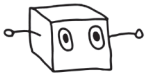
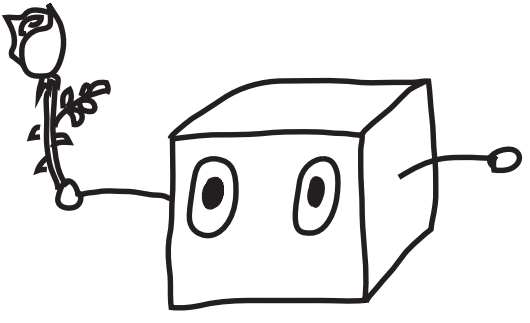


Оглавление

Посвящается читателям моего блога,
которые смеялись над каждой глупостью,
рисовали для меня чумовых существ, нашли всех
жирафов и готовили печенье по рецепту от нейросети.
Спасибо, что смирились даже с брауни из хрена.
Посвящается также моим родным — за то,
что вы самые большие мои фанаты.

ВВЕДЕНИЕ

ИИ повсюду



Я на самом деле не ставила целью проекта научить ИИ* флиртовать.

Вообще-то, я сделала уже порядочно своеобразных проектов на тему искусственного интеллекта. В блоге под названием AI Weirdness** я рассказывала о том, как учила искусственный разум придумывать имена для котов — среди них попадались на редкость неудачные, например Мистер Бубенцы или Рыгуша — и просила сочинять новые рецепты блюд, а получала такие, в которых требовалось взять «почищенный розмарин» или горсть

---

* ИИ — искусственный интеллект. — *Здесь и далее примечания переводчика, если не указано иное.*

** «Странности ИИ», https://aiweirdness.com.

битого стекла. Но попытка натренировать компьютер обольщать людей — уже нечто совсем иное.

ИИ обучается на примерах — в этом случае на пикаперских фразах. Проблема в том, что в обучающий набор входили фразы, собранные мной из разных уголков Интернета, и все они были ужасны. Там попадались высказывания из всех категорий, начиная с тупых каламбуров и заканчивая неприличными намеками. После того как ИИ натренировался сочинять подобные фразы, их можно было бы производить на свет божий тысячу за тысячей одним нажатием кнопки. Однако, как готовый впитать любые впечатления ребенок, искусственный интеллект не знает, что следует повторять, а что нет. Он начинает с чистого листа, ничего не зная о том, что такое пикаперские фразы (и даже что вообще такое язык), и затем учится на примерах, изо всех сил стараясь уловить закономерности и воспроизвести то, что от него требуют. В том числе грубость. Просто он не знает ничего получше.

Я подумывала бросить затею, но пост в блог сам себя не напишет, так что пришлось пересилить себя и потратить уйму времени на сбор образцовых пикаперских фраз. Итак, я начала обучать алгоритм. ИИ принялся искать закономерности и шаблоны в примерах, изобретая и проверяя правила, которые помогли бы ему спрогнозировать, какие нужны буквы и в каком порядке они должны располагаться, чтобы получилась пикаперская фраза. Наконец тренировка окончилась. Испытывая некоторый мандраж, я попросила ИИ выдать несколько фраз.

```
Наверное, ты треугольник? Потому что
  только ты здесь чего-то стоишь.
Эй, детка, ты, должно быть, ключ? Потому
  что я выдерживаю твой свисток.
```

```
Ты свеча? Это потому что ты такая зной-
  ная от видов с тобой.
Ты так прекрасна, что ты говоришь лету-
  чую мышь на меня с деткой.
Ты — нечто, но ты мне нравишься.
```

Я удивилась и порадовалась. Виртуальный мозг ИИ (примерно того же уровня сложности, что и у червя*) неспособен был уловить оттенки смыслов в наборе данных, распознать женоненавистничество или отсеять низкосортные шутки. Но те закономерности, которые ему удалось выявить, позволили ему добиться наилучшего результата… причем неожиданным, может, даже лучшим способом: решая самую главную задачу — как заставить улыбнуться незнакомку.

Хотя для меня итоговые фразы свидетельствовали о бесспорном успехе, скудоумие моего ИИ-партнера, вероятно, удивит того, кто знает об искусственном интеллекте лишь из заголовков современных новостей или из научной фантастики. Очень часто компании заявляют, что ИИ способен воспринимать нюансы человеческой речи так же хорошо, как люди, и даже лучше, или, что еще чуть-чуть, и он заменит людей во многих профессиях. Скоро искусственный интеллект будет повсюду, трубят пресс-релизы. Это одновременно и правда, и нет.

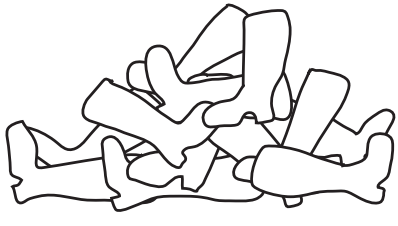
На самом деле ИИ *уже* повсюду. Он формирует все, с чем мы сталкиваемся онлайн, определяет, какую рекламу мы видим в Интернете, предлагает видеоролики и в то же время распознает ботов в социальных сетях

---

* Например, у червя *Caenorhabditis elegans*, геном и строение которого изучены в мельчайших подробностях, мозг содержит ровно 302 нейрона. — *Прим. науч. ред.*

и вредоносные сайты. Компании используют ИИ-программы для сканирования резюме кандидатов на вакансии, а в банках искусственный интеллект решает, кому можно выдавать кредит. Беспилотные автомобили со встроенным ИИ уже проехали миллионы километров (прибегая к помощи человека лишь в моменты замешательства). В наших смартфонах ИИ распознает голосовые команды, автоматически отмечает лица людей на фотографиях и даже применяет к видеопотоку фильтры — благодаря им у нас в кадре, например, вырастают чудесные кроличьи уши.

По опыту мы знаем, что тот ИИ, с которым мы сталкиваемся каждый день, крайне далек от совершенства. Приложения, предлагающие рекламу, без конца преследуют нас в браузерах рекламой ботинок, хотя мы их уже купили. Фильтры в электронной почте иной раз пропускают откровенно мошеннические послания и, наоборот, в самый неподходящий момент убирают в папку спама важное для нас письмо.



ИИ все больше влияет на нашу повседневную жизнь, и его причуды начинают проявляться в последствиях таких масштабов, что уже нельзя говорить о простом неудобстве. Рекомендательные алгоритмы YouTube подсовывают зрителям контент все более полярного характера: от популярных каналов новостей к видеороликам с пропагандой ненависти и теорий заговора

ведет дорога всего из нескольких кликов[1]. Алгоритмы, принимающие решения о досрочном освобождении заключенных, о предоставлении кредитов, программы скрининга резюме не беспристрастны и иногда страдают от тех же предрассудков, что и люди, которых они призваны заменить, может, даже в большей степени. Интеллектуальные системы наблюдения неподкупны, но у них также не возникнет возражений, если от них потребуют сделать нечто аморальное. Кроме того, они могут выдавать ошибки из-за неправильного использования или взлома. Исследователи выяснили, что иногда такая вроде бы малозначительная вещь, как небольшая наклейка, способна заставить систему распознавания принять пистолет за тостер и что сканер отпечатков пальцев, не очень продвинутый в плане безопасности, можно более чем в 77% случаев обдурить с помощью единственного универсального отпечатка*.

Одни люди представляют ИИ более умным, чем он есть на самом деле, и способным делать то, что возможно лишь в научной фантастике. Другие говорят, что создали беспристрастный ИИ, в то время как в его поведении просматриваются явные и измеримые искажения. А еще нередко за работу ИИ выдается результат деятельности людей. Как жителям Земли, нам нужно научиться не попадаться на эту удочку. Нужно понять, как используются наши данные и чем является ИИ, а чем не является.

На сайте AI Weirdness я много времени посвящаю различным забавным экспериментам с искусственным интеллектом. Иногда я заставляю его делать необычные вещи, например, имитировать эти пикаперские фразы. Или стараюсь вывести его из зоны комфорта, как в тот

---

* Аналога так называемого «мастер-ключа». — *Прим. науч. ред.*

раз, когда я «скормила» алгоритму распознавания образов картинку с Дартом Вейдером и просто спросила его, что он увидел: алгоритм объявил, что Дарт Вейдер — это дерево, и начал спорить со мной, отстаивая свою точку зрения. В результате экспериментов я выяснила, что даже самое четкое задание может поставить ИИ в тупик, как если бы вы над ним подшутили. Но оказывается, что разыгрывать ИИ — то есть подсовывать ему задачу и наблюдать, как он ломает о нее зубы, — крайне поучительно и помогает узнавать о нем больше.

Как вы увидите в этой книге, зачастую внутри ИИ происходят настолько странные и запутанные процессы, что анализ выходных данных становится одним из немногих способов выяснить, что искусственный интеллект понял, а в чем ужасно ошибается. Когда вы просите ИИ нарисовать кота или написать шуточную фразу, он начинает делать те же ошибки, что и в процессе распознавания отпечатков пальцев или сортировки медицинских фотоснимков. Вот только тут вы сразу поймете, что что-то пошло не так, если у кота на картинке окажется шесть лап или шутка лишится ключевой фразы в конце. Ну и еще это безумно смешно.

В попытках вывести ИИ из зоны комфорта и заставить его заниматься человеческими делами я требовала от него написать первые строки романа, узнавать на изображениях овец в необычных местах, давать имена морским свинкам и вообще чудить по-всякому. Это позволяет очень многое узнать о том, в чем ИИ хорош, в чем — не очень, а чего ему, скорее всего, не удастся достичь за время моей или вашей жизни.

И что же я узнала?

Пять принципов странности ИИ:

- ИИ опасен не потому что он умен, а потому что он умен недостаточно;

- по силе интеллекта ИИ находится примерно на уровне червя;
- ИИ в действительности не понимает задачу, которую вы перед ним ставите;
- но: ИИ будет делать *в точности* то, что вы от него хотите, — по крайней мере, изо всех сил постарается;
- и еще: ИИ всегда выбирает путь наименьшего сопротивления.

Так давайте же войдем в странный мир искусственного интеллекта. Мы узнаем, что можно назвать ИИ, а что — нельзя. Выясним, в чем он хорош и в чем обречен на поражение. Поймем, почему ИИ будущего, вероятно, будут похожи не на робота C-3PO, а скорее на рой насекомых. Разберемся, почему беспилотный автомобиль не поможет спастись во время зомби-апокалипсиса. Узнаем, почему никогда не надо вызываться проверять работу ИИ, сортирующего сэндвичи, а еще узнаем о ходячих ИИ, которые будут делать что угодно, только не ходить. Все эти истории позволят нам понять, как работает искусственный интеллект, как он думает и почему делает наш мир еще более странным.

ГЛАВА 1

Что такое ИИ?

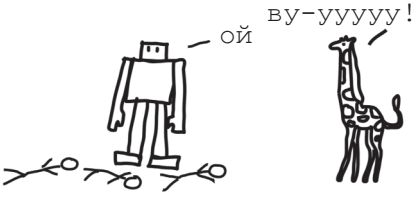
Если вам кажется, что ИИ уже повсюду, то это отчасти потому, что слова «искусственный интеллект» могут означать разные вещи — зависит от того, читаете вы фантастический роман или пытаетесь продать новое приложение для научных исследований. Когда некто заявляет, что у него есть чат-бот с ИИ, надо ли ожидать, что у этого чат-бота будет свое мнение и чувства, как у вымышленного С-3РО? Или это всего лишь алгоритм, научившийся догадываться, как именно люди, скорее всего, отреагируют на ту или иную фразу в диалоге? Или

это электронная таблица, которая ищет слова из вашего вопроса в библиотеке заранее подготовленных ответов? А может, это человек, сидящий где-то в далекой стране на скромной зарплате и печатающий вам сообщения? Или это полностью подчиненный сценарию диалог, где человек и ИИ зачитывают фразы, как актеры в пьесе? Все эти вещи определяли как искусственный интеллект — отсюда и путаница.

В рамках своей книги я буду подразумевать под термином в основном то, что сейчас под ИИ подразумевают программисты, — вид программ, построенных на основе алгоритмов машинного обучения. Ниже я привела целую кучу терминов, о которых мы поговорим дальше, и разнесла их по категориям.

Все, что называют ИИ

ИИ в этой книге	**Тоже есть в книге, но это не ИИ**
Алгоритмы машинного обучения	ИИ из научной фантастики
Методы глубокого обучения	Основанные на правилах программы
Нейронные сети	Люди в костюмах роботов
Нейросети с обратными связями (рекуррентные)	Роботы, действующие по сценарию
Цепи Маркова	Люди, которые за деньги выдают себя за ИИ
Случайный лес	Разумные тараканы
Генетические алгоритмы	Жирафы-призраки
Генеративно-состязательные сети	
Обучение с подкреплением	
Предиктивный набор текста	
Волшебные машины для сортировки сэндвичей	
Невезучие роботы-убийцы	



Все, что я здесь называю искусственным интеллектом, также можно назвать алгоритмами машинного обучения, давайте поговорим о том, что это такое.

Тук-тук, кто там?

Чтобы найти ИИ в дикой природе, важно понять, в чем же разница между **алгоритмами машинного обучения** (именно это понимается здесь под ИИ) и традиционными программами (программисты их называют **основанными на правилах**). Если вы когда-нибудь программировали хотя бы на базовом уровне или обращались к HTML, чтобы создать дизайн сайта, значит, вы использовали основанную на правилах программу. Вы задаете список команд или правил на понятном компьютеру языке, и компьютер делает в точности то, что вы говорите ему делать. Чтобы решить задачу с помощью такой программы, вам потребуется понять, какие шаги должна выполнить программа, чтобы достичь цели, и как именно их описать.

Алгоритм машинного обучения сам додумывается до правил методом проб и ошибок, оценивая, насколько приблизился к поставленным программистом целям. Целью может быть воспроизвести что-то по примерам, достичь определенного счета в игре или что угодно еще. Пытаясь выполнить задачу, ИИ способен выявить такие правила и взаимосвязи, о существовании которых программист даже не подозревал. Программирование ИИ больше похоже на обучение ребенка, чем на разработку программы.

Программирование на основе правил

Предположим, я решила с помощью традиционного программирования научить компьютер выводить шутки «Тук-тук, кто там?». Вначале я должна выявить все правила. Я проанализирую структуру подобных шуток

и выясню, что все они строятся по определенной формуле, вот такой:

```
Тук-тук.
Кто там?
[Имя]
Как[ой/ая/ое] [имя]?
[Имя] [Ключевая фраза]
```

Теперь, когда я определилась с формулой, оказывается, что программа должна заполнить два пропуска: [Имя] и [Ключевая фраза]*. Теперь задача сводится к тому, чтобы произвести эти элементы. Но правила все равно нужны.

Я могу подобрать список имен и подходящих ключевых фраз, например:

Имена	**Ключевые фразы шутки или каламбура**
Рада	вам предложить установку радиаторов отопления!
Тихон	егодяй, сколько можно сверлить!
Яна	вас жаловаться буду!
Арина	тебя не ори, ты ничего не слышишь, глухомань.
Олег	сандр!

Теперь компьютер может выдавать шутки «Тук-тук, кто там?», выбирая пару [Имя] и [Ключевая фраза] из

---

* Актуально для английского языка. В русскоязычном варианте шутки программе, как видно на схеме, придется согласовать в роде местоимение «какой» со словом, подобранным для элемента [Имя]. То есть она должна будет заполнить не два, а три пропуска. — *Прим. ред.*