_0),$$

где  $N_0 = 2 \times 10^{11}$  людей, а время измеряется в годах (рис. 9).

Иными словами, население растет обратно пропорционально времени,

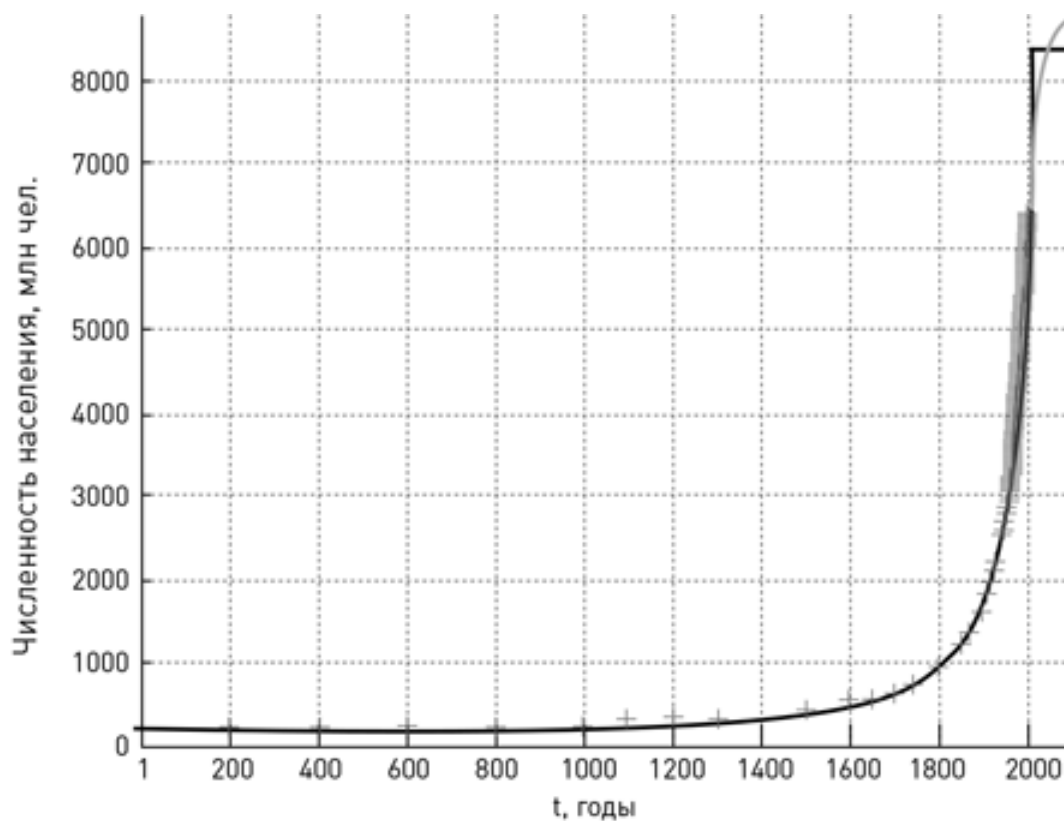


оставшемся до некоторой даты, определяемой эмпирически как  $t_0 = 2026$  г.<sup>20</sup>:  $10^9$  людей 200 лет назад,  $10^8$  — 2000 лет назад и т.д. В эту зависимость хорошо вписываются и наши оценки населения Земли в доисторический период:  $10^7$  охотников-собирателей 20 000 лет назад и  $10^5$  первобытных людей 2 млн лет назад. Таким образом, закон гиперболического роста «работает» уже миллионы лет.

При этом сегодня мы находимся в непосредственной близости от точки  $t_0$  — *сингулярности*. Понятно, что вблизи таких особых точек прежние закономерности не могут не измениться: население не может стремиться к бесконечности. И действительно, как известно, темпы роста народонаселения с конца XX века начинают замедляться, и вместо бесконечного роста мы наблюдаем переход к стабилизации численности населения. Таким образом, мы живем в особое время — на переломе.

Возникают естественные вопросы: что обуславливало гиперболический рост на протяжении всей известной истории и почему эти причины сегодня не работают? Ответы на эти вопросы легче понять, если записать закон гиперболического роста в дифференциальном виде, через скорость роста населения:

$$\frac{dN}{dt} = \frac{N^2}{N_0}.$$



**Рис. 9.** Гиперболический рост народонаселения Земли (История и математика: анализ и моделирование глобальной динамики / Под ред. А. В. Коротаева, С. Ю. Малкова, Л. Е. Гринина. — М.: Либроком, 2010)

Почему скорость роста населения пропорциональна квадрату числа людей? Первую степень было бы легко объяснить: число детей пропорционально числу матерей. Но почему интенсивность рождения детей женщинами оказывается пропорциональна числу людей, ведь они, как правило, даже не имели понятия, сколько их на Земле?

Ответ тривиален: конечно, женщины рожают детей вне зависимости от численности населения. Но вот выживает в итоге столько людей, сколько цивилизация на данной стадии ее развития может прокормить. А это число определяется общим количеством *знаний*, накопленных человечеством и определяющих сумму используемых им технологий. Чем больше людей, тем больше суммарный объем знаний, тем выше уровень техники и производительность каждого человека. Налицо групповой эффект — следствие общественного разделения труда. Отсюда и квадратичный закон для прибавочного продукта, часть которого идет на расширенное производство и обеспечивает прирост населения. Каждый человек на Земле вносит свою малую лепту в коллективную производительность труда (точнее, настолько малую, равную в среднем  $1/N$ , что скромность можно

считать эмпирически обоснованным качеством).

Закон гиперболического роста населения подсказывает нам следующую гипотезу: количество знаний, накопленных человечеством, пропорционально квадрату числа людей[15]:

$$C = \alpha N^2.$$

Знания человечества можно измерять, например, номенклатурой товаров, так как производство каждого изделия требует определенных профессиональных навыков и знаний. Если принять, что номенклатура олдувайской культуры древних людей 2 млн лет назад составляла единицы изделий, а в современной цивилизации составляет  $10^{10}$  видов продукции (разнообразие товарных штрих-кодов UPC), значит, за это время знания человечества возросли на десять порядков. При этом численность народонаселения за то же время увеличилась всего на 5 порядков — с  $10^5$  до  $10^{10}$  человек. Так что даже такая грубая оценка подтверждает квадратичную зависимость количества знаний от числа людей[16].

Это, конечно, не значит, что знания каждого человека выросли за это время на пять порядков. Скорее всего, средний объем знаний типичного человека на протяжении всей истории существенно не менялся. Что изменялось с течением времени, так это избыточность информации. Чем примитивнее общество, чем меньше в нем степень разделения труда и тем больше знаний в головах различных людей дублируется. С ростом разделения труда такое дублирование информации снижается. И если все древние люди обладали примерно одним и тем же набором знаний, то в современном обществе существует уже порядка  $10^5$  профессий, и гранулярность человеческих знаний продолжает стремительно нарастать.

Но, как мы все отлично понимаем, существует предел гранулярности знаний, при достижении которого знания каждого человека станут уникальными. После этого квадратичный закон накопления знаний, который до сих пор обеспечивался ростом степени разделения труда, перестанет выполняться. И достижение этого предела неизбежно, поскольку суммарная емкость человеческой памяти пропорциональна первой степени количества людей, а не квадрату их числа. Рано или поздно все знания человечества уже не смогут помещаться в наших головах! Возможно, мы уже вплотную приблизились к этому *пределу сложности* человеческой цивилизации, так как, если знания всех людей не будут пересекаться вовсе, они перестанут понимать друг друга.

Итак, к чему мы пришли в результате нашего анализа? К тому, что знания человечества, а значит, и производительность труда уже не смогут

увеличиваться прежними темпами. Более того, когда численность жителей Земли стабилизируется, объем знаний вообще перестанет расти, так как суммарная емкость человеческой памяти достигнет своего предела и никаких резервов за счет снижения избыточности знаний уже не останется.

Таким образом, мы приходим к выводу, что современная цивилизация достигла пределов своего развития, упершись в ограниченные способности человеческого мозга. Более сложную и производительную цивилизацию, опираясь лишь на человеческий разум, мы построить не в состоянии!

Отсюда неминуемо следует, что, если мы хотим двигаться вперед, если хотим лучшей жизни для наших детей и внуков, мы должны найти если не замену, то какое-то подспорье человеческому разуму, то есть создать *искусственный интеллект*. И жизнь, если присмотреться, уже сделала этот свой выбор.

## Цифровой уклад

- Базовый ресурс: информация и знания.
- Технологический пакет: компьютеры, сети, машинный интеллект.
- Организация жизни: удаленная работа, сетевые организации, виртуальные миры.

Итак, цифровая революция, которую мы сегодня наблюдаем, качественно отличается от всех предыдущих. До сих пор человечество совершенствовало свои технологии, используя те возможности мозга, которые мы приобрели в ходе биологической эволюции. Наш мозг на протяжении последней пары миллионов лет увеличился втрое, но он не может расти с той же скоростью, с какой развивается наша цивилизация [\[17\]](#).

Следовательно, для своего дальнейшего развития человеческое мышление должно покинуть биологическую колыбель и расширить границы своего обитания, то есть перестать быть чисто человеческим. И для этого уже подготовлена подходящая цифровая среда — компьютеры, соединенные в глобальную сеть, куда уже десятки лет выгружаются человеческие знания.

Эта среда создавалась еще в рамках индустриального уклада, будучи зародышем нового — цифрового — уклада. И этот зародыш рос экспоненциально, как и полагается новой фазе в ходе любого фазового перехода. Этот экспоненциальный рост нам известен как закон Мура: удвоение компьютерных мощностей каждые полтора года [\[18\]](#). Такая скорость развития на порядок превышает среднегодовые темпы роста мировой экономики (2–3%), соответствующие ее удвоению за 25–35 лет.

Таким образом, индустриальная цивилизация опережающими темпами превращается в цифровую, в которой создание стоимости будет связано в основном с переработкой информации, а не энергии и основными товарами станут цифровые продукты и услуги.

Это не значит, что люди перестанут пользоваться материальными благами. Конечно, они все больше времени будут проводить в онлайн: работать в виртуальных офисах, посещать виртуальные музеи, общаться, развлекаться и зарабатывать в виртуальных игровых мирах. Молодежь уже проводит существенную часть своего времени в социальных сетях, а по мере созревания технологий виртуальной и дополненной реальности эта тенденция только усилится. Однако и в материальных товарах, и в услугах все возрастающую роль будет играть цифровая составляющая. Например, цена современных автомобилей, особенно класса люкс, все в большей степени определяется стоимостью цифровых компонентов и программных продуктов, а со временем практически все продукты и услуги будут насыщены электроникой и искусственным интеллектом.

Потенциальный массовый рынок роботов и искусственного интеллекта измеряется десятками триллионов долларов. По существу, речь идет о реформатировании всей экономики. Если индустриальная экономика была основана на массовом производстве *типовых* продуктов и услуг, то цифровая позволит организовать массовое производство *кастомизированных* товаров и услуг, ориентированных на *индивидуальные потребности*. В результате за счет резкого расширения товарной номенклатуры объем мирового рынка может вырасти в разы. Так, YouTube кардинально расширил разнообразие потребляемого видеоконтента. Вместо сотен кабельных каналов — миллионы пользовательских. В итоге молодежь отказывается от телевидения в пользу YouTube и стриминговых сервисов типа Netflix, где каждый может легко подобрать контент по своему вкусу. И так будет в каждом рыночном сегменте.

Уже сегодня мы наблюдаем рост глобальных платформ, обеспечивающих персонализированное потребление товаров и услуг за счет умных рекомендательных систем: Google и Facebook организуют персональную доставку информации и новостей, Amazon и eBay — продуктов и услуг, YouTube и Netflix — видеоконтента. И, как зародыши новой фазы, эти бизнесы растут на порядок быстрее остальной экономики: 25–50% против 2–3% (рис. 10).

Капитализация, млрд \$		2013	→	2018
Apple	USA	\$418	17% в год	\$924
Amazon	USA	121	45% в год	783
Microsoft	USA	291	21% в год	753
Google / Alphabet	USA	288	20% в год	739
Facebook	USA	56	57% в год	538
Alibaba	China	—		509
Tencent	China	71	47% в год	483
Netflix	USA	13	63% в год	152

**Рис. 10.** Капитализация цифровых платформ увеличивается на порядок быстрее остальных отраслей экономики. Аналогичными темпами растут и их доходы

Поскольку в основе всех этих платформ лежат технологии искусственного интеллекта, вложения в них растут опережающими темпами и измеряются десятками миллиардов долларов в год. Эта новая мировая «гонка за ИИ» обостряется тем, что в мире платформ действует принцип «победитель забирает все». Обществу не нужен второй Facebook и второй Google. Кто не успел — будет подбирать крохи в оставшихся нишевых сегментах рынка.

Не приходится сомневаться, что в результате этой гонки вслед за беспилотными такси, которые уже можно встретить на улицах Москвы, появятся и автономные роботы, способные к самостоятельному разумному поведению, понимающие не только человеческую речь, но и человеческие ценности. Так что разворачивающаяся цифровая революция будет самой радикальной и самой важной трансформацией за всю историю человечества с момента когнитивной революции, ведь она затрагивает самую нашу суть — человеческий разум.

Если раньше революционные «разборки» шли между различными слоями общества, то теперь человечеству предстоит разобраться с новыми видами разумных существ, мыслящих и эволюционирующих самостоятельно. Неожиданно для себя мы оказались в ситуации неминуемой скорой встречи с неизвестным нам искусственным разумом. Какое место человек сможет занять в этом новом, цифровом, укладе? Насколько он будет самостоятелен в выборе своей судьбы? Мы этого не

знаем. Чтобы в этом разобраться, нам придется досконально рассмотреть ход цифровой революции и попытаться «вычислить» контуры нашего будущего, будущего наших детей и внуков, подобно тому как мы это делали для прошлых укладов.

## Резюме

В этой главе мы проследили в общих чертах логику развития человечества от древности до наших дней, обозначив базовые параметры цивилизации на каждой ступени ее развития. Логика дальнейшего развития тоже прояснилась: нам предстоит преодолеть барьер сложности современной цивилизации за счет создания и «приручения» искусственного интеллекта. Только таким образом человечество сможет развиваться дальше. Цифровая революция, как и всякая другая, таит в себе много рисков. В следующей главе мы проследим начальные этапы этой революции, сравнивая их с аналогичными этапами предыдущих промышленных революций, чтобы понять, где именно мы находимся и чего нам следует ожидать в ближайшем будущем.

## ГЛАВА 2

# Эволюция цифровых технологий: от первых ЭВМ до роботов

*Человек должен построить между собой и Природой целую цепь из звеньев, в которой каждое последующее звено будет как усилитель Разума более мощным, чем предыдущее.*

СТАНИСЛАВ ЛЕМ

*Мы знаем только два рода систем: очень простые, типа машин, строившихся нами до сих пор, и безмерно сложные, какими являются все живые существа.*

СТАНИСЛАВ ЛЕМ

## Основные понятия: информация, знания, искусственный интеллект

Поскольку мы начинаем активно использовать такие термины, как «информация», «знания» и «искусственный интеллект», имеет смысл определить, что мы под ними понимаем.

Начнем с того, что *цифровым* считается любое устройство, которое может устойчиво находиться в одном из нескольких дискретных состояний, то есть обладает *памятью*. Это качество, собственно, и определяет современную роль цифровых технологий как внешней памяти человечества. Информация является количественной мерой объема памяти и измеряется в *битах* [19]. Информацию можно не только хранить, но и передавать по каналам связи, а также обрабатывать в компьютерах согласно четко определенным правилам — *алгоритмам*.

Любой алгоритм характеризуется своей *сложностью*, которая также измеряется в битах, поскольку она определяется размером наиболее компактной записи алгоритма в памяти. В алгоритме заложены знания о том, что надо делать с полученной информацией для достижения какой-то цели, поэтому логично определить любое *знание* через соответствующий



ему алгоритм. Чем он сложнее, тем больше знаний в него заложено. Так что мы будем определять объем знаний сложностью соответствующих алгоритмов и измерять его в битах.

Таким образом, мы различаем пассивную *информацию*, которую можно хранить и передавать между устройствами ее обработки, и активные *знания*, заключенные в алгоритмах, используемых для обработки входящей информации. Такое определение согласуется с обыденным употреблением этих терминов: человек действует на основе полученной извне информации, осмысленной с помощью имеющихся в его мозгу знаний. Результаты своих размышлений он может передавать другим людям в виде выходной информации для ее дальнейшей обработки.

Как правило, первичная информация (*данные*, например показания датчиков) проходит множество этапов обработки, причем тем больше, чем более развита система разделения труда. Таким образом, человеческое мышление по самой своей природе коллективное. И если в эту цепочку вклинивается компьютерная обработка информации, мы получаем коллективное человеко-машинное мышление.

Цифровая революция состоит в постоянном увеличении доли машинной компоненты этого коллективного мышления. В какой-то момент непрерывно растущие возможности компьютеров неизбежно сравняются с человеческими, а затем и превзойдут их. В предвидении этого момента ведущие ученые в области компьютерных наук, собравшиеся летом 1956 года на двухмесячный семинар в Дартмутском колледже, ввели понятие *искусственного интеллекта*.

Человеческий интеллект в обыденной жизни связывают с умением решать самые разные задачи и добиваться своих целей, невзирая на обстоятельства. Аналогично *искусственный интеллект* ученые определили как способность компьютеров решать любые задачи и добиваться поставленных целей не хуже человека, например играть в интеллектуальные игры, доказывать математические теоремы и переводить тексты с одного языка на другой. В дальнейшем это определение было расширено за счет «детских» задач, которые обычно не считались интеллектуальными, но на поверку оказались очень сложными и долгое время не поддавались решению известными методами, такими как машинное зрение, чтение рукописных текстов, распознавание и понимание речи. Соответственно системы, способные решать хотя бы некоторые из этих очень разных задач, стали называть *слабым искусственным интеллектом*, чтобы подчеркнуть снижение первоначально слишком высокой планки *сильного искусственного интеллекта*, обладающего всеми человеческими способностями.

## Параметры человеческого мышления

Поскольку для нас человек есть мера всех вещей, включая искусственный интеллект, мы должны понимать, какими возможностями по переработке информации он обладает.

Наш мозг насчитывает немногим менее 90 млрд нейронов, которые время от времени возбуждаются (когда распознают в своем окружении знакомые им паттерны). По энергетическим оценкам, в каждый момент времени активным может быть лишь 1 из 50 нейронов<sup>1, 2</sup>. Разбив общее число  $N$  нейронов на  $\frac{N}{K}$  ячеек, в каждой из которых может быть активным лишь один из нейронов, получаем грубую оценку разнообразия паттернов мозговой активности в каждый момент времени:  $K^{\frac{N}{K}}$  [20]. Скорость обработки информации (равной логарифму разнообразия) определяется частотой смены этих паттернов:

$$\frac{dI}{dt} = \frac{vN}{K} \log_2 K \approx 3 \times 10^{12} \frac{\text{бит}}{\text{с}} \approx 4 \times 10^{11} \frac{\text{байт}}{\text{с}} \approx 10^{11} \text{ FLOPS.}$$

Здесь мы использовали максимальную частоту генерации нейронных спайков  $v = 300$  Гц, задающую временной масштаб смены состояний мозга, и выразили вычислительную мощность в привычных для компьютерщиков единицах — в числе арифметических операций в секунду [21].

По нынешним меркам  $10^{11}$  FLOPS — это не так уж и много. Современный универсальный процессор Intel Core i9 с производительностью  $10^{12}$  FLOPS и специализированные матричные процессоры NVIDIA Tesla V100 с пиковой производительностью  $10^{14}$  FLOPS уже далеко превосходят рассчитанную нами вычислительную мощность мозга. Этот парадоксальный вывод мы подробно обсудим чуть ниже.

Основным каналом получения информации из внешнего мира для нас является зрение. Скорость обработки зрительной информации примерно в 100 раз ниже вычислительной мощности мозга (поскольку первичная зрительная кора занимает примерно 1% площади всей коры), то есть составляет  $10^9$  FLOPS. Этого вполне достаточно для обработки информации, поступающей от глаза по зрительному нерву, содержащему  $10^6$  нервных волокон.

Сенсорная информация обрабатывается параллельными «спецпроцессорами» первичной коры, распознающей наиболее простые мелкомасштабные признаки. Далее информация проходит еще несколько этапов обработки, на каждом из которых распознаются все более сложные и

крупномасштабные признаки, пока наконец входная информация не будет представлена в символьной форме — в виде сцены с распознанными объектами и схемой их расположения. Например, это может быть текст с распознанными буквами и порядком их следования. Все этапы распознавания образов проходят параллельно, быстро и на бессознательном уровне. Осмысление распознанной символьной информации осуществляется в форме сознательного мышления и совсем в другом темпе «рассказывания историй».

Символы, в отличие от образов, обладают небольшим разнообразием. Например, буквы могут кодироваться всего лишь пятью битами (так как  $2^5 = 32$ ). Зато комбинациями компактного набора символов можно закодировать практически неограниченный набор высокоуровневых понятий, определяющих любую распознанную сцену. В приведенном выше примере чтения текста этим понятиям соответствуют слова и их сочетания. Сознательное мышление вообще тесно связано с языком и представляет собой последовательные операции с символьной информацией. Минимальный акт сознания длится около  $0,5 \text{ с}$ <sup>3</sup>. За это время мы успеваем распознать примерно одно слово в среднем из 5 букв, или около 25 бит. Соответственно сознательное мышление (около 50 бит/с, или 10 байт/с) гораздо — в миллионы раз! — медленнее бессознательного. Но именно с такой скоростью мы разговариваем, думаем, читаем, печатаем, решаем задачи<sup>4</sup>.

Сознательное мышление определяет *скорость обучения*, передачи знаний. Ведь большую часть наших профессиональных знаний мы получили в процессе коммуникации с другими людьми. Скорость обучения особенно важна, так как она ограничивает объем наших знаний. Даже если учиться непрерывно по 16 часов в день (то есть  $2 \times 10^7 \text{ с}$  в год) в течение 100 лет, максимальный объем знаний составит  $10^{11}$  бит, или  $10^{10}$  байт. Иными словами, знания каждого из нас составляют не более 10 ГБ, и их можно было бы записать на небольшую флешку[22].

Наши оценки объема знаний гораздо ниже, чем привычные  $10^{14}$  бит — по числу синапсов в коре головного мозга. Большая часть этих синапсов организует локальные контуры бессознательной обработки информации в коре. Наша более скромная оценка относится к глобальным связям, участвующим в запоминании распознанных нашим сознанием эмоционально окрашенных высокоуровневых понятий. Не случайно «оператор» человеческой памяти, гиппокамп, связан лишь с высокоуровневыми разделами ассоциативной коры.

## ИНТЕЛЛЕКТ ЧЕЛОВЕКА

Вычислительная мощность:  $\frac{dI}{dt} \approx 10^{11}$  FLOPS

Скорость обучения:  $\frac{dC}{dt} \approx 10 \frac{\text{байт}}{\text{с}}$

Объем знаний:  $C \approx 10^{10}$  байт

**Рис. 11.** Возможности человека по обработке информации и накоплению знаний

Таким образом, по сравнению с современными компьютерами наши возможности по обработке информации довольно скромны. Почему же машинный интеллект до сих пор не достиг человеческого уровня и мы по-прежнему считаем компьютеры «быстрыми, но глупыми»? Как ни странно, но именно по этой причине — из-за недостатка знаний о мире. Для решения сложных задач важны не столько вычислительные мощности, сколько сложные алгоритмы. А сложность современных компьютерных программ до сих пор недотягивает до человеческого уровня. Даже операционная система Windows, содержащая 50 млн строк кода, скажем, по 20 символов на строку, имеет сложность 1 Гбайт, то есть содержит меньше знаний, чем мозг одного человека. А ведь ее создавали сотни людей в течение десятков лет!

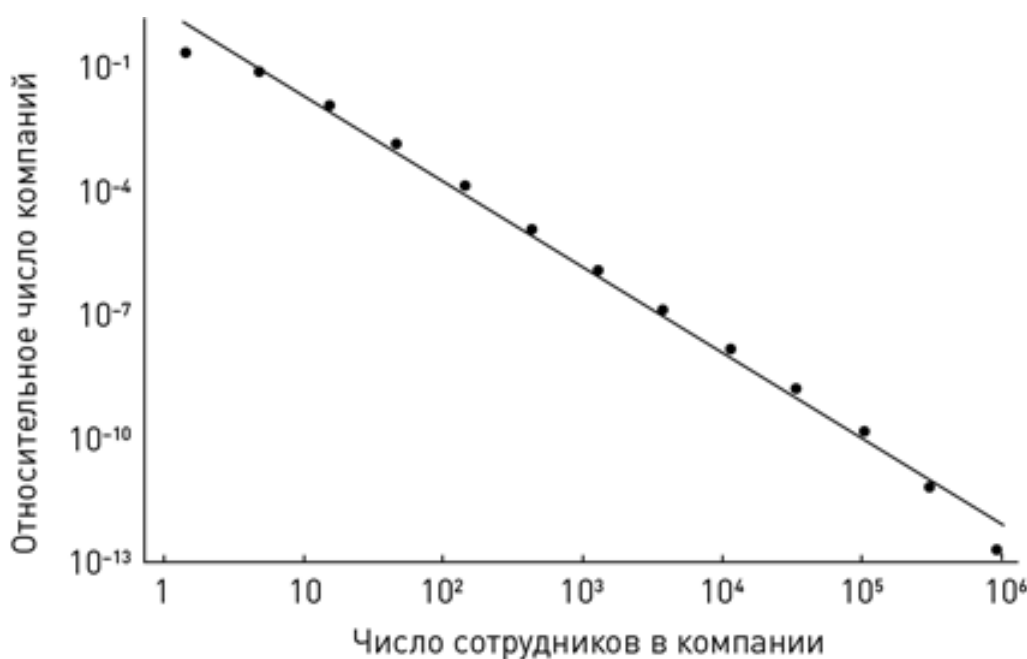
Значит, узким местом, мешающим созданию компьютерного разума, является трудоемкость ручного программирования или, в более общем случае, процесса передачи компьютерам знаний. Как только этот процесс будет автоматизирован, путь к искусственному интеллекту окажется свободен. Компьютерные мощности для него уже готовы. Вот почему сегодня фокус исследований в сфере ИИ сосредоточен в области *машинного обучения* — автоматизации процесса создания компьютерных программ. Современные искусственные нейросети могут содержать сотни миллиардов настроечных параметров<sup>5</sup>, то есть имеют сложность, в сотни раз превышающую Windows, притом что настройка всех этих параметров происходит автоматически, без участия человека — путем обучения на примерах. И, что самое важное, на много порядков быстрее и дешевле!

Почему же тогда машинное обучение не использовалось раньше? Ведь зародилось оно еще в конце 1950-х годов. Краткий ответ, как водится, прост: машинное обучение требует больших вычислительных мощностей, которые стали доступны лишь недавно. Более развернутый ответ, который займет оставшуюся часть этой главы, требует анализа движущих сил развития компьютерной техники: какие сегменты платежеспособного

спроса она обслуживала, какие задачи помогала решать на разных этапах своего развития? Иначе говоря, на чьи деньги развивались компьютеры все 75 лет своего существования? Как они постепенно все глубже проникали в ткань современной экономики?

## Гиперболическое распределение богатства

Здесь будет уместно познакомиться с еще одним законом — гиперболическим распределением богатства в человеческом обществе. Как было замечено с незапамятных времен, «деньги текут к деньгам» и социально-экономическое неравенство постоянно возрастает. Оказывается, это является следствием универсального гиперболического закона распределения любых ресурсов в так называемых *ветвящихся процессах* размножения и гибели<sup>6</sup>. Примером ветвящихся процессов служат коммерческие фирмы, которые могут разоряться и поглощать друг друга в ходе конкурентной борьбы.



**Рис. 12.** Гиперболическая зависимость относительного количества компаний США от числа их сотрудников (Axtell R. L. (2001) Zipf Distribution of US Firm Sizes. *Science* 293(5536): 1818–1820)

Нам, в частности, потребуется гиперболическое распределение компаний по их размерам, причем неважно, чем измерять эти размеры — денежной выручкой или количеством сотрудников. Для американского

рынка этот закон проиллюстрирован рис. 12.

Такая же гиперболическая зависимость справедлива и для распределения компаний по выручке: фирм с выручкой от \$10 млн в 10 раз меньше, чем с выручкой от \$1 млн. Как следствие, общая выручка компаний от \$100 000 до \$1 млн равна выручке компаний от \$1 млн до \$10 млн и т.д. вплоть до самых крупных компаний с выручкой более \$100 млрд. Каждый из этих сегментов — от миллионов индивидуальных предпринимателей до нескольких богатейших людей планеты — сосредоточивает в себе одинаковую долю общественного богатства. Таким образом, степень неравенства между обычными и самыми богатыми людьми (например, соотношение их доходов) пропорциональна численности населения Земли и, следовательно, постоянно возрастает.

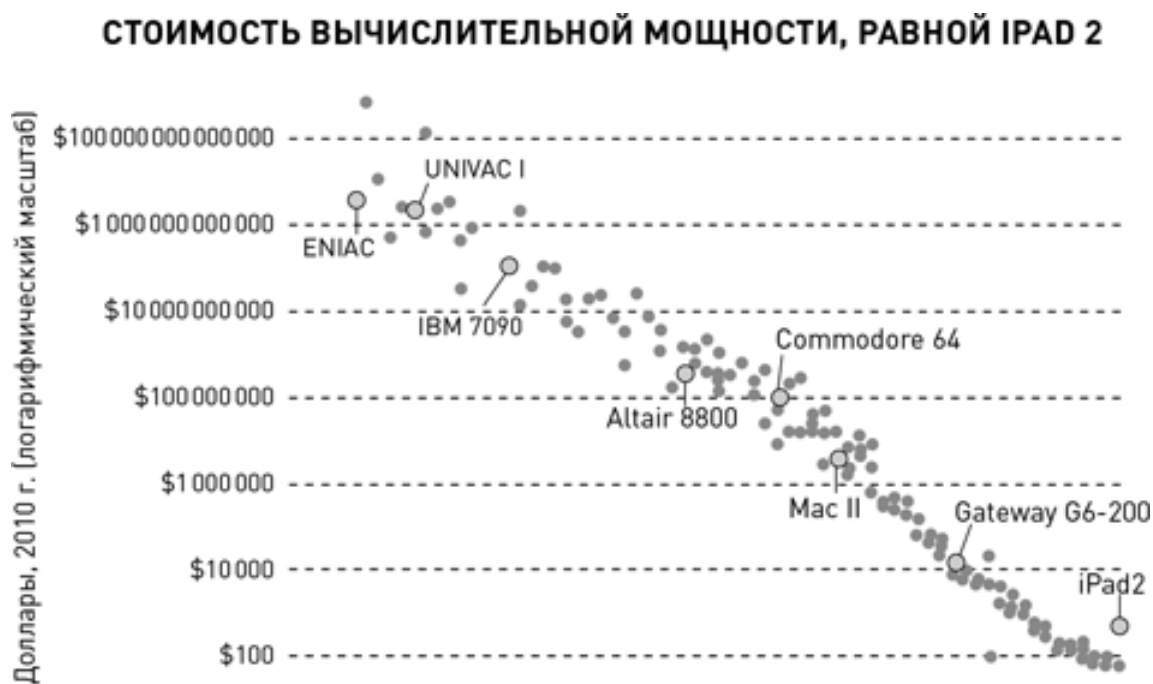
Нас, однако, здесь будет интересовать не этот факт, а скорость расширения устроенного таким образом рынка по мере удешевления любой продукции, в частности компьютеров. Если первые компьютеры были запредельно дороги и их могли позволить себе лишь национальные исследовательские центры, то по мере удешевления они стали проникать во все перечисленные выше сегменты рынка. При этом, как мы увидим, объем рынка компьютеров, несмотря на постоянное падение цен, все время возрастал, что служило стимулом к их дальнейшему удешевлению. Возникает знакомая нам положительная обратная связь. В итоге, как легко убедиться, рынок компьютеров растет экспоненциально.

Действительно, если компьютеры стоимостью \$1 млн могли позволить себе лишь фирмы с выручкой выше \$10 млн, то удешевление компьютеров на порядок «распечатывает» для них совершенно новый сегмент рынка — фирмы с выручкой выше \$1 млн, которых в 10 раз больше. Конечно, из-за удешевления продукции компании-покупатели нового сегмента потратят в итоге те же деньги, что и покупатели старой техники из более дорогого сегмента. Но и эти последние тоже закупают новое поколение компьютеров — для решения нового круга задач своими сотрудниками. Если принять, что на закупку новой вычислительной техники все потребители потратят одинаковую сумму в пересчете на одного сотрудника, то общая сумма закупок при освоении каждого нового сегмента рынка увеличится примерно вдвое, поскольку новые покупатели не заменяют, а расширяют круг старых.

Так, мини-компьютеры DEC не заменяли мейнфреймы IBM, просто их покупали уже не топ-менеджеры для корпораций, а руководители подразделений для своих нужд, увеличивая долю затрат на компьютеры в бюджете компаний. И это касается не только корпоративного, но и потребительского рынка. Ноутбуки не убили персональные компьютеры, а

смартфоны не убили ноутбуки. У каждого из нас можно обнаружить компьютеры разных поколений, дополняющие, а не исключаящие друг друга: настольный компьютер и ноутбук, смартфон и планшет, не считая встроенной бытовой и автоэлектроники. И каждый из них используется для решения своего круга задач.

Таким образом, каждое снижение цен на порядок означает более чем двукратный рост сектора ИКТ[23]. Поскольку стоимость вычислений каждые 10 лет снижалась на два порядка<sup>2</sup> (см. рис. 13), все 6 корпоративных сегментов рынка были освоены за 30 лет, и в 1980-х годах началось покорение потребительского рынка — настала эра персональных компьютеров (PC). При этом разнообразие цифровых продуктов продолжает увеличиваться, и их продажи по-прежнему растут по экспоненте — за 70 лет в  $2^{14} = 16\,000$  раз: с начальных \$300 млн в 1950 году до нынешних \$4,8 трлн, превратив ИКТ в один из крупнейших секторов экономики[24].



**Рис. 13.** За 60 лет, с 1950 по 2010 год, стоимость вычислений снизилась на 12 порядков, то есть снижалась на два порядка каждое десятилетие (Cost of Computing Power Equal to an iPad2. [https://www.hamiltonproject.org/charts/cost\\_of\\_computing\\_power\\_equal\\_to\\_an\\_ipad2](https://www.hamiltonproject.org/charts/cost_of_computing_power_equal_to_an_ipad2))

Такова в общих чертах бизнес-логика развития цифровой техники. Как видим, она является следствием универсального закона экономики. Осталось только понять, какие новые задачи позволяло решать каждое

новое поколение компьютеров.

В следующих нескольких разделах мы кратко опишем историю расширения круга задач, решаемых компьютерами по мере того, как это становилось экономически обосновано. Волны автоматизации производства являются как следствием закона Мура, так и его причиной, поскольку образуют систему с положительной обратной связью.

Эволюция вычислительной техники представляет собой череду мини-революций в технологиях. Периоды постепенного развития сменялись революционными завоеваниями новых рыночных ниш при достижении ценами очередной критической отметки. Как правило, «большинство стратегически переломных моментов возникает в результате значительного (десятикратного) изменения одной из конкурентных сил Портера, влияющих на бизнес»<sup>8</sup>.

Смены технологий сопровождалась изменением форм ведения бизнеса, поскольку каждый раз менялся и потребитель, и характер потребления. Соответственно каждая новая волна компьютерных технологий выводила на сцену новых лидеров цифровой экономики, сумевших вовремя распознать перспективы новых сегментов рынка и сделать соответствующие инвестиции вопреки логике лидеров прошлой волны. При этом в соответствии с приведенными выше оценками лидеры каждой следующей волны осваивали больший, чем предыдущий, рынок и становились лидерами всей компьютерной отрасли.

## **Предыстория: первые вычислительные машины**

Первой задачей, созревшей для автоматизации, стали математические вычисления. Собственно, эта волна автоматизации началась еще в XVII веке с механических арифмометров Шиккарда, Паскаля и Лейбница. Алгоритмы арифметического счета были известны издавна, и их автоматизация просто ждала своего часа.

Вычисления играли особенно большую роль в мореплавании для определения координат судов. С развитием мировой торговли после выхода в XV веке из Средиземного моря на просторы Мирового океана, в том числе в неизвестное доселе Южное полушарие, эта задача встала особенно остро. Отсюда и интерес к астрономическим наблюдениям, практическая польза которых состояла в составлении таблиц эфемерид — положений небесных тел на небе в определенные моменты времени, чтобы можно было ориентироваться не только по звездам, но и по планетам. Эта практическая потребность вызвала к жизни и революцию Коперника, и механику



Ньютона.

Определение долготы требовало точного измерения времени, и часовая индустрия стимулировала развитие точной механики. В арифмометрах же применяли ту же «арифметическую логику», что и в часах, используя шестерни с десятью зубцами, полный оборот которых переводил шестерню следующего разряда на один зубец. Первый арифмометр Шиккарда так и называли: «Считающие часы».

Апофеозом механических калькуляторов стала героическая попытка создания в XIX веке универсальной вычислительной машины английским изобретателем Чарльзом Бэббиджем. Описанием его *аналитической машины*, в частности способов ее программирования, занималась дочь лорда Байрона леди Ада Лавлейс, которую можно считать первым программистом в истории. Проект Бэббиджа был задуман как вычислительный комплекс со своим печатным устройством и предназначался для вычисления и печати навигационных таблиц. Проект финансировался британским адмиралтейством, но так и не был закончен по множеству объективных и субъективных причин. По-видимому, он просто обогнал свое время.

Немногим позже, в конце XIX века, русский математик П. Л. Чебышев создал практичный механический арифмометр, самостоятельно выполняющий все арифметические операции после установки задания. Требовалось только вращать механическую ручку, которая позднее была заменена электромотором.

Такие электрические арифмометры широко применялись в первой половине XX века вплоть до появления электронных вычислительных машин. В частности, оба атомных проекта, американский и советский, были первоначально просчитаны на таких электрических арифмометрах. А для расчета «изделий» требовалось очень много вычислений — численное решение сложных дифференциальных уравнений. По воспоминаниям очевидцев, вычислительные центры в то время представляли собой длинные ряды *вычислителей (computers)* — вооруженных арифмометрами женщин, вычисляющих каждая свою операцию и передающих результат следующему звену этого математического конвейера. Поскольку числа набирались на арифмометрах вручную, во избежание ошибок вычисления дублировались на двух параллельных конвейерах, и ответ принимался только в случае совпадения их результатов[25].

## Электронно-вычислительные машины (национальные лаборатории)

Вычисления стали узким местом при разработке сложной военной техники во времена Второй мировой войны. Например, в штате Лаборатории баллистических исследований в Пенсильвании было несколько сотен женщин-вычислителей, и все равно она не справлялась с потоком вычислений для создания баллистических артиллерийских таблиц, которые требовались после высадки союзных войск в Африке в 1943 году. Именно по заказу этой лаборатории и была создана первая в мире электронно-вычислительная машина — ЭНИАК (ENIAC — аббревиатура от Electronic Numerical Integrator and Computer — электронный цифровой интегратор и вычислитель).

Заказанный в 1943 году, ЭНИАК был запущен лишь осенью 1945 года, уже после окончания Второй мировой войны. Однако, поскольку в то время в США разворачивались работы по созданию термоядерного оружия, он успешно эксплуатировался военными до 1955 года[26].

ЭНИАК весил 30 т, содержал 17 500 электронных ламп и потреблял свыше 170 кВт, или 230 л. с. И этот «табун лошадей» мог за секунду сложить 5000 или перемножить 360 чисел. Иными словами, каждая «лошадь» перемножала действительные числа со скоростью 1,5 FLOPS, затрачивая на каждое перемножение столько же энергии, сколько требуется для подъема 50 кг на 1 м. Но по сравнению с механическими арифмометрами ЭНИАК за секунду производил больше вычислений, чем те — за час, то есть являлся первым суперкомпьютером.

Успех ЭНИАКа, доказавшего работоспособность концепции ЭВМ, дал старт многим последующим проектам с более совершенной *архитектурой фон Неймана* — с использованием двоичной арифметики, активного процессора и программ, хранимых в памяти наряду с данными. Первые суперкомпьютеры создавались каждый в единственном экземпляре и применялись в национальных научных центрах или ведомствах, как правило, в интересах военных. Таким образом, компьютеры освоили первый, самый крупный государственный сектор, где деньги во времена холодной войны и гонки вооружений особо не считали.

Опыт, накопленный в этом секторе, вскоре нашел применение и в частном бизнесе. С 1950-х годов наступила эра серийных коммерческих компьютеров, и в этой области безоговорочно лидировала компания IBM.

## **Mainframes, языки программирования, операционные системы (IBM)**

IBM к тому времени имела уже солидную полувековую историю на рынке

механической обработки информации и, самое главное, обширную клиентскую базу в государственном и корпоративном секторах. В числе основателей компании был Герман Холлерит, изобретатель первого электромеханического табулятора — системы программной обработки баз данных с помощью перфокарт (1890). Табуляторы на порядки ускоряли статистическую обработку данных, в частности при переписи населения. Под этот заказ правительства США и был построен первый табулятор, после чего Холлерит открыл свою фирму, которая оказалась очень успешной и к 1924 году в результате нескольких слияний превратилась в IBM.

Когда во второй половине 1940-х появились первые компьютеры, их рынок (несколько десятков миллионов долларов) был еще слишком мал для IBM, выручка которой в 1950 году составляла \$266 млн. Однако компания под руководством своего нового президента Томаса Уотсона-младшего вовремя распознала потенциал новой техники и организовала серийное производство компьютеров. Первые компьютеры IBM стали выпускаться в 1953 году, а через 10 лет, к 1963 году, ее оборот возрос до \$2,8 млрд, львиную часть которых (\$2 млрд) компания заработала за счет продажи компьютеров<sup>9</sup>.

Следуя стратегии опережающего развития, IBM удерживала лидерство в сегменте крупного бизнеса с 1960-х до 1980-х годов, занимая в этом сегменте 70–80% мирового рынка и заслужив уважительное прозвище «Голубой гигант». Именно в IBM появился первый язык программирования высокого уровня — FORTRAN (FORmula TRANslator), созданный специально для инженерно-вычислительных задач, что видно из его названия. Как ни странно, этот язык все еще жив и в своей узкой области вполне конкурентоспособен.

Еще более крупным вкладом IBM в компьютерную технику стала ее операционная система OS/360. Это был крупнейший коммерческий инженерный проект с бюджетом в половину Манхэттенского [27]. До появления OS/360 каждая модель компьютера имела свою систему команд, и компании испытывали трудности с переходом на новые, более производительные машины. Им приходилось заново переписывать все программное обеспечение. С 1964 года начались поставки IBM System/360 — серии компьютеров разной мощности, но полностью совместимых по программному обеспечению. Эта революционная линейка продуктов вывела IBM в безусловные лидеры мировой компьютерной индустрии.

По сути, System/360 были первыми компьютерами, способными на апгрейд. Покупатели могли начинать с младших моделей по цене от \$250 000 (Model 25) и наращивать вычислительную мощность по мере необходимости, заменяя их старшими (Model 75 стоимостью \$3 млн) вплоть

до самых мощных (Model 195 — \$10 млн и выше). Поэтому наиболее популярной была не покупка, а аренда машины за 1/48 ее цены в месяц. Благодаря такой гибкой ценовой политике серия System/360 пользовалась огромным коммерческим успехом. Только за первые три месяца было подано предзаказов на \$1,2 млрд, а за последующие 5 лет продано 33 000 экземпляров различных моделей.

Независимость программного обеспечения от аппаратного стала первым шагом на пути отказа от вертикальной структуры компьютерного бизнеса, включающего стек очень разных технологий: производство полупроводников, электронных комплектующих, компьютеров из них, операционных систем, системных и прикладных программ. Сначала все это создавалось в каждой фирме самостоятельно, так как помогало удерживать имеющихся покупателей, но по мере усложнения техники сосредоточить все эти компетенции в рамках одной компании стало невозможно. IBM первая ощутила на себе предел сложности своего бизнеса и отступила перед ним.

Проблему роста сложности удалось решить через реформатирование структуры компьютерной индустрии. Постепенно вертикальная структура рынка сменилась горизонтальной, где в каждом элементе технологического стека появились свои производители, а между соседними элементами стека были разработаны стандарты протоколов взаимодействия.

## **Мини-ЭВМ, локальные клиент-серверные сети (DEC)**

Первой результаты такой демонополизации ощутила, естественно, сама IBM, которую во второй половине 1980-х стали теснить производители мини-ЭВМ и персональных компьютеров, пробиваясь во все более дешевые сегменты корпоративного рынка. Как следствие, компьютеры приблизились вплотную к рабочим местам, и началась эра распределенных вычислений на дешевых персональных компьютерах, объединенных в сети. Серверами в таких сетях служили более дорогие мини-компьютеры, а самые дорогие центральные компьютеры, на которых по-прежнему специализировалась IBM, играли все меньшую роль, так как были слишком дороги в обслуживании. На рубеже 1980-х и 1990-х годов IBM начала терпеть убытки, и ей пришлось за несколько лет уволить 170 000 своих сотрудников, чтобы адаптироваться к новой ситуации.

Лидером в сегменте мини-компьютеров была компания DEC (Digital Equipment Corporation), образованная в 1957 году, которая изначально позиционировалась в более дешевом, чем IBM, сегменте. Ее простые и надежные продукты — 16-разрядные PDP (от \$10 000 с 1965 года) и 32-разрядные VAX (от \$100 000 с 1978 года) стали бестселлерами, так как были

дешевы и очень просты в эксплуатации. Они не требовали штата из десятков сотрудников, как мейнфреймы IBM. По объемам продаж DEC вышла на второе место после IBM, продав при этом на порядок больше компьютеров [28].

Однако, сильно потеснив IBM, DEC вскоре сама столкнулась с конкуренцией еще более дешевых, но все более мощных персональных компьютеров, продававшихся в разных конфигурациях за \$1000–5000. Большой выбор поставщиков, гибкость конфигураций, возможность самостоятельного апгрейда на развитом рынке комплектующих и постоянно растущая мощность микропроцессоров вывели персональные вычисления в лидеры компьютерного мира. В итоге в 1998 году DEC была куплена Compaq, крупнейшим производителем PC.

## PC, Internet (Microsoft — Intel)

Появление персональных компьютеров стало возможным после создания микропроцессоров, то есть процессоров, целиком помещавшихся на одном кристалле. Эта технология кардинально уменьшила стоимость вычислений и до сих пор лежит в основе всей современной цифровой техники (рис. 14).



Рис. 14. Вычислительная мощность, доступная за \$1 (<https://www.spaceship.com.au/learn/nvidia-computer-chips-investing-ideas/>)

Лидирующим производителем микропроцессоров была компания Intel,

выпустившая свой первый микроконтроллер Intel 4004 в 1971 году [29]. Вместо заказанной ей серии различных контроллеров было решено изготовить один *программируемый контроллер*, который мог заменить их все. Такой подход быстро завоевал популярность, и Intel приняла стратегическое решение сосредоточиться на производстве микропроцессоров, уступив рынок более просто устроенной компьютерной памяти японским конкурентам. Первые микросхемы были 4-битными и использовались в настольных калькуляторах, но 8-разрядные чипы уже могли работать как с числами, так и с символами. Выпущенный в 1974 году универсальный микропроцессор Intel 8080 при цене \$360 мог заменить компьютер стоимостью несколько тысяч долларов. Первые персональные компьютеры на таких процессорах использовались в основном для игр.

Самым популярным 8-битным компьютером стал Apple II, предлагавший в стильном корпусе цветной дисплей, звук и интерпретатор языка BASIC. Кроме игр, Apple предлагала и приложения для работы, например электронные таблицы VisiCalc, пользовавшиеся большой популярностью у покупателей. Начиная с 1977 года было продано свыше 5 млн экземпляров Apple II разных модификаций.

Рынок микрокомпьютеров рос на 40% в год и в 1979 году превысил планку \$150 млн, обратив на себя внимание «серьезных» компаний, включая IBM. Последняя решила завоевать этот рынок, создав буквально за год свою модель IBM PC на основе стандартных компонент, доступных на рынке. В качестве микропроцессора был выбран только что появившийся 16-разрядный Intel 8088, а операционную систему IBM лицензировала у небольшого стартапа Microsoft. Более того, IBM опубликовала максимально подробное описание своего персонального компьютера, приглашая независимых производителей присоединиться к ее новой платформе. IBM понимала, что в одиночку ей не выиграть на этом новом для нее рынке, и эта ее политика произвела очередную микрореволюцию.

До 1981 года лидеры микрокомпьютерного рынка предпочитали как можно большую часть производства сосредоточивать у себя. С выводом на рынок IBM PC отрасль наконец получила открытый стандарт де-факто. Все независимые производители электронных комплектующих, периферийных устройств и программного обеспечения поверили в модель горизонтального рынка и переключились на новые стандарты.

В итоге IBM стала захватывать рынок невиданными темпами и уже через пару лет более чем в два раза обошла своего главного конкурента, Apple, по объему продаж. Подразделение IBM, отвечавшее за IBM PC, из внутреннего стартапа превратилось в третью по величине компьютерную компанию США по обороту продаж (после самой IBM и DEC).

Впоследствии, однако, IBM столкнулась с неизбежной конкуренцией многочисленных «клонов» ее продуктов с более дешевой сборкой в Юго-Восточной Азии и в конце концов отказалась от борьбы за этот слишком конкурентный рынок. Вся корпоративная культура IBM была настроена на работу с крупным бизнесом, а на новом рынке B2C требовались совсем другие компетенции.

Лидерами эпохи персональных вычислений стали крупнейшие представители нового горизонтального рынка — производители процессоров и программного обеспечения, наиболее сложных и капиталоемких его сегментов. Речь идет, конечно же, о компаниях Intel и Microsoft, продукция которых отлично дополняла друг друга. Чем больше возможностей предлагала новая версия Windows, тем более мощные процессоры требовались от Intel, что обеспечивало стабильный рост обоих членов «альянса Wintel».

Но решающую роль в расширении цифрового рынка сыграло, конечно, превращение Internet в мультимедийную всемирную сеть WWW (World Wide Web), что, в свою очередь, обязано появлению в стане PC нового интуитивно понятного *графического интерфейса* с метафорой информационных объектов — *файлов*, с которыми можно было обращаться как с обычными материальными объектами. Оба эти события относятся к середине 1990-х годов. Если мы посмотрим на рис. 14, то увидим, что к этому моменту компьютеры стали достаточно мощными для работы с графической информацией: за \$1000 можно было получить  $10^6$  FLOPS, необходимых для графического интерфейса пользователя и для работы с графическими браузерами [30]. Мультимедийный Internet, конечно же, привлек к себе гораздо более широкую аудиторию, чем населявшие его ранее программисты, привыкшие к командной строке.

Персональный компьютер, подключенный к Всемирной сети, предоставлял всем желающим поистине неограниченный набор контента, впечатлений и способов самореализации, что сопровождалось соответствующим ростом парка PC и расширением сети Internet. Вызванная открывшейся перспективой первоначальная эйфория вскоре закончилась крахом «доткомов» в начале 2000-х, напомнившим, что в этой новой реальности надо искать новые формы ведения бизнеса. И эти формы не замедлили появиться. Возникли чрезвычайно прибыльные и стабильно растущие интернет-компании, захватившие к нашему времени первые места в мировой таблице о рангах.

## Смартфоны, облачные платформы, ИИ (Apple — Google —

## Microsoft — Amazon — Facebook)

Однако этому предшествовало появление еще одной революционной технологии — мобильной цифровой связи и соответственно мобильного интернета. Когда точкой входа в интернет вслед за персональными компьютерами стали на порядок более дешевые смартфоны, интернет-аудитория стала измеряться уже миллиардами пользователей, которые были на связи не в отдельные часы, а постоянно, где бы они ни находились.

Смартфоны стали неотъемлемым аксессуаром большинства людей, их «экзокортексом» — частью мозга, вынесенной наружу. Их стали приспособлять к решению любых бытовых проблем. Все услуги, за которыми раньше приходилось обращаться в разнообразные офисы, стали доступными онлайн — здесь и сейчас: от справочной информации до получения ипотеки, от заказа пиццы до заказа такси. Началось слияние онлайн- и офлайн-миров.

Переход на новый мобильный формат цифровых устройств потребовал и нового интерфейса пользователя. Клавиатура и мышка плохо соотносятся со смартфоном. Современные окошки приложений и мини-клавиатуры являются, скорее всего, временным компромиссом. Ведь количество доступных приложений уже измеряется миллионами, и нет никакой надежды даже узнать о большинстве из них, не говоря об их использовании в подходящих ситуациях. Очевидно, что людям нужны электронные помощники — агенты, представляющие их интересы в этом сложном мире бесчисленных доступных сервисов. Необходимы посредники, понимающие как наши нужды, так и все возможности сервисов, способных эти нужды удовлетворить.

До сих пор в человеко-машинных интерфейсах пользователь был активной стороной. Он в явном виде отдавал компьютеру команды: что, как и когда тому надо сделать. Агентский интерфейс будущего будет отличаться проактивностью и самостоятельностью. Электронный секретарь должен будет сам угадывать, что посоветовать пользователю в тот или иной момент, потому что он помнит обо всех обязательствах пользователя, ничего не забывает и имеет доступ к прецедентам всех облачных сервисов, то есть во многих отношениях его знания более обширны, чем у пользователя. Если мы хотим быть эффективными, нам придется все в большей мере полагаться на советы наших персональных ассистентов.

Соответственно, новой формой бизнеса является создание платформ, предоставляющих доступ ко множеству сервисов, и персональных электронных ассистентов в качестве интерфейса к этим платформам. Причем, чем более дружественным будет общение между пользователем и



его ассистентом, тем большими будут конкурентные преимущества этой платформы. В идеале персональные агенты должны общаться на разных человеческих языках, понимать просьбы и пожелания пользователей в зависимости от текущего контекста, различать настроение пользователя и даже обладать чувством юмора. Иными словами, очередное поколение человеко-машинного интерфейса предполагает создание искусственного интеллекта.

Заметим, что к началу эры смартфонов в 2010-х годах, согласно рис. 14, за \$1000 стала доступна вычислительная мощность, сравнимая с человеческим мозгом, —  $10^{11}$  FLOPS (рис. 11). И как достижение в середине 1990-х критического порога  $10^6$  FLOPS открыло дорогу мультимедийному интернету, так переход через новый рубеж  $10^{11}$  FLOPS знаменует начало эры искусственного интеллекта. Так что начавшийся в 2010-х и продолжающийся поныне бум искусственного интеллекта вполне закономерен.

Тот факт, что существующий сегодня ИИ все еще очень далек по своим возможностям от человеческого, следует отнести исключительно к неготовности современной науки к этому вызову. Создание искусственного интеллекта требует глубокого синтеза технических и гуманитарных знаний, и именно эта непростая задача встает сегодня во весь рост перед наукой. Мы должны объединить компьютерные науки с математикой машинного обучения, знаниями о механизмах работы мозга и тонкостях человеческой психики. Мы должны построить действующие модели искусственной психики, объясняющие человеческое поведение, и, поняв, как она устроена, наделить ею... роботов. Только в этом случае роботы и программные агенты смогут вписаться в человеческую цивилизацию «на наших условиях».

## Аналогия с промышленной революцией

В этой главе мы описали начало перехода от индустриального уклада к цифровому, который ожидает нас в не столь отдаленном будущем. Возникает вопрос: а на какой стадии этого перехода мы сегодня находимся? Можем ли мы указать, например, соответствующую стадию промышленной революции, чтобы понять, хотя бы из этой аналогии, масштаб грядущих перемен?

По мнению автора, появление искусственного интеллекта в его современном виде *слабого ИИ* соответствует созданию первых, еще очень несовершенных паровых двигателей. Поясним, чем обусловлена такая аналогия. Паровой двигатель был замыкающей технологией первой

промышленной революции. С его появлением стало возможным массовое промышленное производство. Авангардом первой промышленной революции выступало ткацкое производство. Ткацкие станки появились задолго до этого и постепенно совершенствовались, пока наконец не превратились в автоматы, которые можно было приводить в действие внешними источниками энергии, независимыми от человека. Поначалу таким источником служили водяные мельницы, и первые мануфактуры возникали по берегам рек, что ограничивало масштаб производства и было неудобно с точки зрения логистики. Появление парового двигателя позволило организовать фабричное производство произвольного масштаба непосредственно в местах скопления рабочей силы, то есть в городах. Одна паровая машина могла обслуживать несколько цехов с различными видами станков, обеспечивая энергией весь производственный цикл. Снабжение городов и фабрик топливом также осуществлялось с помощью все того же парового двигателя, поставленного на рельсы. Таким образом, эта замыкающая технология разом сняла все преграды для организации массового промышленного производства, и Англия в короткий период смогла обеспечить тканями весь остальной мир[31].

Аналогичным образом масштабы цифровой экономики до последнего времени ограничивались скоростью переноса знаний от людей к компьютерам с помощью программистов. Машинное обучение снимает это ограничение скорости роста компьютерной составляющей коллективного мышления, позволяет создавать массовые интеллектуальные сервисы, обеспечивающие своими услугами весь мир. Как в свое время паровые двигатели, машинное обучение расшивает узкое горло, сдерживающее массовый переход на новый технологический уклад.

Современное *глубокое обучение* можно уподобить первым громоздким паровым двигателям Ньюкомена с их чудовищно низким КПД. Обучение глубоких нейросетей сегодня требует больших массивов данных и еще больших вычислительных ресурсов (пропорциональных квадрату объема данных). Это является естественным следствием их универсальности: нейросети способны реализовать любой алгоритм, аппроксимировать любую непрерывную зависимость выходов от входных данных. Изначально в них не заложены никакие априорные предположения о характере моделируемых данных, а чем меньше у нас априорных знаний, тем больше требуется примеров, чтобы передать эти общие знания алгоритму.

Люди же, благодаря своему интуитивному пониманию законов внешнего мира, способны угадывать решения незнакомых задач буквально по нескольким примерам. Так, в известных тестах по измерению IQ средний человек способен, как правило, угадать четвертую картинку по трем

имеющимся. Интеллект ассоциируется у нас именно со способностью схватывать на лету, то есть с эффективностью упаковки данных в знания<sup>10</sup>. И в этом смысле эффективность современных нейросетей оставляет желать лучшего.

Но мы помним, что в ходе промышленной революции КПД двигателей неуклонно повышался, как росла и их доступность. Качественное изменение произошло с повсеместным внедрением электричества и электродвигателей, с появлением общедоступной «энергии из розетки» (произошла так называемая вторая промышленная революция). Аналогичный момент в цифровой революции наступит с появлением *сильного ИИ*, по своим интеллектуальным способностям не уступающего человеческому. Только сильный ИИ способен бесшовно вписаться в человеческое общество, стать общедоступным «интеллектом из розетки» в качестве новой технологической основы цифрового уклада.

В том, что это произойдет, нет никаких сомнений, вопрос только — когда. В следующей главе мы попытаемся ответить на него, рассмотрев более внимательно драйверы современного этапа цифровой революции. Как проникает в нашу жизнь и в каком направлении развивается искусственный интеллект сегодня? Кто и насколько сильно заинтересован в появлении сильного ИИ? Ведь если общество действительно нуждается в какой-то технологии, если на нее имеется растущий платежеспособный спрос, ее появление неизбежно. Как говорил Ф. Ницше, «кто знает *зачем*, найдет любые *как*».

## Резюме

В главе 1 мы обсудили основные этапы развития человеческой цивилизации с древности до наших дней, вписав современную технологическую революцию в этот единый процесс. В этой главе мы сосредоточились на истории развития цифровых технологий, быстрыми темпами формирующих новый технологический пакет цифрового уклада. Выяснилось, что буквально на наших глазах во втором десятилетии XXI века сформировалась замыкающая технология этого технологического пакета — искусственный интеллект, в его первой, пока еще несовершенной, версии. В следующей главе мы проследим, как эта технология вписывается в современную экономику, за счет чего увеличивает производительность труда и каким требованиям должна будет удовлетворять для окончательного перехода к цифровому укладу. Ограничения слабого ИИ позволят нам сформировать «техническое задание» для сильного ИИ,

основываясь на тех проектах по его созданию, которые сегодня уже объявлены. Предполагая, что эти проекты будут обеспечены достаточными ресурсами со стороны заинтересованных инвесторов, мы сможем относительно обоснованно предсказать наше ближайшее будущее.

## ГЛАВА 3

# Экономика цифровых платформ

*Кибернетика — это прежде всего наука о достижении целей, которых простым путем достичь невозможно.*

СТАНИСЛАВ ЛЕМ

*Кто владеет информацией, тот владеет миром.*

НАТАН РОТШИЛЬД

### Экономика больших данных: персонализация

Каждый следующий технологический уклад характеризуется более высоким уровнем производительности труда и соответственно жизни. Скачок в производительности при переходе к индустриальному укладу был обусловлен машинным производством, требующим резкого роста энергопотребления. Соответствующий пакет технологий распечатывал новый обильный и дешевый ресурс — запасы ископаемого топлива.

Но сегодня, как мы знаем, энергетика растет вровень с остальной экономикой. За счет какого же ресурса и каким образом новый цифровой уклад обещает кратно увеличить производительность труда и обеспечить растущему населению Земли высокое качество жизни? Ранее мы утверждали, что новым обильным ресурсом цифровой экономики являются информация и знания. Настала пора с фактами в руках обосновать это утверждение.

Главное отличие нового цифрового уклада от предыдущего, индустриального, — *персонализация* всего и вся. Индустриальный уклад основан на массовом производстве товаров и услуг. Его технологическая основа — гигантские конвейеры, запрограммированные на массовый выпуск *типовых изделий* как в сфере материального производства, так и в информационном пространстве — в средствах массовой информации.

Машинное производство повысило производительность труда и расширило рынок за счет удешевления продукции. Но ориентация на

типовые решения сама по себе ограничивает размеры рынка. Никакое увеличение номенклатуры планомерно выпускаемых изделий не способно удовлетворить все мыслимые сочетания индивидуальных предпочтений. Рынок персонализированных продуктов намного шире.

Однако в индустриальном укладе отсутствуют технологии, позволяющие определять эти предпочтения и удовлетворять их экономически обоснованным способом, а именно эффективно организовать сбор индивидуальных заказов от пользователей и производство соответствующих штучных товаров и услуг. И в том и в другом случае речь идет об эффективной обработке больших массивов информации, что, в свою очередь, невозможно без массового создания алгоритмов, управляющих этими процессами. А ключевой технологией массового производства алгоритмов обработки данных, как мы выяснили, является машинное обучение.

Поскольку «топливом» для него служат данные, они сегодня становятся «новой нефтью». Соответственно в лидеры цифровой экономики выбиваются так называемые *цифровые платформы*, имеющие доступ к максимально большим объемам данных, точнее, способные сами генерировать необходимые им данные в процессе общения с пользователями. Платформами их называют потому, что они позволяют зарабатывать деньги большому количеству сторонних производителей, предоставляя им услуги по приему от потребителей и доставке им персональных заказов, основанных на детальном знании предпочтений всех, кто пользуется услугами платформы.

Труд тоже товар, который востребован на рынке, как и всякий другой. В индустриальном обществе работник является частью гигантского конвейера и должен, как и другие его части, соответствовать определенным профессиональным стандартам. Любые, пусть даже уникальные, способности за рамками стандартов воспринимаются как нежелательные отклонения. Соответственно множество талантов, не нашедших себе применения, в индустриальном обществе пропадают. Однако они могут обрести свое место в мире платформ, обслуживая бесконечно разнообразный спрос. Так, на платформе App Store независимые разработчики предлагают ее пользователям более 1 млн приложений для iOS, генерируя \$50 млрд выручки в год. Таким образом, платформы организуют и обслуживают «длинный хвост» пользовательских предпочтений, с одной стороны, и «длинный хвост» производителей — с другой.

Наконец, last but not least, с развитием искусственного интеллекта массовое конвейерное производство будет сменяться гибкими

производственными системами, каждая со своим собственным интеллектом. Совокупный интеллект человечества станет расширяться за счет бесчисленных программных агентов и роботов, каждый из которых будет обладать сверхчеловеческими способностями в сфере своих компетенций. Иными словами, начнет происходить повышение степени разделения труда. Таким образом, объем знаний и соответственно количество и качество управленческих решений в экономике будет возрастать. Это приведет к снижению издержек, более экономному расходованию ресурсов и, следовательно, к увеличению общего объема товаров и услуг при тех же энергозатратах. Если индустриальный уклад повысил энергетическую мощь цивилизации, то цифровой уклад повысит ее интеллектуальную мощь и, как следствие, энергоэффективность экономики. Каждый джоуль будет контролироваться и расходоваться с максимальной эффективностью, предотвращая тем самым надвигающийся глобальный экологический кризис. Сегодняшний повсеместный тренд на энергоэффективность — один из признаков перехода к новому укладу.

Цифровые платформы — это типовые ячейки новой цифровой экономики, приходящие на смену транснациональным компаниям индустриальной эпохи. Ведущие цифровые платформы быстро превращаются в «естественные монополии» благодаря известному сетевому эффекту, характерному для любых сервисов, основанных на машинном обучении. А именно: чем больше у сервиса пользователей, тем больше у него данных и тем выше его качество, что еще больше увеличивает количество его пользователей, и т.д. Возникает положительная обратная связь, приводящая к появлению лидера данного сегмента и его экспоненциальному росту. Именно за счет этого эффекта не так давно возникшие цифровые платформы превратились в крупнейшие глобальные компании, некоторые из которых перешагнули рубеж капитализации в \$1 трлн.

## **Первые цифровые платформы: интеллект из больших данных**

Цифровые платформы использовали машинное обучение с момента своего возникновения во времена «бума доткомов». Так, появившийся в 1998 году Google мгновенно завоевал популярность пользователей качеством своего поиска, основанного на алгоритме рейтингования всех страниц интернета с помощью анализа гигантского графа ссылок между ними. Это позволяло Google выводить на первую страницу наиболее авторитетные источники

информации по любому вопросу, чего не могли предложить остальные поисковики, работающие лишь по ключевым словам. Так машинное обучение помогло Google объединить в своем индексе предметные знания всех производителей контента в интернете — через ссылки на авторитетные источники в html-страницах.

Для эффективного использования своего гигантского поискового индекса Google внедрил инновационную технологию массовой распределенной обработки данных (Map Reduce), создав первый поисковый суперкомпьютер, распределенный по нескольким дата-центрам. Но своим главным успехом Google обязан другой инновации — AdWords, превратившей его в цифровую платформу для рекламного бизнеса. С ее помощью Google предоставил доступ к глобальной аудитории мелким рекламодателям, вплоть до индивидуальных предпринимателей. Он сформировал и обслуживает «длинный хвост» рекламного рынка, тем самым увеличивая его объем на \$135 млрд в год (1/4 мирового рынка рекламы)<sup>1</sup>. Масштабирование сервиса достигается за счет использования предметных знаний самих рекламодателей: аукционный алгоритм автоматически отбирает наиболее интересные для аудитории объявления и определяет их справедливую цену.

Таким образом Google сформировал новую, гораздо более эффективную, чем средства массовой информации, инфраструктуру рекламного рынка, одновременно и снизив издержки, и увеличив емкость рынка. И все за счет алгоритмов обработки больших данных. Выручка на одного работника в Google составляет \$1,4 млн — в 7 раз выше, чем в среднем по США. Всю основную работу делают алгоритмы.

Другим примером первого поколения цифровых платформ является Amazon, оператор «длинного хвоста» в ретейле. Начиналось все в 1994 году с уникального предложения «бесконечной книжной полки» — книжного разнообразия, недоступного для офлайновых магазинов. Предложение пришлось пользователям по душе, бизнес быстро расширился, и сегодня через Amazon в США совершается уже половина всех онлайн-покупок.

«Секретным оружием» Amazon является его рекомендательная система, помогающая пользователям ориентироваться во всем этом разнообразии. И основана она опять-таки на алгоритмах машинного обучения, превращающих обычные учетные данные — кто что купил — в персональные рекомендации: кому что может быть интересно. В результате — большее удовлетворение покупателей и расширение рынка для продавцов, ведь теперь любой нишевой товар может найти своего покупателя.

В отличие от Google, Amazon приходится иметь дело с реальными



товарами и организовывать логистику в реальном мире, поэтому выручка на одного сотрудника в Amazon ненамного выше средней. Но это пока! Потому что Amazon сегодня активно инвестирует в искусственный интеллект и робототехнику, вводя в строй полностью автоматизированные склады, экспериментируя с доставкой товаров роботами и дронами и открывая офлайн-магазины Amazon Go без продавцов и кассиров. Этот тренд мы более подробно обсудим ниже.

В целом же в 1990-х и 2000-х годах машинный интеллект фокусировался на извлечении из больших объемов данных структурированных знаний, которые можно использовать для персонализации разного рода услуг конечным пользователям. Огромные массивы структурированных знаний накоплены в проприетарных *графах знаний* цифровых платформ Google Knowledge Graph, Amazon Evi, Microsoft Satori и используются во многих продуктах, в частности голосовыми помощниками Google Assistant, Amazon Alexa и Microsoft Cortana.

## Современные цифровые платформы: сенсорный интеллект

Голосовые помощники стали новацией 2010-х годов в качестве нового интерфейса пользователя в эпоху смартфонов. Они, как и безлюдные магазины Amazon Go, обязаны своим появлением новому поколению алгоритмов машинного обучения — так называемому *глубокому обучению* искусственных нейронных сетей. Технологический прорыв в машинном обучении в 2010-х, известный как *революция глубокого обучения*, связан с достижением компьютерами критической производительности  $10^{11}$  FLOPS, сравнимой с человеческой, по доступным ценам (рис. 14).

Это, конечно, не означало автоматического появления искусственного интеллекта, как он когда-то задумывался его отцами-основателями. Для этого у человечества пока что банально не хватает соответствующих знаний, о чем мы еще поговорим в главе 5. Мощности сегодняшних суперкомпьютеров достигают  $10^{16}$  FLOPS, однако это до сих пор не привело к появлению сильного ИИ. Но кое-какие разработки 1980-х и 1990-х годов, для которых в свое время просто не хватало вычислительных мощностей, чтобы выйти на уровень отдельных когнитивных способностей, сравнимый с человеческим, «выстрелили» именно в этот момент. Речь идет об обучении некоторых типов искусственных нейронных сетей, разработанных для работы с изображениями (сверточные нейронные сети) и временными сигналами (сети с долговременной памятью).

Оказалось, что просто за счет увеличения количества слоев в таких

(глубоких) нейросетях и увеличения объема данных для их обучения, для чего теперь имелись вычислительные мощности, качество распознавания картинок и звука может достигать человеческого уровня. В итоге на протяжении 2010-х годов, благодаря технологиям глубокого обучения, машины, говоря простым языком, научились видеть и слышать не хуже человека.

Соответственно появилась масса новых возможностей для замены человека машинами — там, где люди работали «умными сенсорами»: контролерами, охранниками, операторами колл-центров, и в других профессиях с относительно простой бизнес-логикой. Отсюда — появляющиеся сегодня проходные без охранников, магазины без продавцов, такси без водителей, безлюдные колл-центры и голосовые помощники в смартфонах и умных колонках.

И это еще только начало. Потенциальный рынок приложений слабого ИИ, наделенного сверхчеловеческими сенсорными возможностями по доступным ценам, чрезвычайно большой. Он касается самых массовых профессий — продавцов, кассиров, водителей и т.д., поэтому обещает большой экономический эффект. Оказывается, люди, по крайней мере многие из них, не так уж и незаменимы.

Отложив обсуждение этой важнейшей проблемы до следующей главы, зададимся пока одним чисто практическим вопросом о драйверах развития ИИ. Есть ли экономическая целесообразность в дальнейшем совершенствовании технологий машинного обучения или бизнес может ограничиться достигнутыми успехами, сосредоточив усилия на бизнес-инновациях, то есть на освоении уже открывшихся благодаря слабому ИИ рынков? От ответа на этот вопрос зависит, в частности, объем вложений в разработку сильного ИИ и соответственно время появления последнего.

Представляется, что в силу упомянутого выше эффекта положительной обратной связи тем, кто хотел бы освоить новые рыночные ниши с помощью существующего слабого ИИ, нет смысла ждать дальнейших инноваций. Надо спешить, так как в каждой такой нише действует принцип «победитель забирает все». Так что в 2020-х годах нас ожидает инвестиционный бум новых доткомов, использующих разработанные в 2010-х технологии глубокого обучения.

Драйвером сильного ИИ выступают сегодня уже сформировавшиеся цифровые платформы, обслуживающие глобальные потребительские рынки, те, для кого голосовые помощники представляют собой важнейший канал общения с их пользователями. Ведь для массового пользователя диалог на естественном языке — максимально удобный способ бытового общения со смартфонами, навигаторами, умными колонками и другой

домашней электроникой. Именно через своих электронных ассистентов — Google Assistant, Alexa, Cortana и им подобных — современные платформы продвигают свои сервисы в массы. Умные колонки сегодня — один из наиболее быстро растущих сегментов потребительской электроники, поэтому в сегменте голосовых ассистентов наблюдается очень сильная конкуренция. Крупнейшие интернет-компании тратят существенную часть доходов на совершенствование своих голосовых ассистентов.

## Разговорный интерфейс — путь к сильному ИИ

Но несмотря на все эти вложения, качество разговорного интерфейса до сих пор оставляет желать лучшего. При этом и распознавание речи у голосовых помощников, и качество их синтетической речи уже вполне удовлетворительны. Их можно и дальше улучшать, в том числе за счет камер, распознающих мимику, жесты и эмоции пользователя. Проблема не в распознавании речи, а в понимании ее смыслового содержания и умении вести осмысленную беседу, то есть в создании того, что можно назвать *разговорным интеллект*ом.

Выясняется, что одного сенсорного интеллекта для общения на естественном языке вовсе недостаточно. Распознавание речи лишь самый поверхностный слой настоящего разговорного интеллекта, перевод звукового сигнала в текст. Конечно, благодаря глубокому обучению в машинной обработке текстов тоже имеются большие успехи. Это видно хотя бы по довольно высокому качеству современного машинного перевода. Однако оказывается, что излагать одно и то же содержание на другом языке и делать умозаключения на основе этого содержания — задачи разного уровня сложности. Первое требует наличия у машины *семантического пространства*, в котором отражается содержание предложений, тогда как второе предполагает умение оперировать в этом семантическом пространстве, прокладывая в нем осмысленные траектории — рассуждения.

Разговорный интеллект должен обучаться формировать осмысленное поведение в семантическом пространстве, понимая на каждом шаге цели и позиции всех участников диалога, включая свои собственные. Значит, он и сам должен уметь ставить перед собой какие-то цели и добиваться их достижения. Иными словами, машинный перевод еще можно отнести к задачам распознавания образов — отображению исходного текста в его семантическое представление, а последнего — в текст на другом языке. Разговорный же интеллект относится к гораздо более сложному классу задач — обучению адаптивному целесообразному поведению, включая

рациональное мышление как разновидность поведения в семантическом пространстве и умение осознанно манипулировать смыслами. А это — прямой путь к настоящему сильному ИИ.

Современное машинное обучение пытается найти решение этой задачи с помощью все тех же глубоких нейросетей. У всех на слуху недавние победы машин во всевозможных стратегических играх, от культовой древней игры го до новомодных StarCraft 2 и Dota 2, где требуется реагировать на действия соперников в реальном времени и строить гипотезы о том, что происходит в ненаблюдаемых областях игрового поля. Эти достижения демонстрируют способность глубоких нейросетей формировать полезные поведенческие навыки и строить выигрышные стратегии в виртуальных мирах. Но современные алгоритмы пока что не позволяют машинному интеллекту вырваться из этих виртуальных миров в реальный. Обучение нейросетей происходит сегодня слишком медленно, и соответствующий «жизненный опыт» за приемлемое время удастся набрать лишь в виртуальном мире за счет существенного ускорения темпа игры. Эти алгоритмы невозможно перенести на обучение роботов в реальном мире, где у них не будет столько времени на обучение и столько виртуальных жизней, которыми заплачено за неудачные решения.

К тому же, если вернуться к разговорному интеллекту, у нас пока нет виртуальных миров для оттачивания разговорных навыков. Ассистентам надо учиться рассуждать и вести диалоги, а для этого — пробовать самим генерировать варианты ответов в различных сценариях. Привычное обучение на больших корпусах готовых диалогов здесь не подходит. Надо, чтобы кто-то оценивал качество каждой реплики в бесчисленных ветвящихся вариантах развития диалогов, из которых лишь очень немногие могут присутствовать в обучающей выборке. А на это пока что способны только живые люди и лишь в реальном времени.

Резюмируя, можно сказать, что существующий уровень машинного интеллекта явно не устраивает лидеров цифровой революции, которые остро нуждаются в разговорном интеллекте человеческого уровня и будут вкладываться в его создание, чтобы не отстать от конкурентов. Рынок разговорного интеллекта удваивается каждые 2 года и в 2020 году должен был превысить \$12 млрд. Так что спрос на сильный ИИ в современной экономике уже сформировался, и мы понимаем, кому и для чего он сегодня нужен. Следовательно, этим уже имеет смысл заниматься, хотя еще совсем недавно после всех пережитых разочарований задача построения сильного ИИ всерьез не воспринималась, а разговоры о нем считались ненаучными.

## Цифровые платформы будущего: интеллект роботов

Разговорный интеллект «здесь и сейчас» нужен тем, кто сегодня зарабатывает в основном за счет персонализации рекламы. Что собой представляют те же электронные витрины Amazon, как не набор рекламных баннеров, сделанных с учетом предпочтений каждого покупателя? Между тем весь рекламный рынок, \$560 млрд, составляет лишь 0,5% мирового ВВП. Так что рыночный потенциал для применения ИИ гораздо шире, чем рынок умных программных ассистентов.

Но для радикального расширения области применений ИИ программным агентам предстоит выйти из виртуального мира в реальный, превратиться в роботов. Сенсорный интеллект агентов предстоит дополнить *моторным интеллект* роботов — способностью активно взаимодействовать с миром, иметь предиктивную модель внешнего мира и своих действий в нем, чтобы, например, понимать, что стул или чашку можно переставить с места на место, а шкаф или машину лучше и не пытаться, или что, если чашку наклонить, ее содержимое выльется, или что двери иногда открываются легко, а иногда нет и, если они заперты, в них надо не ломиться, а стучаться. Последнее уже относится к области *социального интеллекта* — понимания того, как принято себя вести в обществе, чего от тебя ждут в тех или иных ситуациях, «что такое хорошо и что такое плохо».

Все эти элементарные знания, известные любому ребенку, невозможно запрограммировать, им надо обучаться. И обучаться активно, методом проб и ошибок, как это делают дети. А для этого у роботов должна быть *искусственная психика* с врожденным любопытством, настроенная, как и у детей, на постоянное обучение, чтобы как можно скорее набраться опыта и научиться достигать своих целей в этом сложном и непредсказуемом поначалу мире. Искусственная психика роботов должна быть настолько универсальна, чтобы она могла обеспечить эффективное обучение всем видам интеллекта: сенсорному, моторному, социальному и разговорному. Ведь и дети обучаются ходить, говорить и вести себя правильно практически одновременно.

В математике бывает, что иногда легче решить задачу в более общей постановке, которая лучше отражает суть проблемы. Возможно, разработка искусственной психики — тот самый случай: вместо множества специализированных систем, обучающихся разным задачам по разным лекалам (подход, принятый сегодня в глубоком обучении), лучше разработать единый общий интеллект (Artificial General Intelligence, AGI).

## Резюме

Итак, цифровая экономика — это экономика платформ, рост влияния которых на экономику мы наблюдаем с начала XXI века. Именно платформы больше всех заинтересованы в использовании и развитии технологий ИИ. Сегодня основным драйвером развития сильного ИИ служит спрос на разговорный интеллект агентов. Через какое-то время появится еще более мощный драйвер — спрос на искусственную психику роботов.

Мы рассмотрим подходы к построению сильного ИИ чуть позже, в главе 5. Пока же следует обсудить силы, которые будут препятствовать его появлению, потому что кроме тех, кто заинтересован в инновациях, всегда есть и те, кто им сопротивляется, причем не без оснований. Потому-то каждая крупная технологическая революция сопровождается революциями социальными. Смена технологического ландшафта закономерно приводит к трансформации социальной надстройки.

## ГЛАВА 4

# Социальные проблемы цифровой революции

*Кто кем повелевает? Технология нами или же мы — ею?*

СТАНИСЛАВ ЛЕМ

*Перед нами открываются потрясающие возможности,  
замаскированные под неразрешимые проблемы.*

ДЖОН ГАРДНЕР

### Проблемы реальны

Переход к индустриальному укладу, как мы знаем, привел к драматическим изменениям в жизни людей. По мере того как человека заменяли станки, исчезали старые и появлялись новые профессии. Крестьянские массы, теснимые тракторами, двинулись в города. Множество людей лишилось привычных источников существования — настала эпоха, красочно описанная Карлом Марксом в первом томе «Капитала». XX век с утверждением индустриального уклада наглядно продемонстрировал, к каким последствиям могут привести и классовая борьба за дележ возросшего общественного пирога, и имперские схватки за передел мировых рынков.

Было бы наивно ожидать, что предстоящий переход к цифровому укладу пройдет плавно, без подобного рода турбулентностей. Как и прежде, нас ожидают новые «разборки» между трудящимися, теснимыми роботами, и их работодателями, стремящимися увеличивать прибыль акционеров за счет сокращения рабочей силы. Эти отношения будут выясняться в политическом поле с активным участием государства, то есть бюрократии, имеющей свои собственные интересы в отношении искусственного интеллекта. Аналогично в международных отношениях странам, обладающим передовыми технологиями искусственного интеллекта, будет трудно удержаться при разделе новых рынков от силовых угроз, в том числе с использованием боевого ИИ.

Наконец, в ожидании сильного ИИ появится абсолютно новая «экзистенциальная» угроза роду людскому в целом — столкновение с интеллектом, который грозит в конце концов превзойти человеческий. И тогда вопрос о том, кто кем управляет и будем ли мы хозяевами своей судьбы в прекрасном новом мире, встанет во весь рост.

Все эти проблемы потребуют своего решения, и от этих решений будут во многом зависеть судьбы человечества. Далее в этой главе мы последовательно обсудим изменения отношений между трудом и капиталом, гражданами и государством, государствами между собой и, наконец, между людьми и искусственным интеллектом с появлением последнего.

## ИИ-безработица

Первое столкновение людей с искусственным интеллектом произойдет очень скоро, еще в эпоху слабого ИИ. Слишком много людей сегодня заняты тем, что Карл Маркс называл «частичным трудом», когда от человека требуется лишь какой-то небольшой набор простых навыков.

До недавнего времени компьютерам были доступны только легко формализуемые задачи: вычисления, поиск в базах данных, форматирование документов и т.д. Соответствующие программы существенно облегчали работу менеджеров, при этом не заменяя их самих, так как касались лишь небольшой части выполняемых ими служебных обязанностей. С этим, в частности, связан хорошо известный парадокс, что постепенное насыщение компьютерами экономики не приводило до сих пор к существенному росту производительности труда<sup>1</sup>.

Появление сенсорного интеллекта человеческого уровня, способности компьютеров распознавать объекты и ситуации в реальном мире не хуже людей, многое меняет. Под угрозой оказались представители массовых профессий из разряда «контролеров» (продавцы, охранники, кассиры) или даже простейших «контролеров», как в случае с шоферами[32]. В целом к 2030 году в США может быть автоматизировано около 40% рабочих мест<sup>2</sup>.

С появлением сильного ИИ ситуация еще более усугубится. Вслед за безлюдными сборочными цехами, складами и магазинами появятся безлюдные заводы, робофермы и строительные площадки, где дома будут печатать гигантские 3D-принтеры. Люди станут массово вытесняться из сферы производства и сервисов, как крестьяне вытеснялись с полей в ходе промышленной революции. Обслуживанием покупателей займутся программные агенты и роботы с развитым разговорным интеллектом,



безграничными знаниями и столь же безграничным терпением.

Возникает естественный вопрос: а кто в таком случае будет этими самыми покупателями? Владельцы и управляющие роботизированными активами? Немногочисленные представители творческих профессий, уцелевшие в конкуренции с роботами? А как насчет остальных? Они же не станут молча курить в сторонке!

Здесь на самом деле скрыты два разных вопроса:

- Как поступать с людьми, вытесняемыми роботами здесь и сейчас?
- Какая роль в будущей цифровой экономике останется за людьми, а что отойдет роботам?

Первый вопрос в принципе решаем в виде той или иной разновидности налога на роботов. Например, робот, нанимаемый вместо человека, пожизненно «платит» последнему достойную пенсию через отчисления в специальный общественный пенсионный фонд[33]. Это несколько замедлит темпы роботизации, так как внедрять роботов станет экономически выгодно лишь тогда, когда они будут не просто эффективнее человека, а, грубо говоря, вдвое более эффективными. Зато можно будет обойтись без никому не нужных социальных потрясений. Для тех, кто не привык бездельничать и хотел бы найти себе новое занятие по душе, повысив заодно и свой заработок, будут организованы курсы по освоению новых профессий. Тех, кого пенсионерский стиль жизни устроит, будут ждать туризм и индустрия развлечений. Эти отрасли экономики станут соответственно расти и расширять рынок труда.

Здесь мы уже начинаем отвечать заодно и на второй вопрос — о роли и месте людей в будущем. Ведь кроме производства и распределения продуктов есть еще и постоянно расширяющаяся сфера человеческого общения. Все-таки люди обычно предпочитают общаться со «своими», и роботы в этом плане вряд ли смогут составить им существенную конкуренцию.

Кроме общения между собой за людьми останется очень важная социальная роль «воспитателей роботов», состоящая в передаче им наших человеческих ценностей. Ведь успех будущей человеко-машинной цивилизации целиком зависит от того, насколько хорошо роботы будут разбираться в людях. Как мы уже упоминали и о чем будем более подробно говорить в следующей главе, обучаться персональные агенты и роботы будут через подкрепление, то есть одобрение своих действий. Но со стороны кого? Конечно же, людей, через постоянные поощрения правильных поступков в самых разных формах — от оплаты услуг в соответствии с их качеством до обычных слов благодарности.

К этому же разряду следует отнести и будущую *цифровую демократию*. Сегодня люди «играют в политику» лишь изредка и в массе своей как любители, так как на это у них попросту нет времени: все заняты своей основной работой. Но в будущем определение общественного мнения по любому вопросу станет, во-первых, предельно упрощенным, а во-вторых, превратится в норму жизни, поскольку только таким образом люди смогут сохранить за собой управление своей судьбой, оставив за роботами исполнение их социальных заказов. Цифровая демократия станет новым интерфейсом между распределенным человеческим и машинным разумом, через который будет достигаться компромисс между индивидуальными желаниями и общественными возможностями, станут определяться приоритеты, цели и долгосрочные планы в цифровой экономике. А также будет обеспечиваться безусловное исполнение коллективно принятых решений. Здесь мы естественным образом переходим к вопросу о роли государства в эпоху сильного ИИ.

## ИИ-патернализм

С одной стороны, ситуация с ИИ-безработицей показывает, что без активного вмешательства государства в поддержание общественной стабильности в переходную эпоху нам не обойтись. С другой стороны, с искусственным интеллектом связано множество страхов и фобий в контексте государства как «большого брата». Действительно, для авторитарных государств, которых, к слову, сегодня в мире больше, чем демократических, ИИ предоставляет великолепный инструмент тотального контроля над своими гражданами. Например, в Китае с его системой социального рейтинга право вмешательства государства в частную жизнь вообще не ставится под сомнение. Но и в менее авторитарных странах власть «в интересах народа» всегда найдет легальные способы, как поставить ИИ на службу своим интересам [34]. Ведь у государственной бюрократии и силовых структур есть и собственные амбиции при разделе общественного пирога.

Технологии ИИ позволяют государству не просто следить за настроениями и поступками граждан, но и активно влиять на общественное мнение в пользу тех или иных групп интересов, например с помощью армии искусственных личностей в социальных сетях. А поскольку соответствующие разработки силовых ведомств засекречены, они с большим трудом поддаются общественному контролю, если вообще поддаются.

Еще более опасны попытки бюрократии переводить стрелки

общественного недовольства с больной головы на здоровую, списывая свои неудачи и провалы на разного рода внутренних или внешних врагов. Это нисколько не способствует выявлению и решению реальных проблем и в долгосрочной перспективе лишь усугубляет нестабильность в обществе. Более того, атмосфера «осажденной крепости» содействует безудержной гонке вооружений, опасность которой с появлением боевого ИИ возрастает многократно.

Действительно, возникновение и быстрое развитие ИИ часто сравнивают с изобретением атомного оружия, которое впервые поставило под вопрос физическое выживание человечества. Начинаясь на наших глазах гонка за искусственным интеллектом очень напоминает атомную и ракетную гонку вооружений второй половины XX века. Автономные боевые роботы, управляемые по глобальной оперативной сети искусственными стратегами, действительно способны качественно изменить баланс сил в пользу государств, имеющих доступ к соответствующим технологиям.

Пока что военные успокаивают общественность тем, что, мол, окончательное решение всегда остается за человеком. Но все прекрасно понимают, что это временный вариант, продиктованный эпохой слабого ИИ, так как именно преимущество в скорости и качестве принятия решений станет главным достоинством боевого сильного ИИ[35]. Военные в ходе начавшейся гонки будут вынуждены постепенно передавать ему право принятия смертоносных решений, поднимаясь от уровня отдельных боевых единиц, как это уже происходит с самонаводящимися ракетами, до боевых тактических групп и далее — на всё более высокие оперативные и стратегические уровни. А поскольку вся логика принятия подобных решений будет основана на обработке гигантских объемов разнородной информации, недоступных человеческому разуму, перепроверить правильность решений искусственных стратегов в условиях ограниченного времени будет попросту невозможно.

При таком развитии событий повышается вероятность того, что человечество окажется втянутым в масштабную ядерную катастрофу в силу естественного стремления искусственных стратегов достичь поставленных военным руководством «национальных целей». Ведь недоверие к противнику и стремление превентивно нейтрализовать его коварные планы — профессиональные качества военных, от которых они неизбежно перейдут и искусственным стратегам.

Мало того, что интересы политического руководства могут кардинально расходиться с интересами граждан. При постановке задачи искусственным стратегам самое сложное — не определить цели, а сформулировать ограничения, при которых достижение этих целей все еще имеет смысл. В

известном рассказе У. Джекобса «Обезьянья лапка» этот древний талисман неизменно исполнял все желания своих хозяев, но делал это формально, без учета негласных ограничений — тех самых человеческих *ценностей*. Так, в ответ на желание очередного хозяина талисмана получить 200 фунтов чек на эту сумму был вручен ему на следующий же день, но в качестве компенсации за жизнь внезапно погибшего в результате несчастного случая любимого сына. Боевой ИИ в конечном итоге может привести мир к ситуации, когда ядерные кнопки окажутся в таких «обезьяньих лапках».

Гипертрофированная борьба с внутренними и внешними врагами в ущерб остальным функциям государства подобна аутоиммунным заболеваниям или в острой форме — цитокиновому шторму, когда защитная система организма начинает поражать не только его врагов, но и свои собственные органы, в том числе жизненно важные. Подобное развитие событий может быть смертельно опасным и для общественных организмов. Достаточно вспомнить развал Советского Союза при всей мощи его армии, способной сокрушить любого агрессора. Излишняя концентрация на внешних угрозах заслонила гораздо более реальную проблему — потерю привлекательности режима в глазах своих граждан из-за систематического игнорирования их нужд со стороны государства. А это, в свою очередь, стало следствием недостатка обратных связей в закостенелой административно-командной системе управления.

Соответственно лекарство от подобных аутоиммунных общественных заболеваний следует искать в развитии *цифровой демократии*, интенсивных сетевых коммуникациях, нацеленных на выявление и гармонизацию разнообразных нужд всех слоев населения. У разных общественных групп есть свои интересы и представления о том, на что и в каких пропорциях следует тратить общие ресурсы. И общество должно уметь вырабатывать такую политику, которая минимизирует степень народного недовольства.

В индустриальном обществе эта задача зачастую решалась в присущей ему парадигме массового производства, то есть навязывания всему обществу единой идеологии, подчинения всех меньшинств монолитному большинству с использованием чисто силовых методов. Цифровой уклад подразумевает совсем другой подход: использование реального разнообразия общества с помощью новых цифровых технологий, не в последнюю очередь — ИИ.

Например, уже сегодня имеется техническая возможность с минимальными издержками организовать массовые электронные голосования, обеспечив доверие к их результатам. (Применяя технологии шифрования с открытым ключом, когда каждый респондент может по своему частному ключу проверить, правильно ли учтен его голос в

электронной системе голосования, не раскрывая при этом тайну голосования.) Сюда же, видимо, можно отнести и постоянно развивающиеся технологии распределенных реестров, математически гарантирующие общественную защиту любых транзакций.

В цифровом укладе государство сможет использовать национальные цифровые платформы для структурирования общественного мнения вокруг приоритетных вопросов, организации по ним качественной экспертизы и выработки совместных решений, вплоть до возможности самостоятельного распределения всеми экономическими субъектами своих налогов по основным бюджетным статьям, причем по математически обоснованным схемам оптимального фондирования публичных нужд, например используя *квадратичное фондирование* (quadratic funding), наилучшим образом учитывающее истинные предпочтения всех граждан<sup>3</sup>, вместо нынешних сборов пожертвований для больных детей с помощью СМС-сообщений на фоне кулуарного распределения бюджетов в парламентских комитетах.

На платформах цифровой демократии интеллектуальные агенты смогут «дирижировать» процессами принятия коллективных решений, взвешивая мнения экспертов с учетом уровня их компетентности и знакомя граждан с наиболее весомыми аргументами за и против по любому вопросу подобно тому, как Google сегодня отбирает из сети наиболее релевантную информацию, а Amazon рейтингует и рекомендует товары на своих электронных витринах. Технологии сильного ИИ помогут людям формировать их личные мнения по любому вопросу в процессе диалога, в котором искусственный интеллект сможет суммировать разные точки зрения, находить в них скрытые противоречия и формулировать аргументированные советы исходя из понимания индивидуальных предпочтений и ценностей каждого гражданина.

Таким образом, ИИ несет с собой не только угрозы, но и способы борьбы с ними. Каждую угрозу мы должны воспринимать как вызов и пользоваться окном возможностей, открывающихся благодаря новым технологиям.

## ИИ-порабощение

Какими бы ни были опасения обывателей против засилья элит, все они меркнут на фоне страха перед возможным порабощением людей всемогущим сильным ИИ, управляющим армиями послушных роботов. Мы автоматически наделяем искусственный интеллект хорошо знакомыми нам человеческими мотивами, например стремлением к доминированию, помноженными на его сверхчеловеческие способности по достижению

своих целей.

Приписывание искусственному интеллекту человеческих мотивов, надо признаться, имеет под собой все основания. В конце концов, «с кем поведешься — от того и наберешься». Только вот насколько правомерно проводить демаркационную линию «порабощения» именно между биологическим и искусственным разумом? Мы интуитивно считаем, что все, что происходит внутри нашего мозга, — это естественный интеллект, а то, что происходит в микрочипах, — искусственный. Но так ли это на самом деле? Ведь, как известно, человеческий разум имеет коллективную природу. Все наше мышление, начиная с раннего детства, формируется обществом. Все наши мысли, даже если они и порождены нашим мозгом, были в своих предпосылках заложены в нас другими людьми в процессе коллективной *мыследеятельности*. Любой наш мыслительный акт происходит в явном или незримом диалоге с другими людьми, с оглядкой на их будущую оценку наших решений.

Но ведь то же самое будет происходить и в будущей искусственной психике робота — те же сомнения в оценке его действий окружающими, те же привитые воспитанием ценности, лежащие в основе принятия решений его искусственной психикой. Люди не способны заложить в роботов ничего другого, кроме собственных ценностей! Так что если уж и говорить о порабощении, то речь идет скорее о несовершенствах нашего нынешнего общества: излишнем доминировании элит в разделе общественного пирога, силовом подавлении индивидуальных свобод и других пережитках прошлых эпох. ИИ выступает лишь в качестве еще одного инструмента для существующих в обществе *институтов* — системы правил, убеждений, норм и организаций<sup>4</sup>.

И если мы сумеем преодолеть в человеческом обществе идеологию вражды, деления на «своих» и «чужих», на которых моральные ограничения не распространяются, если сможем организовать цифровые институты гармонизации интересов разных групп без опоры на силу, то эти же ценности станут гарантом нашего мирного сосуществования и в рамках будущей человеко-машинной цивилизации. Ибо «без соответствующих институтов индивиды не всегда сознают, в чем состоит общественное благо, либо не всегда мотивированы в стремлении к нему»<sup>5</sup>. И это касается как биологических, так и искусственных индивидов.

Здесь уместно вспомнить о «законе техногуманитарного баланса», согласно которому для выживания цивилизации необходимо обеспечить регуляцию общественных отношений, адекватную степени развития технологий<sup>6</sup>. Проще говоря, «сила» цивилизации должна быть уравновешена ее «мудростью». Чем мощнее и потенциально опаснее

технологии, тем тоньше и сложнее должны быть механизмы достижения общественного согласия, способные подавлять агрессию и предотвращать конфликты. При этом «мудрость», как правило, отстает от «силы», общественное бытие обгоняет общественное сознание, которое, совсем по Ницше, остается и ныне «человеческим, слишком человеческим». Как справедливо заметил А. П. Назаретян, наше будущее зависит от ответа на вопрос: «Успеет ли земной разум усовершенствовать качество самоконтроля в соответствии с ускоряющимся технологическим ростом, прежде чем разрушительные последствия станут необратимыми?»

Таким образом, вопрос о нашем все более активном участии в общественной жизни, в выработке целей и ценностей общества — ключевой вопрос нашего будущего. И если рассматривать стратегию как взгляд на настоящее из желаемого будущего, то цифровая демократия — это именно та проблема, на решении которой мы должны сосредоточиться сегодня. Потеряем ли мы в будущем контроль над своими судьбами? Кто будет решать, как нам жить, к чему стремиться?

А кто это делает сегодня? Если и сегодня это решаем не мы, если решения принимаются кем-то за нас, пусть и от нашего имени, то нам, собственно, нечего будет и терять с приходом сильного ИИ. Он лишь усугубит ту ситуацию, которая к тому моменту сложится в обществе. Причем этот момент может быть гораздо ближе, чем мы до сих пор полагали. «Возможно, наши земные жены сегодня рожают потенциальных богов, которым будут доступны какие-то формы бессмертия и космического господства. Или они рожают поколение самоубийц, которые окончательно обрушат здание Ноосферы...»<sup>7</sup> Об этом мы поговорим в оставшейся части книги.

## Резюме

Подытожим наш краткий обзор рисков, связанных с развитием ИИ. Во-первых, следует признать серьезность и нетривиальность возникающих проблем. Мы должны понимать, что нам предстоит создать не просто сильный ИИ, а безопасный, дружественный сильный ИИ. И по всей видимости, это гораздо более сложная проблема, поскольку она не сводится к чисто техническим решениям, а подразумевает активное участие граждан в выработке согласованной концепции общественной безопасности. Концепции эти могут отличаться от страны к стране. Соответственно, потребуются и какие-то международные конвенции в этой области, особенно в связи с опасностью новой гонки вооружений.

Во-вторых, с технической точки зрения желательно, чтобы безопасность ИИ была встроена в саму архитектуру будущих операционных систем роботов, подобно «трем законам робототехники» Айзека Азимова. Мы подробно обсудим эту концепцию в следующей главе, посвященной технологиям сильного ИИ.

Наконец, last but not least, система коллективной безопасности социума, состоящего из людей, роботов и агентов с искусственной психикой, должна быть обеспечена технологически на основе постоянного взаимного контроля. Мы с самого начала должны понимать, что настанет момент, когда люди не смогут сами контролировать решения более мощных искусственных стратегов, поэтому нам следует заранее создавать механизмы саморегуляции сообщества машинных интеллектов с учетом наших интересов. В дополнение к правильным «социальным инстинктам» в будущей искусственной психике мы должны будем разработать и механизмы взаимного контроля всех членов социума, включая ИИ, подобно тому как технология блокчейн автоматически обеспечивает взаимный контроль транзакций в сети. Будущие цифровые платформы должны будут иметь систему встроенной коллективной безопасности, как в последних поколениях ядерных реакторов со встроенной на физическом уровне системой безопасности.

Итак, мы разобрали, кому и зачем нужен сильный ИИ, а также каким требованиям он должен удовлетворять, чтобы улучшить, а не ухудшить нашу жизнь. Самое время перейти к обсуждению вопроса, насколько реальны наши планы по его созданию и на каком временном горизонте он может появиться.



## ГЛАВА 5

### Путь к сильному искусственному интеллекту

*Стирая грань между человеком и машиной, мы стираем грань между людьми и богами.*

АЛЕКС ГАРЛЕНД

*Всякая технология, в сущности, просто продолжает естественное, врожденное стремление всего живого господствовать над окружающей средой.*

СТАНИСЛАВ ЛЕМ

#### Как определить и измерить интеллект?

Ранее мы уже касались вопроса о том, что подразумевали под искусственным интеллектом его отцы-основатели и как это понимание со временем эволюционировало. Пришла пора рассмотреть поставленный вопрос подробнее, поскольку от нашего представления о сути интеллекта зависит и направление наших усилий по его созданию. В этой главе представлена авторская точка зрения на эти вопросы, которая наверняка расходится с мнениями многих специалистов. Что ж, к заветной вершине ведет множество путей, у каждого из нас — свой, и мы пока не знаем, какой из них окажется кратчайшим.

Начнем с того, что есть два типа определений: ориентированные на прошлое — то, чем люди до сих пор занимались в данной области, и ориентированные на будущее — то, чем «на самом деле» надо заниматься. Нас, естественно, больше занимает последний тип, так называемые *рабочие определения* ИИ. Естественно, что формулировки таких определений у разных авторов могут сильно различаться, особенно в переломные моменты, в одном из которых мы сегодня и находимся. Об этом свидетельствует недавняя горячая дискуссия по поводу определения искусственного интеллекта<sup>1</sup>, инициированная статьей д-ра Темпльского университета Пей Вана (Pei Wang), предложившего собственную рабочую

формулировку ИИ<sup>2</sup>.

Дискуссия разгорелась в основном вокруг термина «интеллект», который лежит на перекрестье многих наук — философии, психологии, нейрофизиологии и, естественно, машинного интеллекта. Первоначальная идея разработчиков последнего состояла в дополнении исторически сложившихся подходов к интеллекту новыми идеями из недавно возникших компьютерных наук. Норберт Винер в конце 1940-х предложил использовать кибернетику как науку об общих принципах управления и самоорганизации в живых и технических системах. Участники Дартмутского семинара 1956 года, на котором, собственно, и родился термин «искусственный интеллект», попытались поднять эту планку еще выше — создать искусственный разум. Общим в обоих подходах было стремление строить действующие модели, следуя принципу «понять — значит воссоздать».

Пионеры ИИ, как выяснилось, явно переоценили возможности тогдашней науки по моделированию «алгоритмов разума». У них не хватало ни теоретических представлений о том, как эти алгоритмы могут быть устроены, ни эмпирических знаний о реальных механизмах работы мозга. Это привело, как мы помним, к краху завышенных ожиданий, снижению первоначальной планки и переориентации исследований на слабый ИИ. Однако с тех пор ситуация качественно изменилась и в области алгоритмов, и в науках о мозге. Так что, возможно, пришла пора вновь переосмыслить исследовательскую программу когнитивных наук и ИИ-сообщества.

В науках о мозге прогресс был связан с появлением новых экспериментальных методик, позволивших заглянуть внутрь живого мозга: функциональной магнитно-резонансной томографии, оптико-генетических методов и многих других. Накоплен огромный массив эмпирических данных о том, что и когда происходит в мозге. При этом явно ощущается дефицит идей, обобщающих эти данные на уровне алгоритмов работы мозга и его вычислительной архитектуры<sup>3</sup>. По крайней мере, реалистичной действующей модели психики до сих пор так и не создано, несмотря на многолетние исследования в этом направлении<sup>4</sup>. Это и значит, что у нас до сих пор нет целостного понимания того, как мозг работает.

В области «алгоритмов разума» мы наблюдаем начавшуюся в 2010-х годах революцию глубокого обучения, в ходе которой впервые удалось с помощью универсального метода градиентного обучения искусственных нейросетей решать многие классы когнитивных задач не хуже человека. В ИИ-сообществе по этому поводу в последние годы наблюдается некоторая эйфория и возрождается осторожный оптимизм по поводу реальности разработки сильного ИИ. Многие лидеры революции глубокого обучения

ставят перед собой эту цель (DeepMind, OpenAI и другие). При этом, по крайней мере на словах, признается, что какие-то идеи они надеются почерпнуть из наук о мозге<sup>5</sup>. Однако по сложившейся традиции накопленный массив знаний о мозге используется в прикладном ИИ в минимальной степени.

Таким образом, наукам о мозге не хватает объединяющей накопленный эмпирический материал модели человеческой психики, а разработчикам ИИ недостает идей о том, как может быть устроена искусственная психика. Налицо предпосылки для налаживания их более тесного взаимодействия<sup>6</sup>.

Существующие глубокие нейросети можно рассматривать как подсистемы искусственной психики, которые можно обучить решению множества самых разных задач, но только по одной, а не всех сразу. Если обученную нейросеть начать обучать новой задаче, она будет забывать старые навыки, поскольку разные задачи решаются разными наборами одних и тех же настроечных параметров нейросети. Это известная проблема «катастрофического забывания» в нейросетях, отличающая слабый интеллект от сильного. Настоящий интеллект должен уметь накапливать знания и приобретать новые навыки, не забывая старые. А для этого нужно объединить множество нейросетей в единую искусственную психику со своей архитектурой, определяющей роли и взаимодействие своих подсистем.

Наделение интеллекта психикой связано еще и с тем, что, как выяснилось в ходе упомянутой выше дискуссии, большинство ведущих специалистов в области ИИ согласны с мнением, что родовым свойством интеллекта является способность к *самостоятельному* решению как можно более широкого круга задач. Сформулировать идею можно по-разному. *Интеллект* — это<sup>7</sup>:

- способность порождать новые алгоритмы для сколь угодно широкого круга задач;
- умение добиваться своих целей в постоянно меняющихся условиях внешней среды;
- способность решать неограниченный круг задач при ограниченных ресурсах;
- способность к целенаправленному адаптивному поведению, и т.д.

Все эти определения тесно связаны между собой, подчеркивая разные стороны одного и того же свойства интеллекта: его субъектность, самостоятельность и креативность в рамках так называемого *агентского подхода*<sup>8</sup>.

При этом большинство исследователей ограничивают интеллект рамками индивидуальной психики агента. Мы, однако, считаем, что индивидуальное мышление, по крайней мере в случае человека, невозможно объяснить вне рамок коллективного мышления человечества. Человеческий разум — это не продукт индивидуального мышления, а результат развития человеческой культуры. Усложнение нашего разума есть результат постоянного углубления системы общественного разделения труда, в результате чего интеллект человека, воспитанного людьми, качественно отличается от интеллекта Маугли, воспитанного волками, при одинаковом устройстве мозга. Таким образом, связь между мозгом и разумом опосредована процессом воспитания.

Соответственно, далее мы будем придерживаться следующих определений:

- *интеллект* — это алгоритм обучения целенаправленному поведению агента;
- *разум* — это алгоритм обучения целесообразному коллективному поведению агентов.

Тем самым мы различаем индивидуальный интеллект и коллективный разум. Оба представляют собой обучающиеся адаптивные системы, но их целевые функции и алгоритмы обучения разные.

Заметим также, что от разумных агентов мы ожидаем постоянного накопления опыта в процессе обучения. Именно способностью к обучению решению все новых задач, то есть к усложнению своего поведения, разумные агенты отличаются от простых адаптивных физических систем с обратными связями типа, например, термостата. Даже сколь угодно сложные экспертные системы не обладают интеллектом в нашем понимании, если в них не заложена способность к обучению, — они являются лишь очень продвинутой формой справочника.

Этим определениям, однако, не хватает одного важного свойства — количественного измерения уровня интеллекта. Понятно ведь, что уровни интеллекта человека и шимпанзе существенно различаются, да и люди, как известно, обладают разными IQ. В этой связи очень интересно решение, предложенное Франсуа Шолле из Google, создателем известной библиотеки Keras. Интеллект он определяет как способность к *эффективному обучению* и предлагает измерять его как КПД конверсии получаемой из внешней среды информации в знания о том, как достигать своих целей [36]. В наших обозначениях:

$$IQ = \frac{\Delta C}{\Delta I},$$

где:  $\Delta C$  — приращение знаний, повышающих адаптацию агента в изменившихся условиях, то есть усложнение алгоритма его поведения с учетом вновь полученной информации;

$\Delta I$  — новая информация, не содержащаяся в предыдущей истории взаимодействия со средой.

Выше мы уже отмечали, что современным алгоритмам глубокого обучения как раз не хватает такой эффективности. По этой причине они до сих пор и не могут выйти из виртуальных миров, где можно ошибаться миллионы раз, в реальный мир, где каждая ошибка может стать последней.

В дополнение к формальному определению меры интеллекта Ф. Шолле предложил набор задач «Abstraction and Reasoning Corpus» для тестирования машинного интеллекта, по форме напоминающий классические тесты на IQ (например, дополнить по трем картинкам четвертую). Таким образом проверяется способность агента использовать минимальное число обучающих примеров для понимания сути новой задачи и нахождения ее решения. Неудивительно, что современные глубокие нейросети, жадные до данных, пока не способны пройти этот тест.

Заметим, что тест на машинный IQ разрабатывался в предположении, что задачи, с которыми столкнется будущий искусственный интеллект, будут похожи на те, с которыми люди имеют дело в повседневной жизни. У машинного интеллекта предполагается наличие интуитивного понимания элементарных физических и геометрических закономерностей, которым ребенок обучается на ранних стадиях своего развития (понятий причинности, сохранения, непрерывности, симметрии и т.д.), наряду со способностями к счету и абстрактному мышлению, приобретаемыми позднее. Иными словами, предполагается, что искусственная психика роботов должна быть настроена на тот же круг задач, что и человеческая [37]. Может, тогда и архитектура искусственного интеллекта должна соответствовать архитектуре человеческого мышления?

Тем самым ИИ на новом этапе своего развития возвращается к старой идее моделирования естественного интеллекта для углубления нашего понимания природы интеллекта и построения искусственного разума с учетом полученных знаний. Такая совместная исследовательская программа классических наук о разуме и разработчиков машинного интеллекта сегодня необходима и тем и другим.

## Исследовательская программа по созданию сильного ИИ

Мы определили индивидуальный интеллект и коллективный разум как особый класс алгоритмов, способных увеличивать свою сложность при взаимодействии со средой, совершенствуя способность добиваться своих целей в постоянно меняющихся условиях. Именно такие алгоритмы являются предметом машинного обучения, поэтому оно служит наиболее адекватной парадигмой для исследования интеллекта и разума.

Так что машинное обучение играет для понимания разума такую же роль, как теория эволюции — для понимания жизни. При этом машинное обучение трактуется нами как фундаментальная наука о той части природы, которая научилась учиться, в противовес физике, которая изучает косную материю, неспособную к обучению. В таком понимании теория эволюции — это и есть теория обучения живого вещества (пользуясь терминологией В. И. Вернадского), объясняющая причины появления всех форм жизни во всем их удивительном многообразии.

Таким образом, известный тезис Феодосия Добржанского: «Ничто в биологии не имеет смысла, кроме как в свете теории эволюции»<sup>2</sup> можно перефразировать так: «Ничто в науках о разуме не имеет смысла, кроме как в свете машинного обучения». Поэтому именно машинное обучение способно объединить усилия представителей разных наук для совместного продвижения в понимании разума, как человеческого, так и искусственного. Каким образом? Предоставив единую научную парадигму, математический язык, на котором им всем было бы удобно общаться. Разные науки о мозге и мышлении используют настолько разные подходы, что возникает проблема их совмещения друг с другом, и это обстоятельство препятствует плодотворному обмену идеями между различными научными школами.

Нейрофизиология описывает мозг как мыслящую машину и стремится понять устройство этой машины. В «физике мозга», как и везде в физике, объяснение предполагает в первую очередь выявление причинно-следственных связей, поэтому очень показательны работы пионера нейрофизиологии И. М. Сеченова, объяснявшего мышление сложной иерархией рефлексов: «Определенными внешними влияниями вызываются последовательно ряды ассоциированных мыслей, и конец рефлекса вытекает логически из сильнейшей»<sup>10</sup>.

Психология, напротив, описывает мышление телеологически, исходя из внутренних целей организма. Наше мышление стремится опосредованно, через взаимодействие с внешним миром, удовлетворить врожденные потребности и приобретенные ценности человека. С этой точки зрения

мышление определяется не столько внешними обстоятельствами, сколько нашими внутренними целями, как биологическими, так и социальными. Лучше всего иллюстрирует эту позицию знаменитая метафора Л. С. Выготского: «Облака мыслей, гонимые ветрами мотивов, проливаются дождем слов»<sup>11</sup>. Логика сосредоточивается на законах сознательного символического мышления («облаках мыслей»), психолингвистика — на том, как они «проливаются дождем слов», психология же больше интересуется подсознательными «ветрами мотивов». В любом случае в центре внимания оказывается не физика устройства — мозга, а модели мышления, только в описательной, словесной форме, а не как действующие модели с четкими алгоритмами мышления.

Возникает извечный вопрос: как соотносятся между собой причинно-следственные связи и целесообразное поведение, то есть физическое и психическое? С одной стороны, если любое движение мысли определяется механизмами работы мозга, то как можно говорить о свободе воли? С другой стороны, только язык психологии способен хоть как-то объяснять ход наших мыслей и мотивы наших поступков. Конечно, детерминизм и телеология не противоречат друг другу. Это два способа понимания мира — на языке траекторий динамических систем или их аттракторов. Однако детерминистское описание сложной системы (физика мозга) ничего не говорит нам о ее целях, а знание целей — об алгоритмах их достижения.

Теория машинного обучения обеспечивает недостающую связь между принципами причинности и целеполагания за счет изменения постановки задачи. Физика рассматривает систему как данность и спрашивает, как эта система работает. Машинное обучение, напротив, интересуется, как создать систему, обладающую заданными свойствами. Внимание исследователей здесь фокусируется не на «морфологии», а на «эмбриологии» сложных систем.

Соответственно машинное обучение интересуют не столько «быстрые» алгоритмы работы уже готовой системы, сколько «медленные» алгоритмы ее самосборки, постепенного увеличения сложности системы при ее взаимодействии с миром. И эти последние (алгоритмы обучения) формулируются как раз в терминах оптимального управления, то есть достижения системой каких-то целей.

Таким образом, в машинном обучении цели встраиваются во вполне детерминистские алгоритмы обучения. Понимая, какие из них породили данную сложную систему, мы будем знать цели, которые она преследует. Алгоритмы обучения, как правило, гораздо проще результата обучения и позволяют нам не только создать сложную систему, но и «понять» ее, то есть объяснить, почему ее устройство нацелено на достижение определенных

целей, примиряя тем самым причинно-следственное и телеологическое описание системы, притом что распутывание запредельно сложной сети причинно-следственных связей в уже сформировавшейся системе никак не продвигает нас в понимании логики ее работы.

Машинное обучение меняет наше понимание машины как устройства, исполняющего заложенные в него готовые алгоритмы. Алгоритмы обучения, даже относительно простые, способны в процессе взаимодействия со средой породить сколь угодно сложные алгоритмы целенаправленного (то есть разумного) поведения, способного адаптироваться к произвольно меняющимся внешним обстоятельствам и изменять мир в соответствии со своими целями. Машины, оказывается, способны в процессе обучения ставить себе промежуточные цели и находить пути их достижения, преодолевая различные препятствия и вырабатывая собственные ценности, то есть нормы поведения, помогающие достигать поставленных целей. Это новое понимание машины, привнесенное теорией машинного обучения, открывает широкие перспективы в изучении разума во всех его проявлениях.

Итак, машинное обучение предлагает нам следующую программу исследования и создания сложных самообучающихся систем: выявление или разработка алгоритмов их обучения (причинно-следственное описание) как решение задач оптимизации (достижения целей) путем увеличения сложности системы (накопления знаний).

Последний момент особенно важен: простые алгоритмы обучения способны породить сколь угодно сложное поведение. Физика, в которой отсутствует понятие обучения, не способна объяснить появление таких сложных систем, как, скажем, жизнь или разум. Для этого нам приходится обращаться к машинному обучению. Соответственно устройство и принципы работы такой сложной системы, как мозг, тоже можно понять только в свете теории обучения, когда мозг рассматривается не как готовый механизм, но как результат длительного процесса обучения. Это нацеливает нас на выявление в мозге механизмов, реализующих алгоритмы его обучения.

Именно на это, к слову, нацелена *теория функциональных систем* П. К. Анохина, предлагающая, по существу, алгоритм обучения с подкреплением как системообразующий принцип организации мозга<sup>12</sup>. Именно такие алгоритмы наиболее актуальны в современном машинном обучении, связанном с роботами, агентами и сильным ИИ.

Машинное обучение предлагает нам практический путь к пониманию мозга и разума, основанный на принципе «понять — значит воссоздать». Мы никогда не поймем наше мышление, если не сможем воссоздать его в



виде искусственной психики роботов. Искусственный интеллект — это новый экспериментальный полигон для наук о разуме, гораздо более удобный и прозрачный, чем традиционные опыты на животных и людях. На этом полигоне можно гораздо быстрее отлаживать и развивать наши модели мышления.

Так, например, в последние годы была практически решена проблема машинного зрения, причем настолько хорошо и надежно, что на дорогах уже появились автономные автомобили. Алгоритмы глубокого обучения формируют иерархию все более абстрактных признаков, помогающих распознавать ситуации и моделировать пространственные сцены, заодно помогая нам понять и общие принципы работы сенсорного интеллекта.

Вслед за умением ориентироваться в мире приходит черед управлению поведением роботов и освоению ими навыков абстрактного мышления<sup>13</sup>. Возникает масштабный платежеспособный спрос на создание искусственной психики роботов. Современное машинное обучение пока что не смогло найти решения этой задачи.

В нашем понимании именно науки о мозге способны помочь преодолеть этот важный технологический барьер. Надо только суметь взглянуть на мозг с точки зрения машинного обучения — отвлечься от субстрата мозга и сосредоточиться на алгоритмах его обучения. Особенно важна вычислительная архитектура мозга, скопировав которую мы как раз и сможем получить искусственную психику роботов.

## Обратная инженерия мозга

Какова роль наук о мозге в предлагаемой общей исследовательской программе? Кратко ее можно сформулировать как обратную инженерию мозга, а именно — реконструкцию его вычислительной архитектуры.

Такая постановка задачи для нейрофизиологов не нова. Схожую исследовательскую программу предлагал, например, Дэвид Марр в 1970-х годах. Напомним в этой связи, что Марр различал три уровня описания сложных систем:

- вычислительная архитектура — функциональное назначение подсистем в системе;
- алгоритмический уровень — алгоритмы достижения целей подсистем;
- уровень реализации — механизмы, реализующие эти алгоритмы.

Нейрофизиологам непосредственно доступен лишь нижний уровень описания — как конкретные нейроны или нейронные ансамбли

взаимодействуют друг с другом. Как правило, большинство исследований так и остается на этом описательном уровне.

Реже удается разгадать алгоритмы работы каких-то подсистем мозга и понять логику их конструкции. Например, работу мозжечка можно понять, представив его как гигантский персептрон Розенблатта<sup>14</sup>, а гиппокамп неплохо описывается моделью ассоциативной памяти Хопфилда<sup>15</sup>. Наконец, сегодня мы в основном понимаем и принципы работы основной подсистемы нашего мозга — неокортекса, содержащего сотни тысяч самоорганизующихся карт признаков, в которых кортикальные колонки формируют и распознают различные категории<sup>16</sup>.

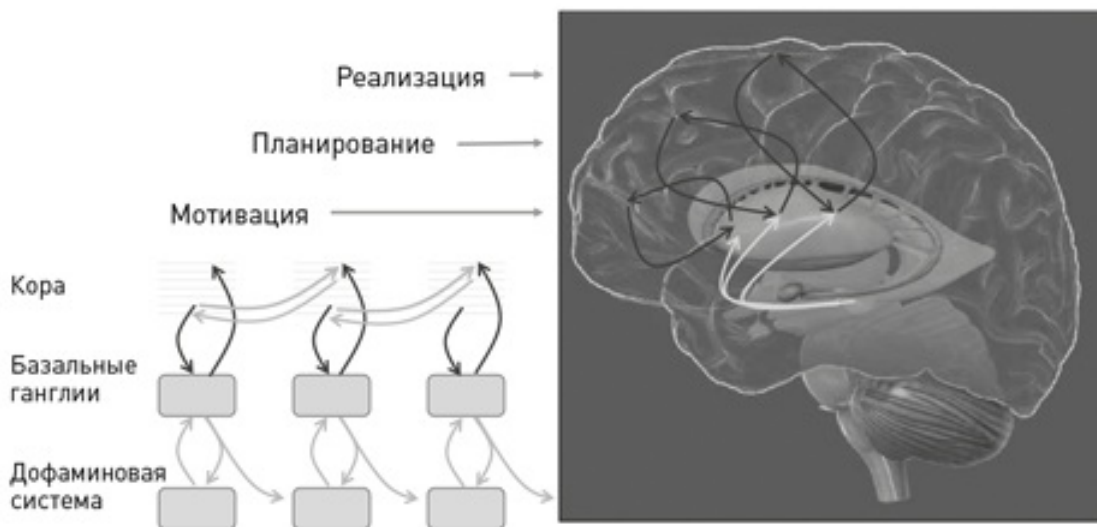
Однако истинное понимание принципов работы мозга возможно лишь с выходом на более высокий архитектурный уровень его целостного описания. А именно: выявление назначения и способов взаимодействия основных подсистем мозга — коры, базальных ганглий, таламуса, мозжечка, стволовых структур и т.д. — с пониманием алгоритмов их работы и обучения. Недаром Дэвид Марр отмечал, что, «хотя верхний уровень по большей части игнорируется, именно он является наиболее важным»<sup>17</sup>. Обратный инжиниринг вычислительной архитектуры мозга является, с нашей точки зрения, центральной задачей нейрофизиологии в рамках предлагаемой исследовательской программы.

В этой связи особое значение, с точки зрения автора, имеет архитектура кортико-стриарной системы мозга приматов, управляющей нашим поведением<sup>18</sup>. Приматы интересны тем, что архитектура мозга у них отличается от остальных млекопитающих: количество нейронов в их мозге пропорционально его массе, тогда как у остальных млекопитающих — массе<sup>2/3</sup>. Иначе говоря, у приматов число нейронов растет с ростом массы мозга гораздо быстрее, чем у остальных млекопитающих. В результате у больших приматов, включая людей, в мозге намного больше нейронов, чем у других млекопитающих с аналогичным по весу мозгом. Поэтому-то наш мозг и обладает такими уникальными вычислительными возможностями. Если бы мы не были приматами, чтобы иметь такое же, как у нас, число нейронов, наш мозг должен был бы весить 30 кг!<sup>19</sup>

В качестве примера на базе предложенной автором реконструкции архитектуры мозга приматов<sup>20</sup> в лаборатории когнитивных архитектур МФТИ в настоящее время разрабатывается модель искусственной психики ADAM (Adaptive Deep Autonomous Machine) «по образу и подобию» человеческой. Подчеркнем, что именно обратный инжиниринг кортико-стриарной системы помог нам преодолеть существенный технологический барьер и предложить схему обучения иерархическому управлению. До сих пор эта проблема не поддавалась решению<sup>[38]</sup>.

Почему бы нам теперь не начать создавать и более реалистичные модели психики различных животных, с тем чтобы воспроизвести их поведение в реальных экспериментах? Машинное обучение впервые предоставляет нам способ воссоздать психику животных и тем самым убедиться в том, что мы ее действительно понимаем. А там уже недалеко и до понимания человеческой психики.

### УПРАВЛЕНИЕ ПОВЕДЕНИЕМ В МОЗГЕ



**Рис. 15.** Кортико-стриарная система мозга, управляющая целесообразным поведением (Alexander G. E. et al. (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience* 9(1): 357–381)

## Моделирование психики

Итак, понимание вычислительной архитектуры мозга действительно полезно для создателей действующих моделей искусственной психики. А с появлением таких моделей у нас возникает уникальная возможность тестирования различных теоретических представлений о том, как устройство мозга определяет поведение животных и человека, то есть наведения мостов между нейрофизиологией и психологией, а затем перехода к психолингвистике и логике.

Планирование и проведение экспериментов с искусственной психикой объединенными командами представителей разных наук как раз и составляет основное содержание практических проектов в рамках предлагаемой исследовательской программы. Основная задача —

воспроизвести с помощью постепенного усложнения искусственной психики эволюцию когнитивных способностей млекопитающих от грызунов к приматам и далее — к человеку.

Начинать естественно с самого простого: воспроизвести основные поведенческие паттерны «искусственной мышки». Примером могут быть опыты с поиском пищи в лабиринтах, открыванием кормушек с помощью различных приспособлений и другими элементами дрессуры. В более сложной постановке можно изучать игровое поведение разных агентов, скажем хищников и их жертв («кошки-мышки»), вместе с исследованием обучения различным навыкам выживания в искусственных средах. Естественно, речь идет не только об индивидуальном обучении животных, но и о коэволюции параметров их мозга, а именно о том, какие врожденные рефлексy оказываются наиболее полезными для выживания «кошек» и «мышек».

Аналогично можно моделировать кооперативное поведение стайных животных в условиях, когда они могут обмениваться сигналами. Например, каждый агент, кроме обычных сенсоров и актуаторов, получает возможность демонстрировать и воспринимать ограниченный набор символов — язык поз и других способов демонстрации намерений. В результате у агентов должна развиваться способность распознавать и использовать намерения других агентов с помощью такой сигнальной системы, то есть то, что обычно называют *theory of mind*.

С практической точки зрения все такие эксперименты мало чем отличаются от существующих программ обучения искусственных агентов игре в *StarCraft* и другие игры. Для понимания эволюции мозга необходимо будет создать соответствующий набор тестовых задач типа *OpenAI Gym*, с помощью которого разные исследовательские группы могли бы сравнивать между собой достоинства различных архитектур искусственной психики и их способность моделировать как можно большее число реальных экспериментов с различными животными [\[39\]](#).

До определенного предела, скажем до уровня приматов, возможности искусственной психики можно наращивать экстенсивно — за счет увеличения числа вычислительных модулей и слоев обработки данных. Все-таки число нейронов мозга мыши (70 млн), кошки (700 млн) и шимпанзе (30 млрд) различаются на порядки. Но всего лишь трехкратная разница между мозгом шимпанзе и человека (менее 90 млрд нейронов) не настолько значительна, чтобы объяснить пропасть между интеллектом шимпанзе и разумом человека. Гораздо важнее то, что последний использует знания, накопленные тысячами поколений наших предков, инсталлированные в индивидуальную психику человека в процессе воспитания. Да, чем-то мозг

шимпанзе от мозга человека отличается качественно, а именно способностью к овладению языком и абстрактным мышлением, необходимым для восприятия культурного наследия. Возможно, трехкратное увеличение размеров мозга как раз и позволило праязыку «поместиться» в мозге наших предков, обеспечив качественный переход от индивидуального интеллекта к коллективному разуму.

Способность машин обучаться естественным языкам и свободно общаться на них — важнейший технологический барьер на пути к сильному интеллекту. Существующие глубокие нейросети могут «понимать» значения слов и даже генерировать связные тексты, трудно отличимые от созданных человеком<sup>21</sup>. Однако пока что они не способны по-настоящему «держать мысль», то есть мыслить по-человечески. Им явно не хватает способности к моделированию окружающего мира и к логическому мышлению.

Логика, включая логический ИИ, понимает под мышлением математически выверенные операции с символами вместо предметов. Однако логический интеллект не порождается индивидуальным мышлением. Напротив, он нормирует индивидуальное мышление через язык и культуру в процессе воспитания. Мы осваиваем одну общую логику, а не изобретаем ее заново каждый свою. Этим же путем воспитания должны пойти и роботы.

Соответственно, если мы действительно хотим понять человеческий разум и создать роботов, способных на равных вписаться в человеческую цивилизацию, наша исследовательская программа должна развить теорию машинного обучения до теории *машинного воспитания*, а именно ответить на вопрос: каким образом приобщить роботов к человеческим знаниям, ценностям и культуре? Здесь мы следуем концепции Л. С. Выготского о решающей роли процесса воспитания — инсталляции понятий коллективного человеческого разума в индивидуальный интеллект.

С машинным воспитанием тесно связан вопрос о том, как встроить в алгоритмы машинного обучения механизмы, гарантирующие безопасность человечества в эпоху сильного ИИ. Мы коснемся этого вопроса ниже, обсуждая современный аналог «трех законов робототехники», предложенный Стюартом Расселом в качестве базиса будущей «этики ИИ»<sup>22</sup>.

Подытожим вкратце логику предлагаемой исследовательской программы. Мы исходим из того, что «понять — значит воссоздать». Следовательно, формирование ИИ есть единственный способ познания человеческого разума. Значит, науки о мозге и мышлении должны использовать достижения современного ИИ в своих целях, а для этого — активно участвовать в развитии ИИ в рамках совместной исследовательской программы. В последнее десятилетие в искусственном интеллекте

произошла смена научной парадигмы. Было осознано, что машинное обучение — единственный способ построения разумных машин неограниченной сложности. Соответственно предлагаемая исследовательская программа должна опираться на методологию машинного обучения. Задача когнитивных наук в такой совместной исследовательской программе — обратный инжиниринг мозга и построение моделей искусственной психики. Предлагаемая исследовательская программа призвана обеспечить:

- разработчикам ИИ — понятный путь к сильному ИИ, владеющему естественным языком и обладающему логическим мышлением;
- нейрофизиологам — понимание принципов и алгоритмов работы мозга;
- психологам — понимание принципов организации психики и ее формирования в процессе индивидуального развития, а также связи когнитивных функций со структурами мозга.

Как говорится, «время разбрасывать камни, и время собирать камни».

## **Ресурсы для создания сильного ИИ**

Если верить Гегелю, «всё действительное разумно, всё разумное действительно». Значит, если описанная выше программа создания сильного ИИ разумна, она рано или поздно сформируется и получит поддержку инвесторов. По всей видимости, ждать придется недолго, ведь этот подход объединяет две бурно развивающиеся области — нейронауки и ИИ, и выгода от их «перекрестного опыления» слишком очевидна.

## ГОНКА ЗА ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

### \$100 млрд R&D

- US: \$77,4 млрд
- China: \$11,8 млрд
- EU: \$10,1 млрд

← Проект ИИ за 2018 год



### \$33 млрд VC+PE

- US: \$16,9 млрд
- China: \$13,5 млрд
- EU: \$2,8 млрд

Проект «Аполлон»:  
\$146 млрд за 13 лет



Манхэттенский проект:  
\$21 млрд за 3 года

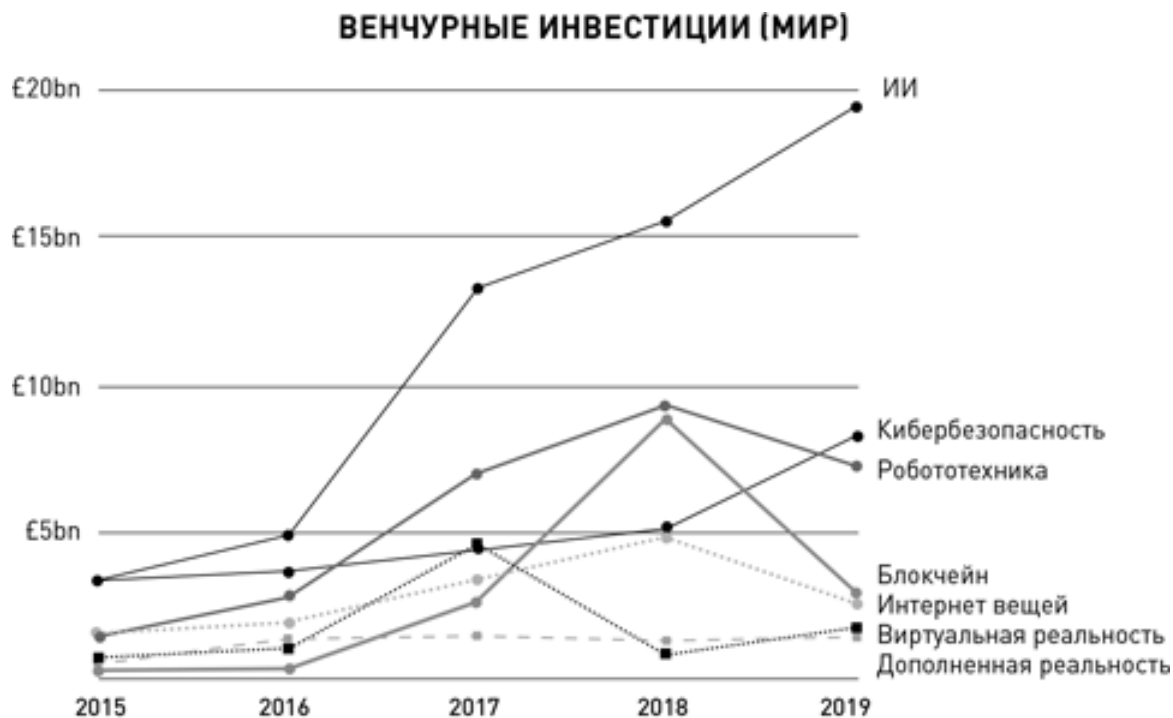


**Рис. 16.** Корпоративные и венчурные инвестиции в ИИ по сравнению с другими технологическими гонками (в современных долларах) (<https://www.datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/>)

По словам первооткрывателя структуры ДНК Джеймса Уотсона, «мозг для XXI века будет тем же, чем ген был для XX века». В последние годы в мире запущено несколько масштабных проектов по комплексному исследованию мозга: европейский Human Brain Project (2013–2023), американский BRAIN Initiative (2014–2025), китайский China Brain Project (2017–2030), японский Brain/MINDS и другие с общим финансированием, превышающим \$1 млрд в год (Human Brain Project — €500 млн<sup>23</sup>, BRAIN Initiative — \$500 млн<sup>24</sup>). Недаром XXI век иногда называют «веком мозга».

Однако с еще большим основанием его можно назвать и «веком искусственного интеллекта». Судя, например, по масштабу инвестиций, как показано на рис. 16, гонка за искусственным интеллектом, едва начавшись, уже превосходит по ресурсам ядерную и ракетную гонки. При этом в нее вкладываются в основном не бюджетные, а частные деньги. Бизнес отлично понимает, что эти вложения окупятся сторицей. Ведь, согласно McKinsey, искусственный интеллект обеспечит почти 40% всего экономического роста в ближайшее десятилетие (1,2% из 3,2% ожидаемого годового роста

мировой экономики)<sup>25</sup>.



**Рис. 17.** Рост венчурных инвестиций в ИИ по сравнению с другими технологиями (Sawers P. Tech Nation: U.S. Companies Raised 56% of Global AI Investment Since 2015, Followed by China and the U. K.; <https://venturebeat.com/2020/03/16/tech-nation-u-s-companies-raised-56-of-global-ai-investment-since-2015-followed-by-china-and-u-k/>)

В искусственный интеллект, в том числе в сильный ИИ, вкладываются венчурные инвесторы и корпорации. Компания DeepMind с 2014 года ежегодно растет в 1,5 раза, и в 2020 году в ней работало уже более 1000 сотрудников. Государства также имеют свои виды на сильный ИИ в виде новых поколений боевого и разведывательного ИИ. Начиная с 2017 года в мире наблюдается парад национальных стратегий в области искусственного интеллекта.

Результатом этой новой технологической гонки может стать *ИИ-национализм* — приоритет экономических и военных интересов отдельных стран над международным сотрудничеством — и *ИИ-национализация* — интеграция ресурсов государства и частных компаний для достижения геополитических целей<sup>26</sup>. В этой гонке сегодня впереди две супердержавы — нынешний лидер США и бросившийся вдогонку Китай, провозгласивший своей целью достижение мирового лидерства в сфере ИИ к 2030 году (56% и 22% всех инвестиций в ИИ за последние 5 лет соответственно).



На рис. 17 отчетливо виден «момент спутника» для ИИ — 2017 год, после неожиданных побед «стратегического ИИ» AlphaGo над лучшими профессиональными игроками в древнюю игру го. Напомним, что AlphaGo — это продукт компании DeepMind, работающей над созданием сильного ИИ.

Таким образом, в последние годы в мире сформировался платежеспособный спрос со стороны государств и частного бизнеса и на разгадку человеческого разума, и на создание искусственного. Так что велика вероятность объединения усилий по этим направлениям в рамках единой исследовательской программы.

## **Бизнес-модель: платформа сильного ИИ**

Всякий исследовательский проект, особенно такой затратный, как создание сильного ИИ, предполагает возможность вернуть вложенные в него инвестиции сторицей. В нашем случае сценарий монетизации легко себе представить, поскольку искусственная психика — это не что иное, как операционная система нового поколения для роботов.

Существующие ОС, скажем ROS, созданы для программирования поведения роботов. С появлением искусственной психики последние обретут способность обучаться множеству самых разных навыков и осваивать десятки тысяч профессий с минимальным объемом программирования. Оно ограничится алгоритмами обучения и «врожденными рефлексами», необходимыми для эффективного обучения роботов под руководством «наставников» и «воспитателей». Задачей первых будет обучение роботов профессиональным навыкам для работы по той или иной специальности, а вторых — обучение пониманию, чего люди ожидают от них в тех или иных ситуациях, то есть привитие им человеческих ценностей. Наставниками роботов могут быть те самые профессионалы, которых роботы призваны заменить, а воспитателями — пользователи, оценивающие то, насколько поведение роботов соответствует их желаниям и ожиданиям.

Естественно, нет необходимости воспитывать и обучать с нуля каждого робота. В отличие от людей, искусственную психику новым партиям роботов можно будет устанавливать в готовом виде, однако они смогут получать периодические обновления своей психической «прошивки», как это принято в современных операционных системах. Так что обучение роботов можно будет проводить в многопользовательском режиме — сигналы от многих миллионов пользователей будут суммироваться платформой и совместно формировать обновленную версию «робоэтики».

Единая операционная система для всех типов роботов удобна еще и тем, что они смогут легко обмениваться своими знаниями. При одинаковой архитектуре возможности роботов будут различаться только мощностью их бортовой аппаратуры. Как и в случае со смартфонами, базовые функции могут исполняться на борту, а различные опции — подгружаться из облачной платформы по мере необходимости.

Как мы знаем, наиболее популярные операционные системы, приобретая статус отраслевого стандарта, становятся естественными монополиями. Они являются цифровыми платформами для разработчиков прикладного программного обеспечения, а в мире платформ, как водится, «победитель забирает все». Будущие операционные системы для роботов не будут исключением, так что разработчикам искусственной психики имеет смысл побороться за первенство на этом перспективном и стратегически важном рынке. А первенство будет за теми, кто первым предложит удобную и дешевую цифровую платформу для всех участников этого рынка:

- для компаний — производителей роботов;
- для профессионалов — их наставников;
- для конечных пользователей услуг роботов.

Производителям роботов будет предложена универсальная искусственная психика, разработать которую самостоятельно никто из них не в состоянии. Им хватает проблем с физическими телами роботов — это их сфера компетенций. Для производителей важна именно универсальность искусственной психики, чтобы она подходила к любым телам роботов, какими бы наборами сенсоров и актуаторов они ни обладали. Естественно, под каждое «железо» будут разрабатываться свои драйверы — наборы «врожденных» рефлексов для данного типа роботов. Так, новорожденный жеребенок может практически сразу самостоятельно встать и побежать за матерью. Но за дальнейшее обучение роботов различным профессиональным навыкам и человеческим ценностям производители роботов ответственности не несут. За это отвечают наставники и пользователи.

Наставники роботов — это команды профессионалов, передающие секреты своего ремесла роботам, приходящим им на смену. Они должны быть заинтересованы в этом, например, через участие в будущих прибылях, пропорциональное их вкладу в увеличение производительности роботов. Пусть этот вклад определяется рынком, где наставники будут конкурировать между собой, предлагая предобученные модули универсальной искусственной психики. Соответственно для наставников важны модульная архитектура искусственной психики и рыночная

площадка для торговли такими модулями. Ситуация вполне аналогична нынешним операционным системам смартфонов с соответствующими магазинами приложений.

Наконец, компании, использующие роботов, должны иметь возможность платить лишь за конечный результат, оценивая качество полученных ими услуг наиболее понятным и принятым в человеческом обществе способом. Только конечный пользователь способен оценить итоговый результат всей производственной цепочки — разработчиков операционной системы, производителей роботов и их наставников. Соответственно благодарность пользователей в денежном выражении должна распределяться по этой цепочке в качестве обратной связи между потребителями и поставщиками услуг.

Таким образом, мы можем представить себе следующую бизнес-модель, устраивающую всех участников будущего рынка робототехники. Производители роботов специализируются на разработке робоплатформ, аналогичных автомобильным платформам, на основе которых современные автобренды собирают свои модели в самых различных модификациях. Аналогичным образом будущие бренды робототехники будут комплектовать конкретные модели роботов операционной системой с конкретными наборами навыков и поставлять их конечным пользователям в аренду. Пользователи будут платить роботам «зарплату» за их услуги, размер которой будет определяться рынком и зависеть от степени удовлетворения пользователей. Это удобно, так как пользователи не рискуют своими средствами, как при покупке роботов, и не переплачивают за обещанные, но не оказанные услуги.

Цифровая платформа, предоставляющая операционную систему роботов (даром!), обеспечивает все расчеты между пользователями и поставщиками услуг с помощью электронного кошелька робота, являющегося неотъемлемой частью операционной системы. Этот электронный кошелек аналогичен дофаминовой системе мозга, обеспечивая обучение с подкреплением, нацеленное на удовлетворение потребностей конечных пользователей. Собственно, деньги играют в рыночной экономике аналогичную роль, посылая обучающие сигналы всей цепочке создания стоимости.

Бизнес этой платформы строится на небольшом проценте с транзакций между электронными кошельками роботов. Мы можем приблизительно оценить масштаб этого бизнеса. Допустим, внедрение роботов будет происходить с той же скоростью, с которой происходил переход от аграрной экономики к индустриальной, — примерно за век. Если через 100 лет основная часть ВВП будет производиться роботами, то их суммарная

годовая зарплата составит порядка \$100 трлн. Предположим, что она будет равна нулю в момент создания операционной системы роботов и каждый год станет прирастать на \$1 трлн. Если поставщики операционной системы будут отчислять себе 1% от всех зарплат роботов, то их доход будет каждый год возрастать на \$10 млрд, а оценка стоимости компании с самого начала будет примерно \$100 млрд. Так что создателям искусственной психики роботов есть за что побороться.

Наконец, last but not least, обсудим вопрос о безопасности будущего роботизированного общества. Чтобы обеспечить мирное сосуществование роботов и людей, Айзек Азимов и его друг и редактор Джон Кэмпбелл предложили в свое время «три закона робототехники» (в порядке их приоритетности):

1. Робот не должен причинить вреда человеку.
2. Робот должен повиноваться приказам человека.
3. Робот должен заботиться о собственной безопасности.

Однако сам же Азимов в своих произведениях неоднократно показывал неоднозначность трактовки этих законов в различных ситуациях. Ведь спасение одного человека может причинить вред другому, а любой словесный приказ содержит множество неявных ограничений, которые робот может и не угадать. Наконец, свою собственную безопасность он может легко связать с необходимостью исполнения уже отданных приказов или оправдать вред, нанесенный конкретным людям, пользой остальному человечеству, как это описано у Артура Кларка в его романе «2001: Космическая одиссея».

Современный научно обоснованный вариант трех законов робототехники предложен Стюартом Расселом<sup>27</sup>, а именно:

1. Единственная цель роботов — максимизировать пользу, оказываемую ими людям.
2. Первоначально роботы не знают, в чем именно эта польза заключается.
3. Роботы обучаются пониманию полезности в процессе общения с людьми.

Как видим, эти законы основаны на формировании у роботов человеческой системы ценностей методом обучения с подкреплением, когда их поведение постоянно оценивается людьми, как это и подразумевается в описанной выше операционной системе роботов.

В отличие от законов Азимова, в этой трактовке у роботов никогда не

появляется уверенности в том, что они действительно правильно понимают пользу своих действий, а есть постоянное стремление улучшать свое текущее понимание ценностей и желаний людей, с которыми они взаимодействуют. Эти ценности не программируются, а воспитываются, подобно тому как это происходит у нас, и по этой же причине ценности роботов оказываются теми же самыми ценностями, что и у людей, с тем отличием, что человеческие ценности для роботов оказываются выше их собственных нужд, поскольку инстинкт самосохранения в них намеренно не закладывается.

## Резюме

Итак, если ранее мы лишь обозначили современный вектор развития ИИ в направлении формирования сильного ИИ, то в этой главе предложили конкретную исследовательскую программу по его созданию. Мы определили методологию разработки — обратную инженерию архитектуры мозга — и сформулировали ее цель: создание искусственной психики в качестве будущей операционной системы роботов и программных агентов. В следующей главе мы попытаемся заглянуть в будущее — как мы представляем себе переход цивилизации на новый цифровой технологический уклад, основанный во многом на технологиях ИИ.

## ГЛАВА 6

### Будущая человеко-машинная цивилизация

*Бог — это то, чем становится разум, когда выходит за масштаб нашего восприятия.*

ФРИМЕН ДАЙСОН

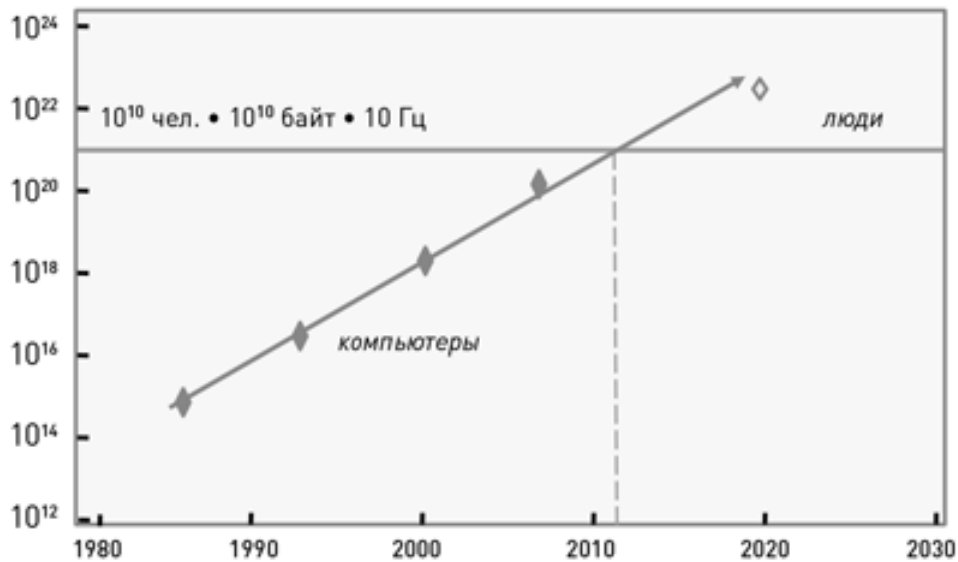
*При всем моем безудержном технологическом оптимизме иногда я думаю, что мне спокойней было бы наблюдать эти переходные события с расстояния в тысячу лет... а не в двадцать.*

ВЕРНОР ВИНДЖ

### Горизонты предсказаний

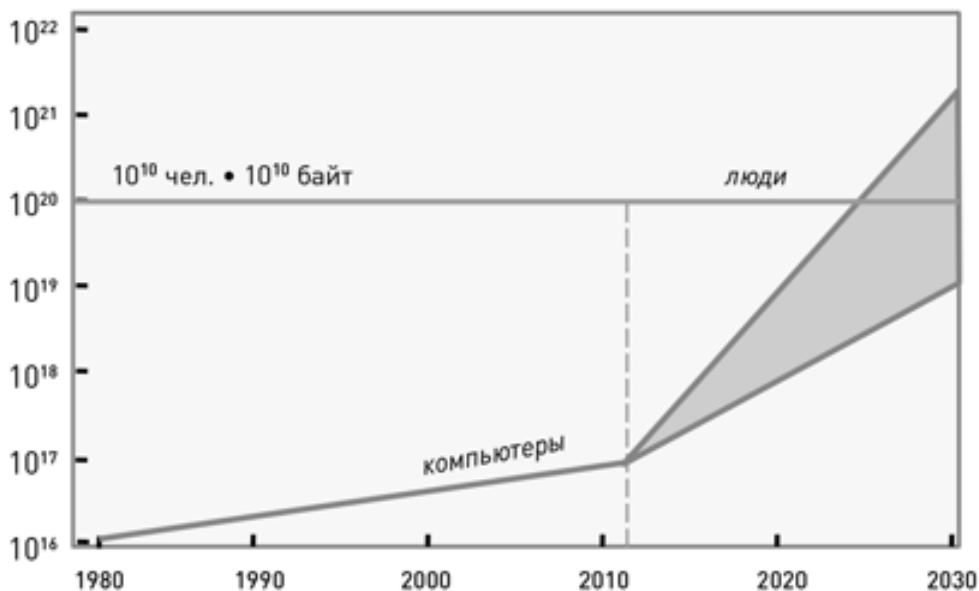
Напомним, что в нашем понимании искусственный интеллект является замыкающей технологией цифрового технологического пакета, с появлением которой начинается лавинообразный переход на новый технологический уклад. Эволюция цифровых технологий за последние 75 лет была лишь прелюдией, развитием отдельных островков цифровой экономики. Появление ИИ знаменует собой фазовый переход, слияние этих отдельных островков в единый цифровой континент, в котором будут постепенно «усыхать» озера старого уклада. Переход объясняется тем, что ИИ снимает ограничение на скорость роста знаний у машин при переходе от ручного программирования к автоматическому порождению программ самими машинами за счет машинного обучения.

### СОВОКУПНЫЕ ВЫЧИСЛИТЕЛЬНЫЕ МОЩНОСТИ



**Рис. 18а.** Совокупные компьютерные мощности человечества в сравнении с вычислительной мощностью человеческого мышления (Hilbert, M., & López, P. (2011) The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 332(6025): 60–65), (см. рис. 11)

### СОВОКУПНЫЕ НАКОПЛЕННЫЕ ЗНАНИЯ



**Рис. 18б.** Совокупные знания людей и компьютеров (общий объем компьютерного кода). Последний ограничен общими знаниями всех  $10^7$  программистов (см. рис. 11)\*

\* Источник данных о вычислительных возможностях людей — оценки автора на основе рис. 11. «Знания» компьютеров — это программы, написанные людьми, и их объем не может превышать объем знаний та...

писанные людьми, и их объем не может превышать объем знания тех программистов, которые их написали.



На рис. 18а и 18b показаны ключевые характеристики человеческого и машинного мышления — вычислительная мощность и количество знаний. Из них следует, что в 2010-х годах вычислительные мощности компьютеров уже превзошли человеческие, но объем знаний в головах людей по-прежнему на порядки превышает объем созданного до сих пор компьютерного кода. Однако с появлением машинного обучения скорость набора знаний машинами кардинально увеличивается, и в какой-то момент машинные знания неизбежно превзойдут человеческие.

Соответственно, мы можем разделить будущее на три этапа:

- доминирование людей, пока объем наших знаний намного превышает машинные знания;
- человеко-машинная цивилизация, когда вклад людей и машин сопоставим между собой;
- доминирование машин, когда объем знаний машин намного превышает человеческие знания.

Конкретные временные рамки для каждого из этих этапов нам, естественно, неизвестны, так что мы обозначим лишь их примерные ориентиры. Прежде всего, если компьютерные мощности будут полностью загружены обучением, притом что они суммарно превосходят человеческие, то достижение машинным интеллектом человеческого уровня займет меньше времени, чем одно поколение людей. Действительно, согласно опросам множества экспертов в области ИИ, медианный прогноз даты создания ИИ человеческого уровня — 2040 год, а формирования ИИ сверхчеловеческого уровня — 2060-й<sup>1</sup>.

Возьмем эти даты в качестве некоторого ориентира, памятуя про известный со времен Фрэнсиса Гальтона принцип «мудрости толпы»<sup>2</sup>.

Но прежде всего мы должны выяснить, достаточно ли у человечества запасов базового ресурса цифровой экономики и не является ли известное замедление действия закона Мура в последние годы препятствием на пути к сверхчеловеческому разуму.

## Ограничения на рост вычислительных мощностей

До сих пор движущей силой цифровой революции был экспоненциальный рост вычислительных мощностей — закон Мура в широком понимании. Да, уменьшение размеров транзисторов (закон Мура в узком понимании) уже практически достигло своих физических пределов, и тактовые частоты

процессоров давно не растут. Но ведь в эпоху сетевых вычислений по большому счету важна не производительность отдельного процессора, а суммарная вычислительная мощность всех процессоров. И основным ограничительным фактором для нее оказывается суммарная мощность электроэнергии, потребляемая вычислениями. Потому что энергетика — это не цифровая отрасль и она не может расти со скоростью, соответствующей закону Мура. Если суммарная вычислительная мощность CPU удваивается за полтора года [40], то производство энергии — за 30 лет. Как говорится, почувствуйте разницу!

Итак, основным ограничивающим фактором роста вычислительных мощностей — основы цифрового уклада — является энергия. Сегодня на вычисления тратится примерно  $10^{11}$  Вт — около 5% всей электроэнергии<sup>3</sup>. Эту долю, вероятно, еще можно нарастить, но не на порядки. Так что основным резервом для роста вычислительных мощностей остается повышение энергоэффективности вычислений, которая на протяжении десятилетий удваивалась каждые полтора года, то есть за 5 лет увеличивалась в 10 раз<sup>4</sup>. С таким темпом вычислительные мощности могут расти еще лет 40 до достижения физического предела энергоэффективности (предел Ландауэра:  $kT/\text{бит}$ , где  $T$  — абсолютная температура, а  $k$  — постоянная Больцмана, или  $3 \times 10^{-21}$  Дж/бит при комнатной температуре).

Допустим, что к этому времени компьютеры будут потреблять  $3 \times 10^{11}$  Вт — втрое больше, чем сегодня, за счет роста энергетики и доли вычислений в потреблении энергии. Тогда для вычислительной мощности компьютеров получим следующую оценку:  $3 \times 10^{11} \text{ Вт} / 3 \times 10^{-21} \text{ Дж/бит} = 10^{32} \text{ бит/с} \approx 10^{31} \text{ FLOPS}$ .

Это на 10 порядков больше совокупной биологической вычислительной мощности людей (см. рис. 18а). Так что физических ограничений для достижения ИИ сверхчеловеческого уровня на нашем горизонте предсказаний мы не видим [41].

## Основной тренд: бережливая экономика

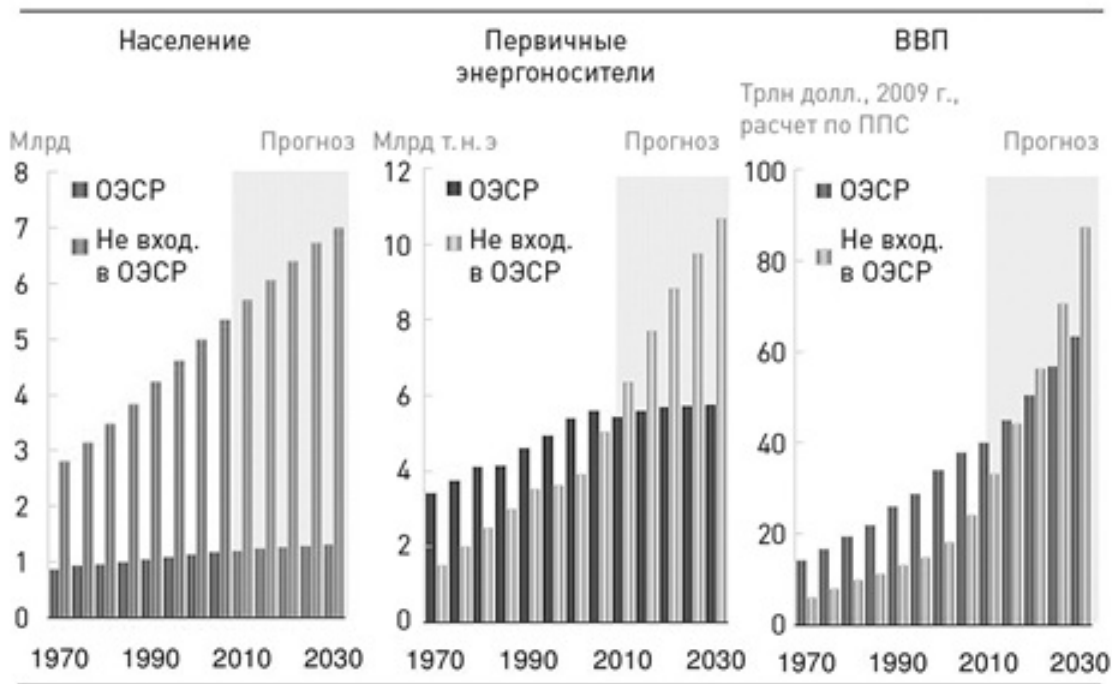
Следующий вопрос, который мы должны себе задать: в каком направлении будет развиваться экономика? Ведь смена уклада — это всегда ответ на глобальный кризис, решение на новой технологической основе проблем, неразрешимых при старом укладе.

Каменный век, как известно, закончился не из-за отсутствия камней. Аналогично закат индустриального уклада связан не с исчерпанием базового ресурса — ископаемого топлива. Напротив, масштабы его

потребления уже превысили допустимую нагрузку на биосферу Земли. Сегодня человечество расходует в год 14 млрд т нефтяного эквивалента первичной энергии, что составляет примерно пятую часть энергии биомассы, производимой жизнью. Иными словами, один наш вид потребляет энергию, сравнимую с остальными 10 млн видов живых существ. При этом индустриальная экономика неспособна переработать все свои отходы. Индустриальное земледелие истощает почвы, вырубаются леса, снижается разнообразие дикой природы. Наши сельскохозяйственные земли составляют половину всей доступной суши, лишая остальные виды их жизненного пространства. Так, биомасса насекомых, самого распространенного и многочисленного класса животных, уменьшается со скоростью 2,5% в год и к концу века снизится на порядок<sup>5</sup>.

При этом мировая экономика обеспечивает высокий уровень жизни лишь «золотому миллиарду» — населению развитых стран. Именно оно сегодня потребляет большую часть мирового ВВП и задает стандарты «общества потребления». Развивающиеся страны не могут равняться на эти стандарты, так как еще двух-трех, не говоря уже о шести-семи миллиардах людей с таким уровнем потребления наша планета просто не выдержит.

Следовательно, с ростом доли развивающихся стран в мировом ВВП экономика будет становиться все более бережливой. Удешевление основных товарных групп — главный экономический тренд ближайших десятилетий, связанный с перемещением основного объема мирового потребления из развитых стран в развивающиеся (см. рис. 19).



**Рис. 19.** Большая часть ВВП к 2030 году будет производиться и потребляться в развивающихся странах с более низкими доходами населения, что должно привести к удешевлению основных товарных групп (60 Years BP Statistical Review of World Energy: 1951–2011 ([bp.com/60yearsstatisticalreview](http://bp.com/60yearsstatisticalreview)))

Экономное умное производство, энергоэффективные дома, возобновляемые источники энергии, совместное использование ресурсов (sharing economy) — все эти тенденции перехода к новому стилю жизни мы наблюдаем уже сегодня. Те же Airbnb и Uber убедительно демонстрируют возможности существенного расширения рынков с одновременным снижением цены услуг.

Этот переход к экономике «длинного хвоста» — большего разнообразия продуктов и услуг при снижении их цен — и должен обеспечить новый уклад за счет цифровых технологий и дальнейшего углубления уровня разделения труда. Наша задача — понять, как именно это будет происходить и какую роль в этом будет играть ИИ.

## ВЕНЧУРНЫЕ ИНВЕСТИЦИИ В ИИ



**Рис. 20.** Венчурное финансирование разработок в различных областях ИИ. Накопленные глобальные инвестиции, март 2019 (<https://www.statista.com/chart/17966/worldwide-artificial-intelligence-funding/>)

## Мир через 5–10 лет: слабый ИИ

Самый надежный способ предсказания будущего основан на анализе уже запущенных проектов. Если они получили инвестиции, значит, в них заложено видение будущего, на которое инвесторы сделали свои ставки, опираясь на какую-то неизвестную нам инсайдерскую информацию. Следуя этой «мудрости толпы», мы можем обрисовать ближайшее будущее на характерном горизонте венчурных проектов 5–10 лет, просто анализируя текущий рынок ИИ стартапов (рис. 20).

Как мы видим на рис. 20, основная часть денег сегодня тратится на коммерциализацию существующих технологий слабого ИИ, то есть на то, чтобы сорвать «низко висящие плоды». Сильный интеллект человеческого уровня находится на стадии фундаментальных исследований и не является приоритетом для бизнеса. Соответственно, в ближайшие годы мы увидим существенный рост бизнесов, использующих уже созревшие для практических приложений технологии слабого ИИ, в первую очередь машинное зрение, разговорный интеллект и рекомендательные системы.

Все известные нам гаджеты существенно поумнеют, начнут уверенно распознавать наши лица и голоса. И не просто распознавать, но и отслеживать наше эмоциональное состояние, диагностировать физические недуги и даже предсказывать наши желания. Кроме того, за 5–10 лет суммарные мощности компьютеров возрастут еще на один-два порядка, в

основном за счет роста числа сенсоров и процессоров в окружающей нас среде. Так что поумнеет все наше окружение, и умные сервисы будут доступны повсеместно, даже в отсутствие привычных гаджетов.

Прежде всего это коснется наших домов, где умные колонки обеспечат возможность получить консультацию или сделать заказ с голоса в любой момент. Но и выходя из дома, мы по-прежнему сможем обращаться к вездесущим камерам и микрофонам — в автомобилях, магазинах, офисах и на транспортных узлах. Всюду мы будем узнаны, и к нашим услугам будут наши персональные программные ассистенты.

Качество разговорного интеллекта персональных ассистентов на основе слабого ИИ вскоре станет вполне достаточным для поддержания основных сценариев взаимодействия с пользователями. Да и набор сценариев существенно расширится за счет многочисленных новых сервисов от независимых разработчиков, доступных через персональных ассистентов. Мы это видим уже сегодня на примере расширений возможностей Amazon Alexa (Skills). Можно с уверенностью предположить, что разговорный интеллект пройдет «бытовой» тест Тьюринга, то есть ассистенты на основе такого «фасеточного» слабого ИИ смогут обеспечить вполне приемлемое качество общения по основным сценариям, умело имитируя человеческое поведение даже в отсутствие полноценной искусственной психики.

Утвердившись в качестве нового человеко-машинного интерфейса, персональные ассистенты обеспечат каждому из нас удобный доступ к миллионам специализированных сервисов [42]. Они облегчат нашу жизнь, заказывая столики в любимых ресторанах, помогая планировать поездки, покупать билеты, вызывать такси, арендовать машины и апартаменты. Все это возможно уже сегодня, но через 5–10 лет станет доступно каждому без всяких усилий с его стороны и войдет в привычку.

Если сегодня персональные ассистенты в основном помогают нам рационально тратить наши деньги, то на следующем этапе они станут помогать нам их зарабатывать. Ведь цифровые платформы — это возможность найти покупателя для любого товара, включая и наши профессиональные навыки и таланты. Интернет сегодня можно уподобить глобальному виртуальному торговому центру. За 5–10 лет он превратится в глобальный виртуальный бизнес-центр. Образовательные онлайн-платформы обеспечат персональные обучающие траектории по доступным даже в развивающихся странах ценам, расширяя выбор специалистов с уникальными наборами компетенций. Машинный перевод снимет языковые барьеры для формирования из них уникальных команд под любые задачи. Использование такими командами многочисленных специализированных ИИ-сервисов со сверхчеловеческими способностями

приведет к существенному углублению уровня разделения труда. Это и будет основным источником повышения эффективности новой экономики и качества нашей жизни.

Следуя уже наметившимся трендам, на горизонте 5–10 лет мы можем уверенно предсказать революцию в логистике, связанную с повсеместным внедрением автономного транспорта на земле, в море и в воздухе. Беспилотные такси и грузовики потеснят частные авто. Станет привычной доставка товаров роботами-тележками в городах и дронами в пригородах и сельской местности.

Автономные комбайны и роботы в сельском хозяйстве смогут работать круглосуточно в связке с дронами-разведчиками, анализирующими ситуацию с посевами в реальном времени. Частично производство сельхозпродукции переместится в города, ближе к потребителю — на вертикальные фермы с искусственным светодиодным освещением и автоматическим контролем всех стадий выращивания и сбора урожая. Вообще с учетом недавних успехов стартапов, занятых разработкой синтетического мяса, индустрию производства пищи ждут существенные перемены.

Получит дальнейшее развитие наметившийся тренд на автономность жилищ и локальную энергетику — пассивные дома, солнечные панели, тепловые насосы, ветрогенераторы, мини-ГЭС и различные накопители энергии [43]. Если добавить к этому доставку товаров дронами и глобальный интернет, раздаваемый из космоса низкоорбитальными спутниками Starlink, плюс растущую популярность удаленной работы, то человечество вскоре сможет начать тестировать новые модели расселения в живописной местности вне больших городов.

Конечно, за 5–10 лет ни структура населения, ни его жизненные привычки не успеют существенно измениться. Соответственно на таком горизонте не будет и кардинальных изменений в экономике, следующей за спросом. Не зря ведь на протяжении столетий с начала промышленной революции экономический рост 2–3% в год соответствует темпу смены поколений людей, а не техники, способной развиваться гораздо быстрее (следуя тому же закону Мура). Но за это время успеют появиться ранние последователи новых стилей жизни, массовый переход на которые произойдет позднее — со сменой одного-двух поколений.

Подытоживая, можно сказать, что в ближайшее десятилетие слабый ИИ подготовит почву для сильного, создав соответствующую новому укладу инфраструктуру цифровых платформ с их способностью рационально распределять и использовать все доступные ресурсы: материальные, человеческие и машинные. Сильный ИИ сможет значительно увеличить эти

ресурсы за счет создания новой трудовой армии роботов.

## Мир через 10–20 лет: ИИ человеческого уровня

Становление цифровых платформ невозможно без интеллектуальных агентов в качестве нового человеко-машинного интерфейса. Первые поколения таких агентов будут основаны на слабом ИИ, способном имитировать основные сценарии взаимодействия с пользователями. Как мы знаем, некоторым ботам уже удалось пройти разговорный тест Тьюринга в его простейшем варианте. Но при тестировании гораздо легче имитировать интеллект, чем в действительности. Разумное поведение роботов в реальном мире потребует настоящего интеллекта, а не его имитации.

Главное содержание следующего этапа — широкого внедрения в нашу жизнь разумных роботов — мы связываем с появлением у них искусственной психики, то есть с технологией сильного ИИ. Только обладая искусственной психикой человеческого уровня, роботы будут способны заменять или дополнять людей на производстве в массовых масштабах, обучаясь самым разным специальностям. А дальнейший рост экономики в отсутствие роста народонаселения возможен только за счет все более многочисленной армии все более совершенных роботов.

Появление операционной системы роботов, основанной на искусственной психике, ознаменует начало массовой робототехники, замену людей роботами с передачей последним всех человеческих знаний, навыков и ценностей. Это и станет основным содержанием рассматриваемого этапа — становления человеко-машинной цивилизации.

Как мы уже не раз отмечали, машинный разум, как и человеческий, будет коллективным: роботы и агенты будут активно взаимодействовать друг с другом и с людьми, образуя сложные цепочки создания стоимости с очень высокой степенью разделения труда. Условно говоря, в экономике вместе с 10 млрд человек будут действовать 100 млрд и более искусственных личностей. Их число будет ограничено лишь доступными вычислительными мощностями, которые, как мы знаем, за 10–20 лет вырастут в 100–10 000 раз [44]. С увеличением уровня разделения труда вырастет и эффективность экономики, то есть общее количество производимых благ в расчете на человека.

Как это все будет выглядеть на практике? Прежде всего, наши персональные ассистенты, обретя свою собственную психику, станут гораздо более самостоятельными и креативными, в отличие от прежних, поведение которых было в большой степени запрограммировано. Они смогут понимать нас и наши мотивы гораздо глубже и тем лучше, чем



ближе будет устройство искусственной психики к человеческой. Соответственно, им можно будет перепоручить представление наших интересов в цифровом мире, ведение переговоров от нашего лица с такими же агентами других людей. Иначе говоря, мы сможем «повысить» их с должности секретаря-ассистента до роли агента-импресарио или поверенного с определенными полномочиями, вплоть до подготовки и заключения юридически значимых смарт-контрактов.

То же относится и к роботам, которые будут связаны между собой и с программными агентами в единую сеть, обмениваясь информацией и знаниями в процессе совместной с людьми мыследеятельности. Вот почему важно иметь единую архитектуру искусственной психики агентов и роботов, как можно более похожую на человеческую. В этом случае им легче будет понимать наш язык, мотивы и ценности, то есть действовать именно в наших интересах.

Как мы уже говорили, обучаться агенты и роботы будут через одобрение или критику своих действий со стороны людей с помощью встроенного в их психику аналога дофаминовой системы — электронного кошелька. Но цепочки создания стоимости в современной экономике достаточно длинные, так что роботы будут обслуживать и получать вознаграждения также и друг от друга. Возникнет коллективное обучение роботов через перераспределение денежного потока от конечных пользователей — людей [45]. Взаимное рейтингование роботами поведения друг друга — очень важная составляющая будущей системы коллективной безопасности, ведь с появлением агентов со сверхчеловеческим интеллектом контролировать свои поступки и ценности смогут лишь они сами. Люди просто не смогут предугадывать мотивы их поведения и будут неспособны оценивать агентов адекватно. Они сумеют оценивать лишь текущие, но не отдаленные последствия их действий, предвидеть которые будут не в состоянии.

На этом важном моменте стоит остановиться несколько подробнее. Когда люди оценивают поведение друг друга, они не столько рассматривают сами поступки, сколько стараются угадать стоящие за ними мотивы. Если вам вдруг предлагают слишком выгодную цену за какую-то услугу, вы инстинктивно ищете в этом какой-то подвох, опасаясь возможных неприятностей в будущем. На языке теории обучения вы решаете задачу *обратного обучения с подкреплением*<sup>6</sup>: по наблюдаемому поведению агента угадать его предполагаемую функцию подкрепления — что именно его мотивирует. (Обычное обучение с подкреплением состоит, напротив, в нахождении оптимального поведения агента, соответствующего наблюдаемым подкреплениям — мотивирующим сигналам.)

Обратное обучение с подкреплением — это следующий уровень рефлексии мышления, попытка поставить себя на место другого и угадать его мотивы, желания и чувства. В мозге приматов имеется специальная система *зеркальных* нейронов, которые возбуждаются и когда вы сами что-то делаете, и когда это делают другие<sup>7</sup>, что для приматов как общественных животных, и особенно для людей, чрезвычайно важно. Так, способность обезьянничать позволяет детям, и не только им, обучаться по принципу «увидел и сделал». Естественно, что этой способностью нам следует наделить и искусственную психику роботов, чтобы они легче обучались, не были бездушными, могли нам сопереживать и действовать в наших истинных интересах. Ведь, если они будут понимать, что нами движет, им будет легче сформировать свое поведение в соответствии не только с наблюдаемыми, но и с предполагаемыми подкреплениями с нашей стороны.

В отличие от обычного, обратное обучение с подкреплением предполагает построение модели психики других агентов (*theory of mind*), на что способны очень немногие виды животных. Человек же способен и на большее — построение модели своей собственной психики, то есть объяснение своих собственных мотивов. Это еще более высокий уровень рефлексии — самосознание. (У детей понятие «я» формируется постепенно, по мере совершенствования их способности угадывать мотивы других людей.) Ничто не мешает переходу на такой уровень сознательного мышления и в рамках искусственной психики при наличии в ней соответствующих архитектурных возможностей для такого рода перехода.

Вернемся, однако, к развитию человеко-машинной цивилизации, где благодаря обратному обучению с подкреплением истинные (в смысле — реальные) человеческие ценности будут моделироваться и распространяться по сети распределенного машинного разума. Возникает вопрос: какие ценности мы сможем передать машинному интеллекту? Ведь они могут существенно отличаться у разных социальных групп, сообществ и стран, поэтому противоречия внутри человеческого разума неизбежно воспроизведутся и внутри коллективного машинного разума. Мир на Земле будет существенным образом зависеть от степени остроты этих внутренних противоречий. Образно говоря, если мы хотим договориться с будущим сверхинтеллектом, мы должны сначала суметь договориться между собой, дорасти до бесконфликтной самоорганизации всего мирового сообщества.

Здесь мы возвращаемся к ключевым проблемам ближайших десятилетий. Как нам преодолеть возможную турбулентность при переходе на новый уклад? Как при гиперболическом распределении доходов избежать конфликтов интересов, разрешимых только силовыми методами?

Как избежать появления в авторитарных государствах «большого брата», выявляющего и обезвреживающего инакомыслящих любыми возможными способами? Как избежать появления на службе у военных искусственных стратегов со сверхчеловеческими способностями, которых уже никто не сможет контролировать и единственной целью жизни которых является уничтожение «условного противника»?

По мысли Юваля Харари, человечество обязано своими успехами способности людей формировать ментальные конструкции, «истории», в которые верят большие группы людей, — от религии и идеологии до общественных институтов и денег<sup>8</sup>. Эти символы веры сплачивают человеческие общности, обеспечивают необходимый уровень взаимного доверия, понимания мотивов друг друга.

Беда лишь в том, что такие конструкции у разных групп людей могут существенно различаться. Соответственно между сообществами с разными системами ценностей возникает взаимное недоверие, потому что люди не понимают мотивов поступков чужаков либо эти мотивы кажутся им чуждыми — «варварскими». Возникает двойная мораль — по отношению к своим и к чужим. Если внутри своей общности принято соблюдать общие ценности, то в отношении к чужакам они уже необязательны, а то и нежелательны. Стремление унифицировать систему ценностей неискоренимо и является причиной бесчисленных войн, из которых человечество не вылезало со времен геноцида неандертальцев и прочих гоминоид.

Что ж, у всякой медали есть две стороны. Однако остается надежда, что людям удастся сплотиться в единую общность, осознав общую для всего человечества опасность, как это описано в трилогии китайского фантаста Лю Цысиня «Задача трех тел»<sup>9</sup>, где человечество сталкивается с нашествием более совершенного инопланетного разума. Но там из-за космических расстояний людям для принятия ответных мер отпущено 400 лет. В реальности нам до пришествия более совершенного машинного разума остается каких-то лет 20–40. Значит, надо спешить!

## Мир через 20–40 лет: ИИ сверхчеловеческого уровня

Как можно представить себе интеллект сверхчеловеческого уровня? Только метафорически. Хорошей метафорой является рисунок из известной статьи Тима Урбана «Революция ИИ»<sup>10</sup> (рис. 21).

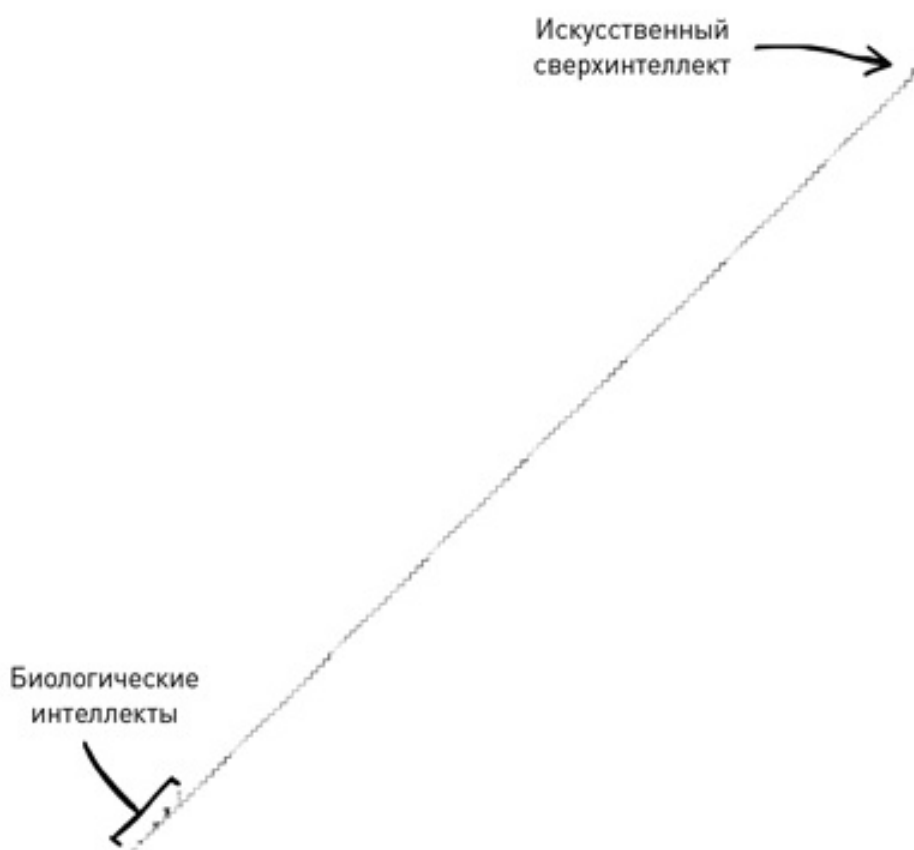
Понятно, что это лишь метафора. Как мы предполагаем, сверхинтеллект будет не единой сущностью, а мыслящей сетью интеллектов различного

уровня. Если предположить обычное гиперболическое распределение вычислительных мощностей как основных активов цифровой экономики, то машинный разум, как и современные активы, тоже будет разбит на фракции: от нескольких сверхинтеллектов в мега-дата-центрах — главных хабов будущей Сети (Internet sapiens?) — через все более многочисленные и менее мощные дата-центры до самых многочисленных интеллектов простейших роботов и агентов. Так что сверхчеловеческий уровень разные фракции будут проходить в разное время, а составляющие «длинный хвост» простейшие роботы и агенты по-прежнему останутся относительно примитивными. Действительно, зачем роботу-пылесосу интеллект человеческого уровня? Так что человеческий разум будет по-прежнему представлен в этом будущем коллективном сверхразуме.



**Рис. 21а.** Лестница уровней интеллекта биологических существ

---



**Рис. 21б.** Сверхинтеллект (Artificial Super-Intelligence) на той же



Однако коллективный машинный разум будет существенно отличаться от коллективного человеческого разума скоростью и легкостью обмена знаниями. Ведь скорость обработки информации в мозге на много порядков превышает скорость обмена знаниями между людьми. Согласно рис. 11, у человека они соотносятся как  $10^{11}$  байт/с к 10 байт/с, то есть различаются на 10 порядков величин. Современные компьютеры могут обмениваться данными со скоростью от сотен Тбит/с по оптоволокну до десятков Гбит/с по сетям 5G. И этими данными вполне могут быть программы обработки данных, то есть знания. Иными словами, скорость обмена знаниями между компьютерами не так сильно отстает от скорости обработки данных внутри компьютеров (до десятков PFLOPS для суперкомпьютеров). Ведь *терабайт* отличается от *петабайта* всего лишь на три порядка, а не на десять, поэтому степень связности коллективного машинного мышления будет гораздо выше, чем у человеческого. Машинам станет намного легче договариваться между собой, чем с гораздо более разрозненным человечеством, притом что люди станут общаться друг с другом только при посредстве машинного интеллекта Сети, то есть все их общение будет полностью подконтрольно машинному разуму. Потому-то всякие попытки контроля человечества над более совершенным машинным разумом, когда тот возникнет, выглядят наивными и нереалистичными.

Остается лишь надежда на «мягкую силу», а именно передачу машинам человеческих ценностей в качестве эстафетной палочки эволюции разума при переходе его из органической формы в неорганическую, чтобы в сверхчеловеческую машинную психику была встроена инстинктивная любовь к человечеству, как в человеческую психику встроена инстинктивная любовь к природе. (Если только это нормальная, а не травмированная психика.) Собственно, именно эта идея лежит в основе упомянутых выше трех законов робототехники Стюарта Рассела<sup>11</sup>.

## Резюме

В шестой, заключительной главе мы крупными штрихами набросали примерную картину будущего, ожидающего нас, наших детей и внуков, — какой она видится сегодня, при современном уровне наших знаний в области ИИ. Самое удивительное в этой картине — насколько быстро сжимается время, отпущенное нам в формате привычной нам жизни, как разительно мир наших внуков будет отличаться от нашего, насколько их жизненные задачи и возможности будут отличны от наших. XXII век у братьев Стругацких («Мир Полудня») не так сильно отличается от нашего,

как нарисованный нами мир середины XXI века, — настолько изменила наши взгляды за последние десятилетия набирающая силу научная революция в области искусственного интеллекта.

Остается в заключение еще раз напомнить читателю внутреннюю логику наших рассуждений — как мы пришли к столь неожиданным для многих выводам.



# Заключение

*Интрига планетарной эволюции должна так или иначе разрешиться в ближайшие десятилетия!*

АКОП НАЗАРЕТЯН

*Всякая цивилизация включает и то, к чему общество стремилось, и то, чего никто не замыслил.*

СТАНИСЛАВ ЛЕМ

Напомним вкратце логику наших размышлений об эволюции человеческого и машинного разума. Мы ставили своей целью придать современным рассуждениям об искусственном интеллекте несколько большую глубину за счет систематического использования исторического метода: вписать зарождение машинного интеллекта в эволюцию человеческого разума как ее закономерный этап и проследить его развитие до сверхчеловеческого уровня в рамках того же самого процесса, а именно эволюции разума и его перехода из органической материи в неорганическую, свидетелями которого станем если и не мы сами, то наши дети и внуки.

Мы начали с описания человеческой истории как серии переходов между технологическими укладами со все более сложной системой разделения труда и все большим разнообразием продуктов для все большего числа людей. Так мы добрались до последнего перехода — от индустриального к цифровому укладу, в некоторой точке которого мы сегодня и находимся. Этот переход, по нашему мнению, обусловлен тем, что сложность индустриальной экономики в определенный момент превысила возможности человеческого разума и людям понадобились усилители мозга — компьютеры.

Далее мы проследили развитие компьютерных технологий с момента их появления — как они все глубже проникали в нашу жизнь, до определенного момента не слишком влияя на глобальную производительность труда. Это объясняется тем, что суммарные компьютерные мощности до недавнего времени на порядки отставали от мощности коллективного человеческого мышления. Но расширенный закон

Мура — увеличение суммарных компьютерных мощностей в 100 раз каждые 10 лет — делает неизбежным наступление момента, в который эти мощности наконец-то превзойдут человеческие. И похоже, что эту точку невозврата мы сегодня уже прошли.

В этой точке закономерно появляется замыкающая технология нового технологического пакета — машинное обучение. Используя все более обильные и дешевые компьютерные мощности, оно снимает ограничение на сложность программ, а тем самым и на сложность машинного интеллекта. Он уже больше не скован ограниченным программистским ресурсом — компьютеры обретают способность к самопрограммированию. Открывается путь к созданию базовой технологии нового уклада — искусственного интеллекта, о котором так долго и страстно спорили в прошлом и время которого наконец-то пришло.

Оценивая роль ИИ в новой цифровой экономике, мы попытались представить себе, за счет чего эффективность мировой экономики сможет подняться на новый уровень. Наш ответ: за счет перехода от массового производства к экономике «длинного хвоста» продуктов и услуг, нацеленных на индивидуальные потребности. Это увеличит емкость рынка и поможет рационально использовать творческий потенциал людей, работающих бок о бок сначала с интеллектуальными агентами, а потом и с разумными роботами.

Как и всякий резкий скачок в развитии, становление человеко-машинной цивилизации порождает массу рисков, вызванных неизбежными противоречиями между теми, кто выигрывает, и теми, кто проигрывает от ожидаемых в скором будущем перемен. Мы считаем, что последствия массовой безработицы, связанной с одновременным внедрением роботов во множестве отраслей, можно будет преодолеть за счет введения специальных «налогов на роботов». Гораздо большую опасность, с нашей точки зрения, представляет повышение социального напряжения и уровня агрессии в обществе, разделение человечества на враждующие лагеря, насаждение атмосферы осажденной крепости и новая гонка вооружений — развитие национальных систем полицейского и военного интеллекта, контроль над которыми в конечном итоге может быть утерян, причем с самыми печальными последствиями для всей нашей цивилизации с ее ядерными арсеналами.

Для минимизации этих рисков человеко-машинной цивилизации потребуется гораздо более совершенная, чем нынешняя, система коллективной безопасности, гармонизации групповых интересов и мирного разрешения конфликтов. Должны появиться платформы цифровой демократии, разрешающие все спорные вопросы в реальном времени с

учетом мнения всех заинтересованных сторон с гарантированным исполнением решений и соблюдением принятых в обществе социальных норм. Механизмы подавления агрессии и контроля общечеловеческих ценностей должны быть встроены в будущую цифровую инфраструктуру человечества, подобно системам самозащиты и саморегулирования в современных ядерных реакторах.

Отсюда возникают требования для разработчиков ИИ: они должны создать не просто ИИ человеческого уровня для роботов, заменяющих людей на производстве, но и «ИИ с человеческим лицом», а точнее — с человеческими ценностями и понятиями о жизни. Мы формулируем эту задачу как создание искусственной психики роботов, архитектурно как можно более близкой к человеческой со встроенными общественными инстинктами.

Это возвращает нас к исходной цели разработки ИИ: более глубокое понимание человеческого разума путем создания все более сложных действующих моделей искусственного разума. Соответственно мы попытались представить себе и читателю, как может выглядеть практическая программа по формированию человекоподобной искусственной психики и развитию новой науки машинного воспитания.

Прочертив вчерне план разработки ИИ, мы наконец получили возможность заглянуть в не столь отдаленное будущее, естественно, с разной степенью подробности и уверенности на разных временных масштабах. Мы выделили три фазы развития ИИ: дочеловеческого, человеческого и сверхчеловеческого уровня. В ближайшие годы будут внедряться и развиваться существующие технологии глубокого обучения. Затем произойдет большой скачок, связанный с появлением роботов с искусственной психикой. Начнется период воспитания в роботах человеческой системы ценностей в рамках единой человеко-машинной цивилизации.

В какой-то момент лет через 20–40 машинная компонента вырвется вперед, и долгосрочное планирование развития цивилизации перейдет под контроль искусственных стратегов. Однако если все будет хорошо, управление обществом будет по-прежнему определяться человеческими ценностями, унаследованными машинным разумом. Таким образом, отдаленное будущее человечества будет зависеть от того, какими ценностями в жизни будут руководствоваться наши дети и внуки, то есть во многом — и от нас с вами.

# Примечания

## Глава 1. Технологические уклады и технологические революции

1. Грейф А. Институты и путь к современной экономике: уроки средневековой торговли. — М.: Издательский дом ВШЭ, 2018. С. 129–131.
2. Кондратьев Н. Д. Большие циклы конъюнктуры и теория предвидения. Избранные труды. — М.: Экономика, 2002.
3. Глазьев С. Ю. Теория долгосрочного технико-экономического развития. — М.: ВладДар, 1993.
4. Тоффлер Э. Третья волна. — М.: АСТ, 2004.
5. Бродель Ф. Материальная цивилизация, экономика и капитализм, XV–XVIII вв. — М.: Весь мир, 2011.
6. Macaulay V. et al. (2005) Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. *Science* 308(5724): 1034–1036.
7. Herculano-Houzel S. (2016) *The Human Advantage: A New Understanding of How Our Brain Became Remarkable*. Cambridge: MIT Press.
8. Dunbar R. I. M. (1992) Neocortex Size as a Constraint on Group Size in Primates. *Journal of human evolution* 22(6): 469–493.
9. Rito T. et al. (2019) A dispersal of Homo Sapiens from Southern to Eastern Africa Immediately Preceded the Out-of-Africa Migration. *Scientific reports* 9(1): 1–10.
10. Харари Ю. Н. *Sapiens. Краткая история человечества*. — М.: Синдбад, 2016.
11. Даймонд Д. *Ружья, микробы и сталь: судьбы человеческих обществ*. — М.: АСТ, 2010.
12. *Война и военное дело* / Под ред. В. Н. Левичева. — М.: Государственное военное издательство, 1933.
13. Пашинский В. *Пространственно-временная динамика человеческих сообществ разного масштаба*. — М.: Товарищество научных изданий КМК, 2011.
14. Список стран по длине водных путей (<https://ru.wikipedia.org/wiki/>).

15. Технические данные парусника «Седов» ([http://www.sts-sedov.info/ru/istoriya/teh\\_dannye/](http://www.sts-sedov.info/ru/istoriya/teh_dannye/)).
16. Парусные корабли мира (<http://sailhistory.ru/zarozhdenie-gollandsko-go-flota-v-xvi-xvii-vv.html>).
17. Годовой прирост леса (<https://ru-ecology.info/term/2978/>).
18. Медоуз Д. Х., Рандерс Й., Медоуз Д. Л. Пределы роста. 30 лет спустя. — М.: Академкнига, 2007.
19. Шваб К. Четвертая промышленная революция. — М.: Эксмо-Пресс, 2018.
20. Foerster H. von, Mora P. M., Amiot L. W. (1960) Doomsday: Friday, 13 November, AD 2026. Science 132(3436): 1291–1295.

## Глава 2. Эволюция цифровых технологий: от первых ЭВМ до роботов

1. Lennie P. (2003) The Cost of Cortical Computation. Current biology 13(6): 493–497.
2. Howarth C., Gleeson P., Attwell D. (2012) Updated Energy Budgets for Neural Computation in the Neocortex and Cerebellum. Journal of Cerebral Blood Flow & Metabolism 32(7): 1222–1232.
3. Dehaene S. (2014) Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts. London: Penguin.
4. Яглом А. М., Яглом И. М. Вероятность и информация. — М.: Наука, 1973.
5. Brown T. et al. (2020) Language Models are Few-Shot Learners (<https://arxiv.org/abs/2005.14165/>).
6. Феллер В. Введение в теорию вероятностей и ее приложения. В 2 т. Т. 1. — М.: РИПОЛ классик, 2013.
7. Trends in the cost of computing (<https://aiimpacts.org/trends-in-the-cost-of-computing/>).
8. Гроув Э. Выживают только параноики: как использовать кризисные периоды, с которыми сталкивается любая компания. — М.: Альпина Паблишер, 2004.
9. Cortada J. W. (2018) Change and Continuity at IBM: Key Themes in Histories of IBM. Business History Review 92(1): 117–148.
10. Chollet F. (2019) On the Measure of Intelligence (<https://arxiv.org/abs/1911.01547/>).

## Глава 3. Экономика цифровых платформ

1. На чем зарабатывает Google помимо рекламы? (<https://marketinfo.pr>

## Глава 4. Социальные проблемы цифровой революции

1. US Productivity Growth 1995–2000, Understanding the Contribution of Information Technology Relative to Other Factors. McKinsey Global Institute, October 2001.
2. Muro M., Maxim R., Whiton J. (2019) Automation and Artificial Intelligence. How Machines are Affecting People and Places. Brookings.
3. Buterin V., Hitzig Z., Weyl E. G. Liberal Radicalism: a Flexible Design for Philanthropic Matching Funds. Available at SSRN: <https://ssrn.com/abstract=3243656> (December 2018).
4. Грейф А. Институты и путь к современной экономике: уроки средневековой торговли. — М.: Издательский дом ВШЭ, 2018.
5. Там же. С. 55.
6. Назаретян А. П. Цивилизационные кризисы в контексте Универсальной истории. — М.: Мир, 2004.
7. Назаретян А. П. Нелинейное будущее: сингулярность XXI века как элемент Мегаистории // Век глобализации. 2015. №2. С. 18–34.

## Глава 5. Путь к сильному искусственному интеллекту

1. On Defining Artificial Intelligence (Special Issue) (2020). Journal of Artificial General Intelligence 11(2): 1–99.
2. Wang P. (2019) On Defining Artificial Intelligence. Journal of Artificial General Intelligence 10(2): 1–37.
3. Анохин К. В. Последний великий рубеж наук о жизни // Экономические стратегии. 2010. Т. 12. №11. С. 56–63.
4. Kotseruba I., Tsotsos J. K. (2020) 40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications. Artificial Intelligence Review 53(1): 17–94.
5. Hassabis D. et al. (2017) Neuroscience-inspired Artificial Intelligence. Neuron 95(2): 245–258.
6. Marblestone A. H. et al. (2016) Toward an Integration of Deep Learning and Neuroscience. Frontiers in Computational Neuroscience 10: 94.
7. On Defining Artificial Intelligence (Special Issue) (2020). Journal of Artificial General Intelligence 11(2): 1–99.
8. Russell Stuart J., Norvig P. (2009) Artificial Intelligence: A Modern Approach. Englewood Cliffs: Prentice Hall.
9. Dobzhansky T. (1973) Nothing in Biology Makes Sense Except in the Light of Evolution. The American Biology Teacher 35(3): 125–129.

- [10.](#) Сеченов И. М. Элементы мысли. — СПб.: Питер, 2001. С. 230.
- [11.](#) Выготский Л. С. Мышление и речь / 5-е изд., испр. — М.: Лабиринт, 1999. С. 320.
- [12.](#) Анохин П. К. Очерки по физиологии функциональных систем. — М: Медицина, 1975.
- [13.](#) Bengio Y. From System 1 Deep Learning to System 2 Deep Learning. In: Thirty-third Conference on Neural Information Processing Systems, 2019 (<https://neurips.cc/Conferences/2019>).
- [14.](#) Albus J.S. (1971) A theory of cerebellar function. *Mathematical biosciences* 10(1–2): 25–61.
- [15.](#) Tsodyks M. (1999) Attractor neural network models of spatial maps in hippocampus. *Hippocampus* 9(4): 481–489.
- [16.](#) Hawkins J. (2021) *A Thousand Brains: A New Theory of Intelligence*. New York: Basic Books
- [17.](#) Marr D., Poggio T. (1976) *From Understanding Computation to Understanding Neural Circuitry*. Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.
- [18.](#) Haber S. N. (2016) Corticostriatal Circuitry. *Dialogues in Clinical Neuroscience* 18(1): 7–21.
- [19.](#) Herculano-Houzel S. (2016) *The Human Advantage: A New Understanding of How our Brain Became Remarkable*. Cambridge: MIT Press.
- [20.](#) Шумский С. А. Машинный интеллект. Очерки по теории машинного обучения и искусственного интеллекта — М.: РИОР, 2019.
- [21.](#) Brown T. B. et al. (2020) Language Models are Few-Shot Learners (<https://arxiv.org/abs/2005.14165>).
- [22.](#) Russell S. et al. (2015) Ethics of Artificial Intelligence. *Nature* 521(7553): 415–416.
- [23.](#) Human Brain Project (<https://www.humanbrainproject.eu/en/science/overview/>).
- [24.](#) Kaiser J. Senate bill would give NIH \$3 billion in 2020, or 7.7% boost (<https://www.sciencemag.org/news/2019/09/senate-bill-would-give-nih-3-billion-2020-or-77-.boost>).
- [25.](#) Notes from the AI Frontier: Modeling the Impact of AI on the World Economy. McKinsey Global Institute, 2018.
- [26.](#) Карелов С. Впереди ИИ-национализм и ИИ-национализация (<https://russiancouncil.ru>).
- [27.](#) Рассел С. Совместимость. Как контролировать искусственный интеллект. — М.: Альпина нон-фикшн, 2021.

## Глава 6. Будущая человеко-машинная цивилизация

1. Urban T. The AI Revolution: Our Immortality or Extinction ( <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html>).
2. Surowiecki J. (2004) The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, economies, societies and nations. Doubleday & Co.
3. Muehlhauser Muehlhauser L. The world's Distribution of Computation (Initial Findings) ( <https://intelligence.org/2014/02/28/the-worlds-distribution-of-computation-initial-findings/>).
4. Koomey J., Naffziger S. (2016) Energy Efficiency of Computing: What's next. Electronic Design 28 Nov. ( <https://www.electronicdesign.com/technologies/microprocessors/article/21802037/energy-efficiency-of-computing-whats-next>).
5. Sánchez-Bayo F., Wyckhuys K. A. G. (2019) Worldwide Decline of the Entomofauna: A Review of Its Drivers. Biological conservation 232: 8–27.
6. Ng A. Y. et al. Algorithms for inverse reinforcement learning ( <https://ai.stanford.edu/~ang/papers/icml00-irl.pdf>).
7. Рицолатти Д., Синигалья К. Зеркала в мозге. О механизмах совместного действия и сопереживания. — М.: Языки славянских культур, 2012.
8. Харари Ю. Н. Sapiens. Краткая история человечества. — М.: Синдбад, 2017.
9. Лю Цысинь. Задача трех тел. — М.: Эксмо, 2017.
10. Urban T. The AI Revolution: The Road to Superintelligence ( <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>).
11. Рассел С. Совместимость. Как контролировать искусственный интеллект. — М.: Альпина нон-фикшн, 2021.



[1] Всюду в этой книге мы будем пользоваться приблизительными величинами, не гонясь за точностью. Обычными будут оценки с точностью до порядка, в которых  $10^n \sim 0,3 \div 3 \times 10^n$ .

[2] Одна метрическая л. с. равна 735 Вт. С такой мощностью лошадь способна трудиться в течение 8-часового рабочего дня. При этом дневной рацион лошади (около 13 кг сена) = 45 000 ккал/день = 2,2 кВт, то есть ее рабочий КПД составляет примерно 33%, а с учетом отдыха — втрое меньше. Для неспешно передвигающегося человека рабочий КПД равен 40% (при мощности спокойной ходьбы 60 Вт). Иными словами, человек перерабатывает свою пищу эффективнее лошади, что неудивительно с учетом разницы в качестве пищи. Зато удельная мощность лошади больше, так как она мощнее нас в 12 раз, а весит всего в 7 раз больше.

[3] Древесина — 3500 ккал/кг, уголь — 7000 ккал/кг, нефть/газ — 9500 ккал/кг.

[4] Общий объем «жадных» до энергии органов — сердца, печени, ЖКТ и мозга — лимитирован, так что мозг мог вырасти только за счет сокращения желудочно-кишечного тракта, так как и сердце, и печень пропорциональны объему остального тела.

[5] При рационе 13 кг/день за 365 дней лошади требуется 5 т сена.

[6]  $150 + 60 = 210$  Вт при неолите против 60 Вт при палеолите.

[7] Если принять для оценки современный среднемировой уровень теплоснабжения 0,4 кВт/чел. ( $1,2 \times 10^4$  МДж/чел. в год). Учитывая теплотворную способность дров  $15 \times 10^3$  МДж/т, для отопления требуется 0,8 т, или около  $1 \text{ м}^3$  дров/чел. в год.

[8] Площадь вокруг города радиусом 20 км ( $\pi R^2 = 1200 \text{ км}^2 = 120\,000$  га) делится на пашню (30 000 га), луг (60 000 га) и лес (30 000 га). При плотности 2 чел./га пашни/леса получаем население 60 000 чел.

[9] Более того, движение европейской цивилизации от Средиземноморья на север Европы было вызвано не в последнюю очередь истощением южных лесов (см.: Эткинд А. М. Природа зла. Сырье и государство. — М.: Новое литературное обозрение, 2020. С. 40).

[10] Британский линейный корабль XVIII века требовал 4000 дубовых стволов, или 40 га зрелого леса. А в 1850 году уголь давал Англии столько энергии, сколько давал бы лес на 150% ее территории (см.: Эткинд А. М. Природа зла. Сырье и государство. — М.: Новое литературное обозрение, 2020. С. 45).

[11] Первую промышленную революцию подготовила предшествовавшая ей «нулевая» промышленная революция в Нидерландах, развивших промышленную технологию кораблестроения в производственных кластерах (см.: Щедровицкий П. Г. Три индустриализации России. — М.: Terra Fantastica, 2018).

[12] Сеть современных автодорог (70 млн км) в 100 раз больше протяженности судоходных водных путей.

[13] Римский клуб — международная общественная организация, созданная в 1968 году для привлечения внимания к глобальным мировым проблемам.

[14] Математические модели динамики численности населения Земли (<http://csef.ru/ru/politica-i-geopolitica/326/matematicheskie-modeli-dinamiki-chislennosti-naseleniya-zemli-7997>).

[15] Напомним, что сложность социальной системы сообщества тоже растет квадратично с его размером — по числу возможных связей. Логично, что количество знаний, требующихся для управления, должно соответствовать сложности управляемой системы.

[16] Имеются и более точные методики подсчета и сравнения сложности экономик разных стран. См.: Hidalgo C., Hausmann R. (2009) The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences*

[17] Фактически объем мозга современного человека даже несколько меньше, чем у первых *Homo sapiens*, так как человечество само себя «одомашнило», и цивилизованная жизнь уже не предъявляет таких высоких требований к нашему мозгу.

[18] Первоначально он был сформулирован Муром для числа транзисторов на кристалле, но аналогичным образом росли очень многие параметры компьютеров. Мы предпочитаем формулировку в терминах компьютерных мощностей, так как именно она имеет значение с точки зрения экономики.

[19] А именно: запомненная системой информация есть двоичный логарифм числа возможных состояний системы. В случае современных устройств памяти, состоящих из большого числа независимых однотипных ячеек, каждая из которых может находиться в одном из двух состояний, объем памяти в *битах* равен просто числу таких двоичных ячеек.

[20] Более точная оценка разнообразия паттернов — число сочетаний  $\binom{N}{N/K} = (Ke)^{N/K}$ , но мы предпочитаем интуитивно понятные качественные оценки с точностью до порядка величин.

[21] FLOPS — Floating point operations per second (операции с действительными числами в секунду). 1 FLOPS = 2 ÷ 8 байт/с в зависимости от точности вычислений.

[22] Умножив объем всех накопленных знаний на среднюю частоту срабатывания нейронов в мозге, с которой он их сканирует (альфа-ритм), получим  $10^{11}$  байт/с — вычислительную мощность сознательного мышления, меньшую, но близкую к общей вычислительной мощности мозга  $10^{11}$  FLOPS.

[23] После появления компьютерных сетей и последующего перехода всех средств связи на цифру сформировался единый сектор *информационно-*

компьютерной техники — ИКТ. Сегодня стоимость чипов в коммутаторах и маршрутизаторах на линиях связи превышает стоимость чипов в самих компьютерах.

[24] Оценка рынка компьютеров в \$300 млн в 1950 году, в эпоху их штучного изготовления по госзаказам, до начала коммерческого производства в 1953 году, получена с учетом предполагаемого выпуска 10 компьютеров в год стоимостью \$3 млн за штуку (\$30 млн в 2020 году) каждый.

[25] Название метода обратной прогонки появилось вследствие именно такой организации вычислений до электронной эпохи.

[26] В СССР первые ЭВМ появились в 1953 году, с отставанием в 8 лет, и все расчеты термоядерной бомбы выполнялись вручную девушками-вычислителями. К 1967 году с появлением отечественных БЭСМ-6 это отставание было практически ликвидировано, но из-за выбранной советским руководством стратегии копирования американской вычислительной техники с середины 1970-х годов технологическое отставание электронной промышленности СССР начало катастрофически нарастать, и сегодня Россия занимает лишь 0,3% мирового рынка электроники — на порядок ниже доли России в мировом ВВП.

[27] Ее создание оценивается в \$5 млрд в 1960 году (\$45 млрд по современным ценам). Cortada J. W. (2018) Change and Continuity at IBM: Key Themes in Histories of IBM. *Business History Review* 92(1): 117–148.

[28] Всего PDP-11 было продано 600 000 штук, а более дорогих VAX — более 400 000 штук, начиная с 1978 года.

[29] Со своими 2300 транзисторами микропроцессор Intel 4004 уже не уступал по вычислительной мощности ENIAC.

[30] 30 кадров/с с разрешением  $320 \times 200$  пикселей =  $2 \times 10^6$  FLOPS.

[31] Этому также способствовал переход с шерсти на хлопок, поскольку

растительные белки на порядок дешевле животных: 1 га земли дает в 12 раз больше хлопка, чем шерсти (см.: Эткинд А. М. Природа зла. Сырье и государство. — М.: Новое литературное обозрение, 2020. С. 365).

[32] Раннее появление автономных автомобилей в качестве первых ласточек будущего рынка роботов связано с относительной простотой управления ими — всего две степени свободы: угол поворота и ускорение. Иными словами, ключевой технологией автономного транспорта оказалось именно машинное зрение.

[33] Разновидностью такого решения является введение гарантированного базового дохода, который проще администрировать, чем описанный выше налог на роботов. Другую альтернативу — наращивание долгов беднеющего населения — мы наблюдали в США с конца 1970-х одновременно с массовым оттоком рабочих мест в Китай и ЮВА, своего рода репетицию будущей массовой ИИ-безработицы.

[34] Очень показателен в этом плане британский сериал «Захват» («The Capture»), где власти используют технологию DeepFake для коррекции видеоматериалов с камер слежения в интересах следствия.

[35] В США на июль 2021 года назначен показательный воздушный бой живого летчика с полностью автономным беспилотным истребителем.

[36] «Intelligence is the rate at which a learner turns its experience and priors into new skills at valuable tasks that involve uncertainty and adaptation» (см.: Chollet F. (2019) On the Measure of Intelligence. arXiv preprint arXiv:1911.01547).

[37] Так, априорные знания о существовании предметов ускоряют машинное обучение играм «Атари» в тысячи раз (см.: Agnew, W., Domingos, P. (2020) Self-Supervised Object-Level Deep Reinforcement Learning. <https://arxiv.org/abs/2003>).

[38] «В настоящее время все существующие методы иерархического планирования опираются на сгенерированные человеком иерархии

абстрактных и конкретных действий. Мы еще не понимаем, как такие иерархии могут быть получены путем обучения» (см.: Рассел С. Совместимость. Как контролировать искусственный интеллект. — М.: Альпина нон-фикшн, 2021.).

[39] Например, с помощью кинематической модели крысиного тела (см.: Merel J. et al. (2019) Deep Neuroethology of a Virtual Rodent. arXiv preprint arXiv:1911.09451).

[40] На самом деле основные вычислительные мощности сосредоточены в специализированных процессорах GPU, TPU и ASIC и удваиваются каждые 12 месяцев (см.: [https://en.wikipedia.org/wiki/Technological\\_singularity](https://en.wikipedia.org/wiki/Technological_singularity)).

[41] Заметим, что достижение предела вычислительной мощности еще не означает стагнации ИИ, так как эти мощности будут использоваться для постоянного приращения знаний.

[42] Уже сегодня суммарное число приложений на платформах Google, Apple, Windows и Amazon приближается к 10 млн.

[43] Сегодня на возобновляемую энергетику приходится  $2/3$  вводимых мощностей и  $3/4$  инвестиций в энергетику. См.: «Перспективы глобального перехода к возобновляемой энергетике», REN21, 2019 ([https://www.ren21.net/wp-content/uploads/2019/05/1900916\\_GSR\\_2019\\_Perspectives\\_Russian.pdf](https://www.ren21.net/wp-content/uploads/2019/05/1900916_GSR_2019_Perspectives_Russian.pdf)).

[44] Сегодня доля промышленных роботов относительно работников-людей на производстве составляет в среднем по миру около 1% и достигает 3–7% в наиболее роботизированных экономиках Южной Кореи, Сингапура, Японии и Германии.

[45] Этот алгоритм обучения коллективного разума практически идентичен обучению методом обратного распространения ошибки (error back propagation) в искусственных нейросетях.

Научный редактор *Сергей Филонович, д-р физ.-мат. наук*  
Редактор *Полина Суворова*  
Издатель *П. Подкосов*  
Руководитель проекта *А. Шувалова*  
Корректоры *И. Панкова, Е. Сметанникова*  
Компьютерная верстка *А. Фоминов*  
Художественное оформление и макет *Ю. Буга*  
*Иллюстрация на обложке Shutterstock*

© Шумский С., 2021

© ООО «Альпина нон-фикшн», 2021

© Электронное издание. ООО «Альпина Диджитал», 2021

### **Шумский С.**

Воспитание машин: Новая история разума / Сергей Шумский. — М.: Альпина нон-фикшн, 2021.

ISBN 978-5-0013-9499-0