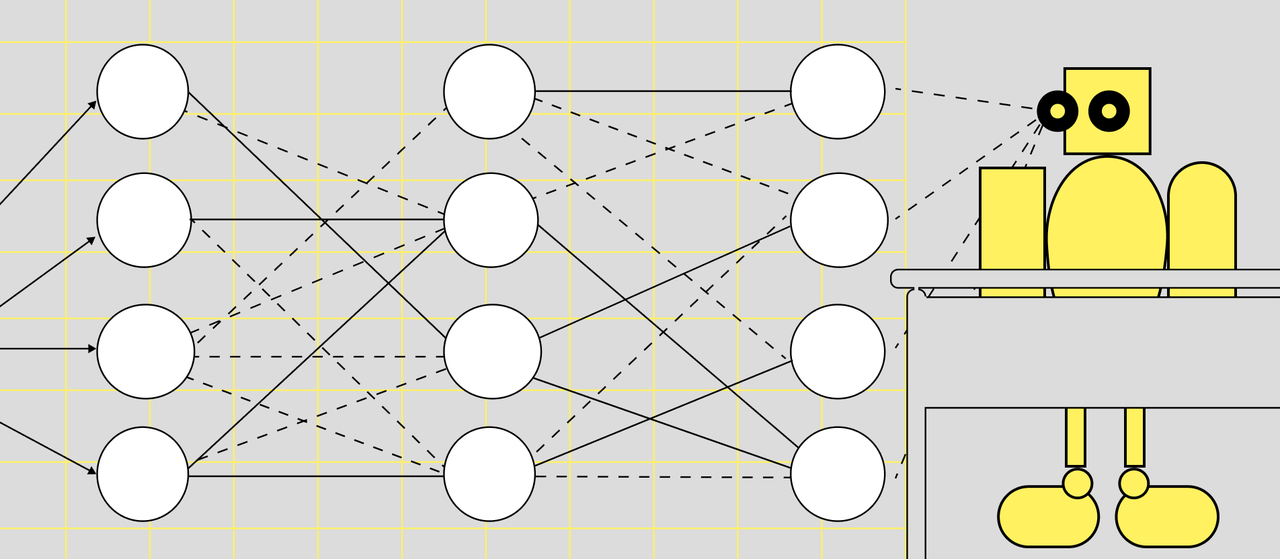
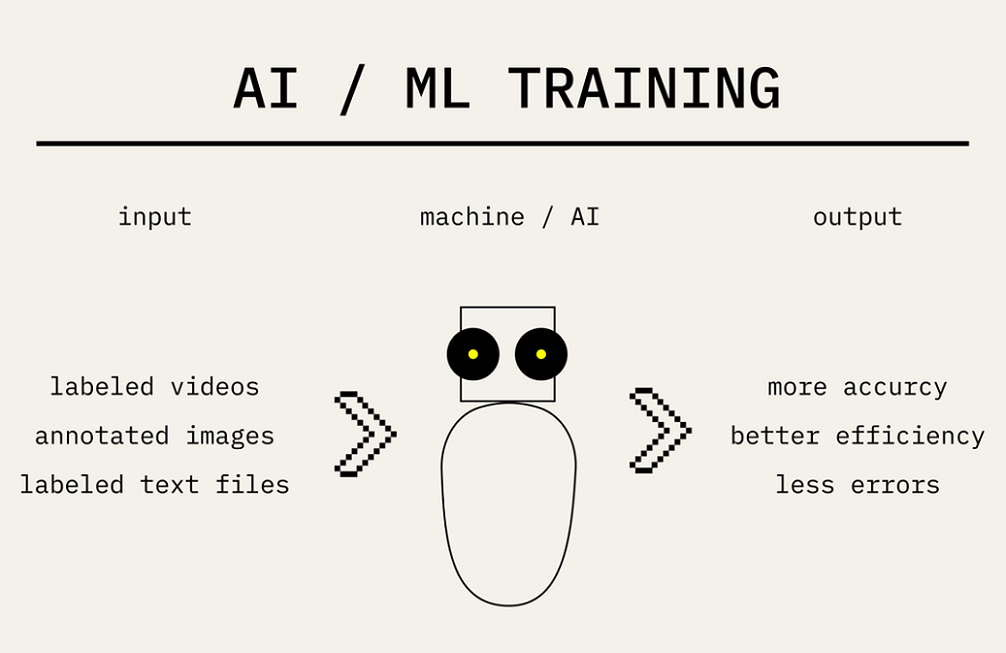
**Что такое машинное обучение**



* [Как работает машинное обучение](https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/#kak-rabotaet-mashinnoe-obuchenie)
* [Как и когда появилось машинное обучение](https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/#kak-poyavilos-mashinnoe-obuchenie)
* [Почему машинное обучение оказалось так важно](https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/#vazhnost-mashinnogo-obucheniya)
* [Где используется машинное обучение](https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/#gde-ispolzuetsya-mashinnoe-obuchenie)
* [Кто и как обучает модели](https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/#kto-i-kak-obuchaet-modeli)

**Как работает машинное обучение**

Когда мы записываемся к врачу через чат-бота или просим Алису включить рок, то почти не задумываемся, как это происходит: действия кажутся очень простыми. На самом деле за каждым из них стоит сложный процесс, включающий элементы машинного обучения.  
  
Представьте, что вас послали в магазин с поручением купить три килограмма картошки. Для взрослого человека это максимально простая задача. Но что будет, если отправить за килограммом картошки двухлетнего ребёнка?   
  
1. Скорее всего, сначала придётся рассказать и показать, что такое картошка и как отличить её от морковки.   
2. Потом объяснить, что такое килограмм и чем отличается от других весовых категорий.   
3. Научить присваивать числа товарам: килограмм картошки — 50 рублей.  
4. Рассказать правильную последовательность действий: прийти, взять, взвесить, принести к кассе, оплатить.   
  
И даже при самой последовательной и точной инструкции вероятность, что ребёнок с первого раза справится с задачей, будет очень маленькой. Ведь он ко в сему прочему ещё и плохо говорит. 

Пример с ребёнком и килограммом картошки — то, как работает машинное обучение, или ML, Machine Learning. ML — это класс методов, в которых машину, алгоритм, учат думать и действовать как человек на основе полученного опыта или данных. Читать, писать, рисовать, отличать рэп от рока, а картошку от морковки.

Машинное обучение — не то же самое, что программирование. Программист создаёт для машины алгоритм: прописывает чёткую последовательность действий, которая приведёт к нужному результату.

ML-специалист, который занимается обучением модели, не пишет для машины программу. Он передаёт данные и пытается объяснить, что хочет получить на выходе. У алгоритма нет заданного ответа, к которому нужно прийти, он знает лишь, как построить модель, отвечающую на поставленный вопрос. Цель машинного обучения — научить модель саму находить решение.

**Как и когда появилось машинное обучение**

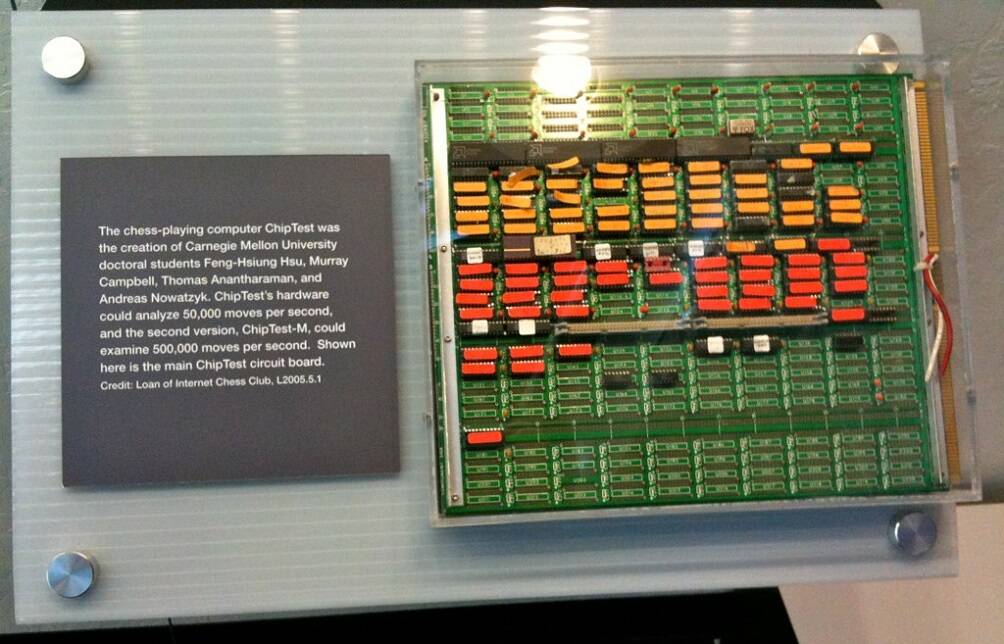
Понятие машинного обучения придумал американский исследователь Артур Самуэль, работавший в IBM. **В 1959 году он создал первую программу по игре в шашки, которая умела играть сама с собой и обучаться самостоятельно**.

Самуэль был не первым, кто начал работать с машинным обучением. До него уже существовал целая искусственная нейросеть: её создал в 1958 году американский нейрофизиолог Фрэнк Розенблатт.

Новый нейрокомпьютер мог распознавать образы и предсказывать погоду. Но эта технология не получила развития: коллеги-учёные раскритиковали перцептрон Розенблатта и назвали неспособным к обучению в большинстве случаев.

А вот игровое направление, освоенное Самуэлем, наоборот, стало развиваться очень активно.

**В конце 80-х аспиранты университета Карнеги-Меллон спроектировали машину для игры в шахматы ChipTest**. Она стала прототипом легендарного суперкомпьютера Deep Blue, обыгравшего Гарри Каспарова в 1996 году.



  ChipTest мог анализировать до 50 тысяч движений в секунду. [Источник](https://www.flickr.com/photos/cgriego/3693301436/) — flickr

Матч с Каспаровым стал настоящей революцией. До Deep Blue алгоритмы не могли тягаться с человеком. Они хорошо считали ходы, но были лишены интуиции и импровизации. Из-за чего не умели распознавать те из них, в которых соперник жертвует материал, чтобы получить преимущество.

Конечно, IBM, создавшая Deep Blue, не одарила машину настоящей интуицией. Но она создала компьютер мощности, которая позволила рассчитывать ещё большее количество вариантов. Что, в итоге, помогло «почувствовать» маневр Каспарова в очередной партии матча. Когда гроссмейстер «подставил» фигуру, Deep Blue вместо того, чтобы принять жертву за чистую монету, ответным ходом отклонил ее. Это был настоящий шок: никто не ожидал, что машина сможет просчитать тактику противника и среагировать так, как среагировал бы человек.



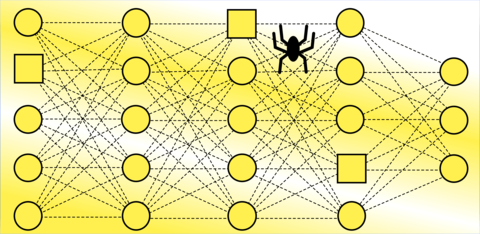
Так выглядел суперкомпьютер Deep Blue, выигравший у Каспарова. [Источник](https://ru.wikipedia.org/wiki/Deep_Blue)— Wikipedia

Вскоре после этого в истории искусственного интеллекта и машинного обучения начался новый виток. В 2011 году Google основал Google Brain — подразделение, занимавшееся проектами в области ИИ. Спустя ещё три года свои платформы по машинному обучению появились у Amazon и Microsoft, а Facebook внедрила в работу DeepFace — алгоритм, умеющий распознавать человеческие лица.

Материал по теме:

Для чего строят и обучают нейросети в IT

[Подробнее](https://practicum.yandex.ru/blog/chto-takoe-neyronnye-seti/?from=blog-practicum-perelinkovka" \t "_blank)

[[](https://practicum.yandex.ru/blog/chto-takoe-neyronnye-seti/?from=blog-practicum-perelinkovka)](https://practicum.yandex.ru/blog/chto-takoe-neyronnye-seti/?from=blog-practicum-perelinkovka)

**Почему машинное обучение оказалось так важно**

Принцип машинного обучения позволяет создавать компьютеры и программы, которые думают как человек. Но при этом, в отличие от человека, они не устают, могут меньше ошибаться, работают с любым количеством данных и оценивают их непредвзято. Это открывает огромное поле для [возможностей искусственного интеллекта](https://practicum.yandex.ru/blog/chto-takoe-iskusstvennyi-intellekt/).   
  
С помощью машинного обучения можно создавать множество сервисов и программ. От самые простых, которыми мы пользуемся в повседневной жизни — навигаторов, камер с распознаванием лиц, сервисов рекомендаций, — до сложных инструментов, которые нужны в промышленности или безопасности.



Так работает технология распознавания лиц, разработанная китайской компанией Hanvon — она способна распознавать даже людей в масках. [Источник](https://www.csis.org/blogs/trustee-china-hand/coming-focus-chinas-facial-recognition-regulations)— Hanvon

Например, когда стриминговый сервис предлагает вам подборку фильмов, а дорожная камера распознаёт преступника в толпе, это работа машинного обучения. Именно оно отвечает за то, чтобы получить данные, классифицировать их и сделать точный прогноз или вывод.

**Где используется машинное обучение**

Практически везде. Способности искусственного интеллекта запоминать информацию, анализировать и предсказывать данные используются в маркетинге, финансовой сфере, медицине, демографии и сфере безопасности.

Вот несколько примеров, как и где используются модели машинного обучения.

**Банки**. Программы банковского скоринга решают вопрос с обработкой огромного количества кредитных анкет. Специалисты создают модель, которая автоматически рассчитывает кредитный рейтинг, оценивает платёжеспособность клиента и определяет, одобрить выдачу кредита или отказать в ней.

**Маркетинг**. Когда Алиса предлагает персональный плейлист, а Facebook показывает персональную ленту, это классическое применение машинного обучения в задаче рекомендации. Другой пример — магазины без касс и продавцов, в которых за счёт машинного обучения алгоритмы учатся соотносить клиента с его виртуальной корзиной и отслеживать перемещения товаров на полках.



25 января 2022 года ВкусВилл открыл в Москве первый магазин без касс и продавцов, где всю работу по обслуживанию выполняет искусственный интеллект. [Источник](https://vkusvill.ru/news/budushchee-bez-kass-vkusvill-zapustil-novyy-format-magazina.html)— ВкусВилл

**Медицина**. Один из самых громких примеров — фундаментальное открытие, которое в 2020 году совершил алгоритм AlphaFold. Он смоделировал процесс сворачивания белка, решив одну из самых сложных биохимических задач столетия. Благодаря модели учёные получили возможность предотвращать развитие инфекций, когнитивных и нейродегенеративных заболеваний — Паркинсона, Альцгеймера и других.

**Сельское хозяйство**. С помощью машинного обучения созданы модели, умеющие анализировать состав почвы, рассчитывать нужное количество удобрений, предсказывать урожайность и даже прогнозировать надои молока у коров.

**Кто и как обучает модели**

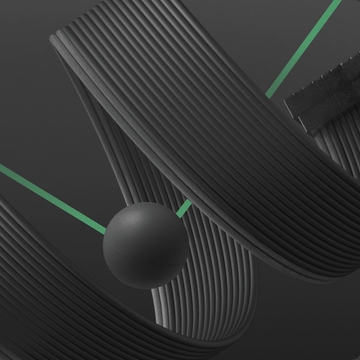
Сам по себе искусственный интеллект не способен что-то оценить или предсказать. Чтобы модель поняла, что вы любите триллеры, или рассчитала количество удобрений на гектар почвы, её нужно обучить работать с данными.

Этим занимается специалист по машинному обучению — его ещё называют «Специалист по Data Science». В Яндекс Практикуме есть целый курс по Data Science и машинному обучению: там учат анализировать большие объёмы данных, применять ML и за счёт этого инструмента создавать и улучшать продукты для бизнеса, промышленности и науки.

Решайте задачи с помощью машинного обучения

Попробуйте себя в роли специалиста по Data Science: находите неочевидные закономерности в данных, стройте гипотезы, обучайте алгоритмы. Начните курс с бесплатной вводной части.

[Попробовать](https://practicum.yandex.ru/data-scientist/?from=blog-practicum" \t "_blank)



Основы машинного обучения выглядят так. Модель похожа на чёрный ящик: она принимает на вход данные — условие задачи, — а на выходе выдаёт некий ответ.

Но у ящика есть много дополнительных параметров. Они влияют на то, как именно будут обрабатываться данные. Способы машинного обучения и типы задач влияют на то, как именно будут настраиваться эти параметры.

Мы сами, когда учимся или выполняем какие-то задания, действуем похожим образом. Например, самый простой способ выучить стихотворение — повторять его, пока не запомнишь. Чтобы решить математическую задачу, нужно понять условие и разложить его на несколько действий. Написать пейзаж проще, если найти образец и срисовать с фотографии.

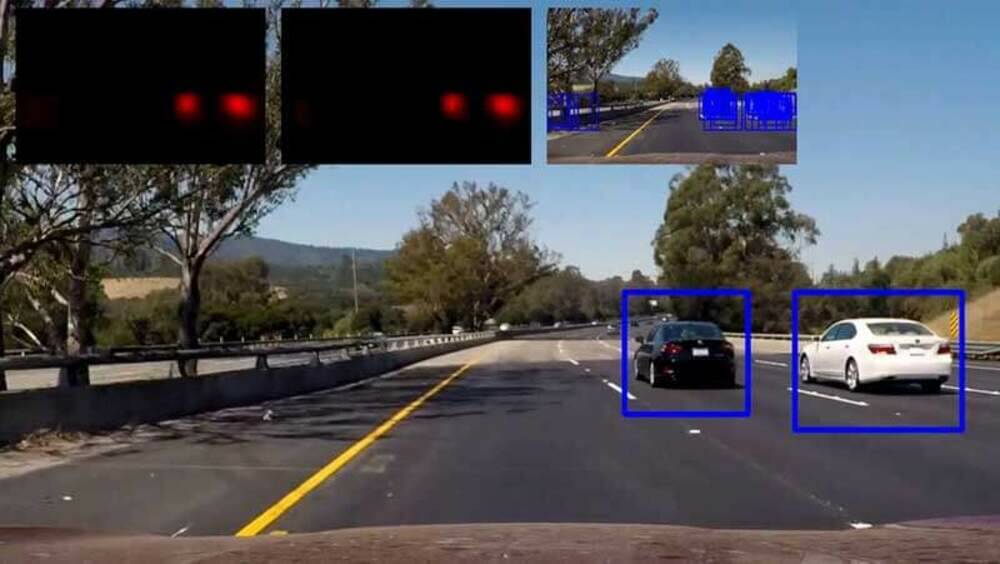
То, какой подход к обучению в итоге выбран, похоже на подбор, который настраивается в зависимости от типов машинного обучения или типов задач.

Какие способы машинного обучения бывают.

**Обучение с учителем**

Специалист использует для обучения модели размеченные данные. Размеченные — значит, что на каждый пример уже есть правильный ответ. Задача алгоритма — понять, почему он правильный.

Например, нужно научить алгоритм распознавать на улицах легковые автомобили. Для обучения будут использоваться много фотографий, на которых все области, включающие машины, выделены прямоугольниками. Модели нужно самой найти взаимосвязи между областями в прямоугольниках и установить, что в них находится один и тот же объект — автомобиль.

\*

Детекция автомобилей с помощью компьютерного зрения и машинного обучения используется при конструировании автопилотов. [Источник](https://towardsdatascience.com/teaching-cars-to-see-vehicle-detection-using-machine-learning-and-computer-vision-54628888079a)— TowardsDataScience

**Обучение без учителя**

Для обучения используются неразмеченные данные — то есть такие, по которым нет никакой оценки. Машине нужно самой определить закономерности. Например, проанализировать всех пользователей интернет-магазина, классифицировать их по поведению и выявить среди них тех, кто чаще всего покупают детские товары.

**Обучение с подкреплением**

В этом случае задача — подобрать подходящий ответ в зависимости от действий пользователя. Самый популярный пример — боты для игры в шашки или шахматы, когда алгоритм пытается принять решение в зависимости от действий другого игрока.

Как работает обучение с подкреплением, видно на примере Каспарова и Deep Blue. В первый раз Каспаров сделал привычный тактический ход — пожертвовал фигуру. Алгоритм принял жертву за чистую монету. Но во второй раз в сходной ситуации машина действовала по-другому, так как сумела проанализировать действия противника и последствия.

Задачи машинного обучения.

* **Обучение с учителем — используется для классификации и регрессии**.

Классификация — задача, при которой нужно присвоить объекту класс: кошка, собака, машина, дома.

Регрессия — задача, в которой необходимо присвоить числовое значение. Например, просчитать ожидаемый доход от инвестиций через 6 месяцев.

* **Обучение без учителя — подходит для классификации**.

Кластеризация — это группировка на основе набора признаков. Например, если покупатель каждый месяц покупает игрушки, пелёнки, детское мыло и подгузники — у него есть дети. А это значит, что он будет покупать примерно те же товары, что и другие родители.

* **Обучение с подкреплением — оптимально для создания ботов**.

Обучение с подкреплением помогает [чат-ботам](https://practicum.yandex.ru/blog/chto-takoe-chat-bot/) подобрать правильный ответ на вопрос, а роботу-манипулятору открывать различные ручки дверей.

Как работает ML‑специалист.

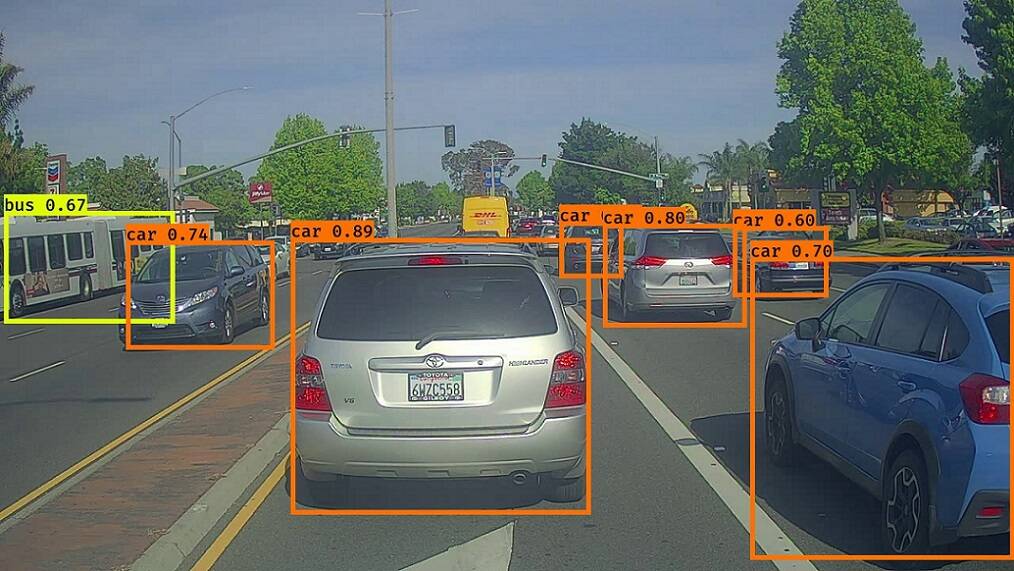
Работа с обучением модели делится на пять этапов:

**1. Сбор данных**

На этом этапе нужно собрать информацию, которая будет использоваться для обучения модели.  
Но собрать — не значит просто найти много подходящих цифр или картинок.  
Специалист должен определить, какие именно данные нужны, какие критерии будут критичными, а какие нет. Именно поэтому специалисты по ML с самого первого этапа плотно взаимодействуют с заказчиком модели. Например, если это банк, то эксперт может участвовать в составлении анкеты для клиента.

**2. Разметка данных**

Если обучение модели предполагает работу с размеченными данными, нужно провести подготовительную работу — выделить области или критерии, которые нужны для обучения машины или дать правильный ответ для каждого случая.



Существует несколько типов разметки изображений: ограничительные рамки, как на этой иллюстрации, или трёхмерные кубы, помогающие ИИ определить размер объекта. [Источник](https://medium.com/nerd-for-tech/what-are-the-training-data-sets-used-to-train-the-ai-model-for-self-driving-cars-540a45bf64dc)— Medium

Это трудоёмкий и рутинный процесс, и далеко не всегда все данные без исключения размечаются ML-специалистами вручную. Иногда к разметке привлекают сторонних подрядчиков или поручают разметку специально обученному алгоритму. Но полностью автоматизировать этот процесс невозможно: участие эксперта нужно для того, чтобы проверить, нет ли ошибок в разметке.

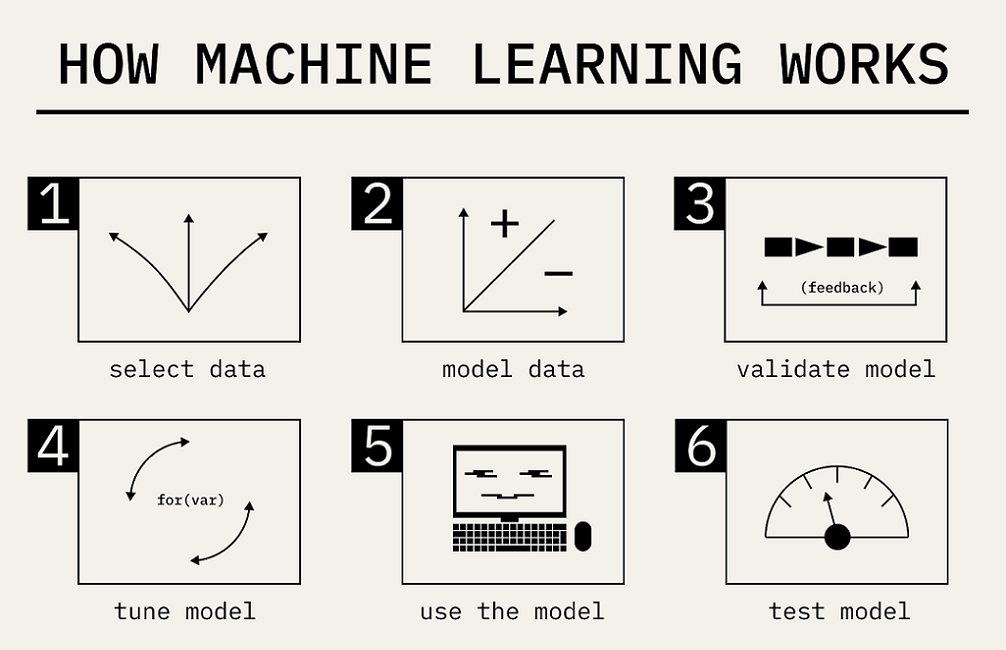
А ещё есть специфические данные, для разметки которых нужно глубокое знание отрасли. Например, если размечают данные для медицинской модели, которая учится распознавать снимки и определять на них опухоли. Правильно обучить такую модель можно, только если сами эксперты уверены, что на каждом выделенном фрагменте действительно новообразование, а не просто засвеченный участок.

**3. Исследовательский анализ**

Этап контрольной проверки, при котором специалист сверяет, как распределены данные, как зависят друг от друга различные признаки и нет ли в них ошибки или нетипичных случаев. Например, для некоторых классических алгоритмов нужно, чтобы все цифры находились в диапазоне от 0 до 1 или от −1 до 1. Задача ML-специалиста на этом этапе — посмотреть вглубь данных и увидеть основные закономерности.

**4. Обучение модели**

На этом этапе ML-специалист выбирает подходящие алгоритмы для решения задачи, обучает несколько перспективных моделей.



**5. Тестирование и оценка**

Итоги обучения нужно оценить и понять, что делать дальше: собрать недостающие данные и продолжить обучение, заменить параметры модели или пересмотреть алгоритм.

Как в любой профессии, в ML бывают и нестандартные ситуации, не вписывающиеся в эту схему работы. Например, когда для данной задачи нет подходящих алгоритмов и нужно спроектировать новый. Или создать новую архитектуру [нейросети](https://practicum.yandex.ru/blog/chto-takoe-neyronnye-seti/), обучить её и оценить результат. Машинное обучение — быстро развивающееся направление. Вычислительные мощности растут, появляются новые задачи, требующие нетипичного подхода. А значит, у ML-специалиста всегда есть пространство для творчества и профессионального развития.