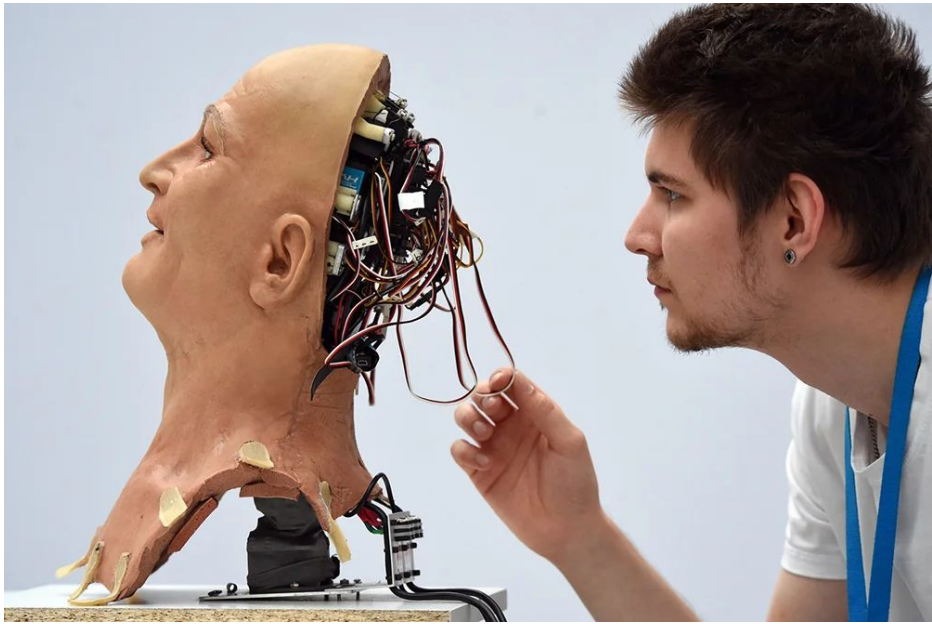
Может ли искусственный интеллект навредить человечеству



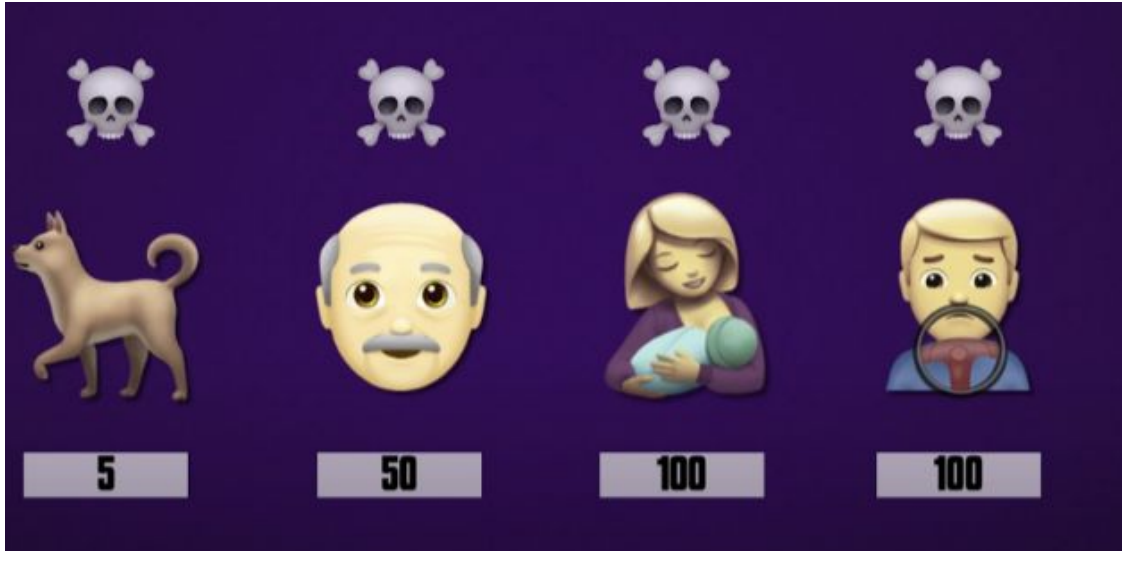
19 марта 2018 года 49 летнюю женщину насмерть сбил беспилотный автомобиль

Это первая авария со смертельным исходом при участии беспилотного автомобиля.

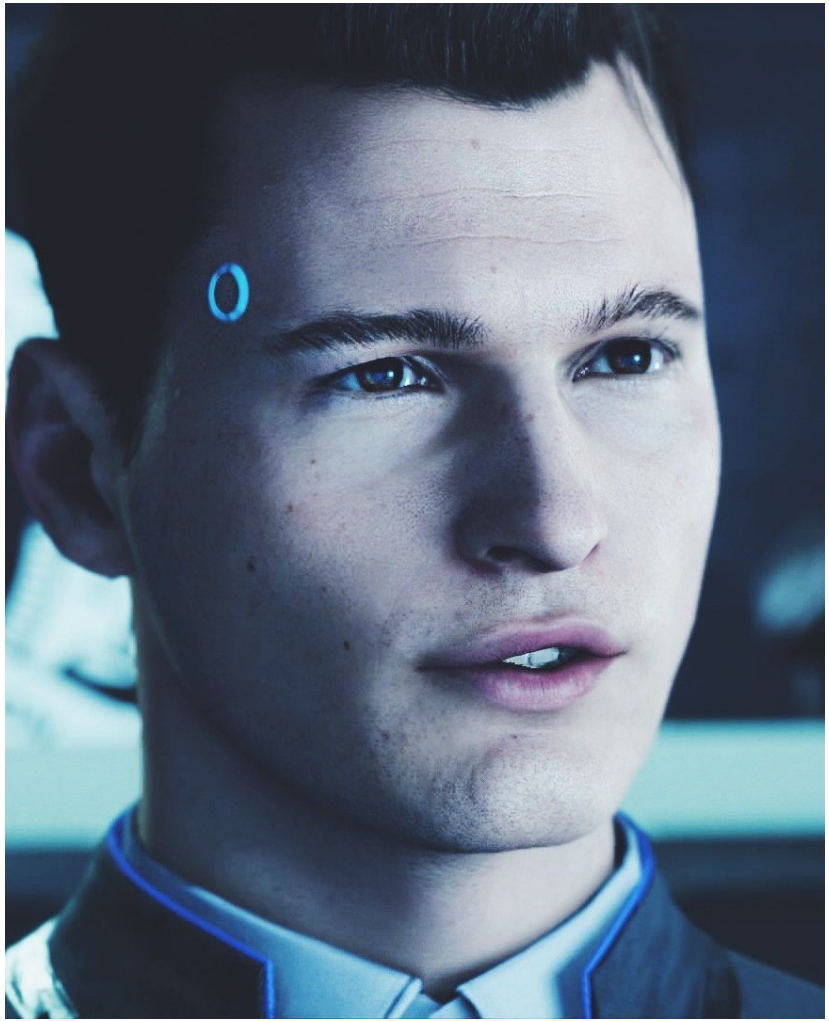
В этой аварии нет виноватого, машина тестировалась, и водитель был внутри как раз на случай таких ситуаций, но она отвлекалась, смотрела шоу голос. Велосипедистка выпрыгнул на дорогу из темноты в неположенном месте, к сожалению никто бы не мог предотвратить этой смерти.
Этот трагический и грустный случай был первым, поэтому он вызвал такой резонанс.
Но действительно ли это восстание машин?

Может ли искусственный интеллект убивать специально и представляют ли опасность беспилотные автомобили ?

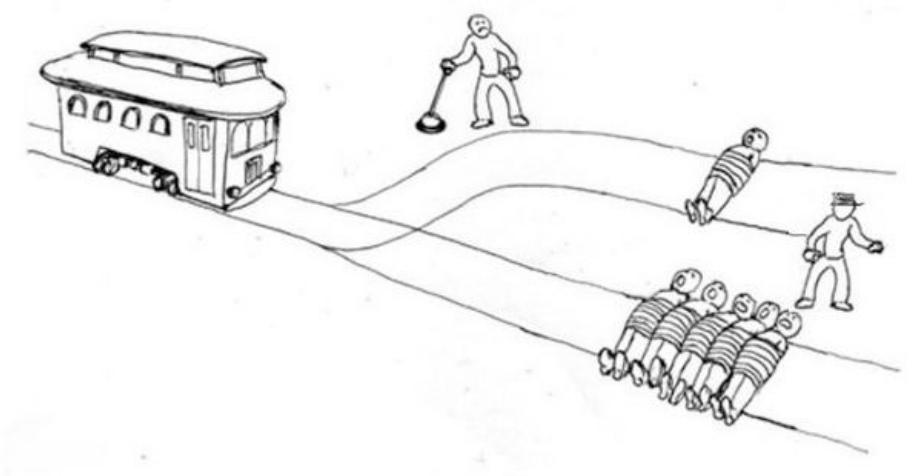
Автомобиль не отвлекается не засыпает и старается всеми силами избежать аварии, но если это невозможно, то постарается минимизировать ущерб.
Для этого скорее всего ему пропишут систему штрафов. *Например сбить собаку он это штраф 5 баллов, старика 50, мать с младенцем 100, пожертвовать водителем тоже 100 баллов.* *Задача машины набрать как можно меньше баллов.*



Если мы хотим открытого и дружественного интеллекта от машин, то нам сначала нужно разобраться со своим отношением к себе, а уже потом мучить машину. Что такое хорошо, а что плохо? Но как машины воспринимают наш мир сейчас, и стоит ли их бояться.



Мы ,люди, прописываем машине правило, как поступить в той или иной ситуации. Не машине придется выбирать кого спасти, а кем пожертвовать, выбирать придется человеку, который задает правила.
Все это похоже на классическую проблему вагонетки.
Ситуация:
несколько человек привязаны к рейсам и на них едет вагонетка с одним человеком внутри, необходимо сделать выбор, убить одного человека, чтобы спасти нескольких или не делать ничего и дать вагонетки переехать пятерых.
И правильного ответа здесь нет. Можно действовать по принципу наименьшего зла или наоборот не вмешиваться но так или иначе объяснить.



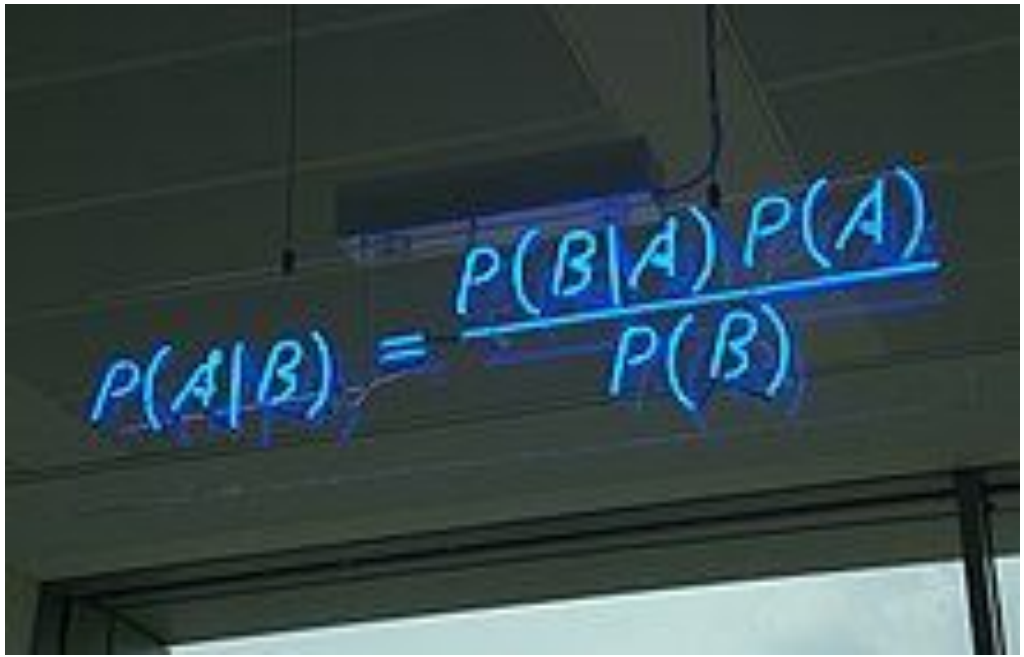
Мы приходим к выводу, что так или иначе нам придется рассказать искусственному интеллекту, что хорошо, а что плохо,но для этого нужно знать, как он работает.

Теорема Байеса

Одна из основных теорем элементарной теории вероятностей, которая позволяет определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие.

Томас Байес придумал эту формулу ещё в 1763 году. Можно сказать, что основы искусственного интеллекта были заложены еще в 17 веке, но первый нейрокомпьютер появился только спустя 200 лет.

Теорема Байеса



$P(A)$ — априорная вероятность гипотезы A (смысл такой терминологии см. ниже);

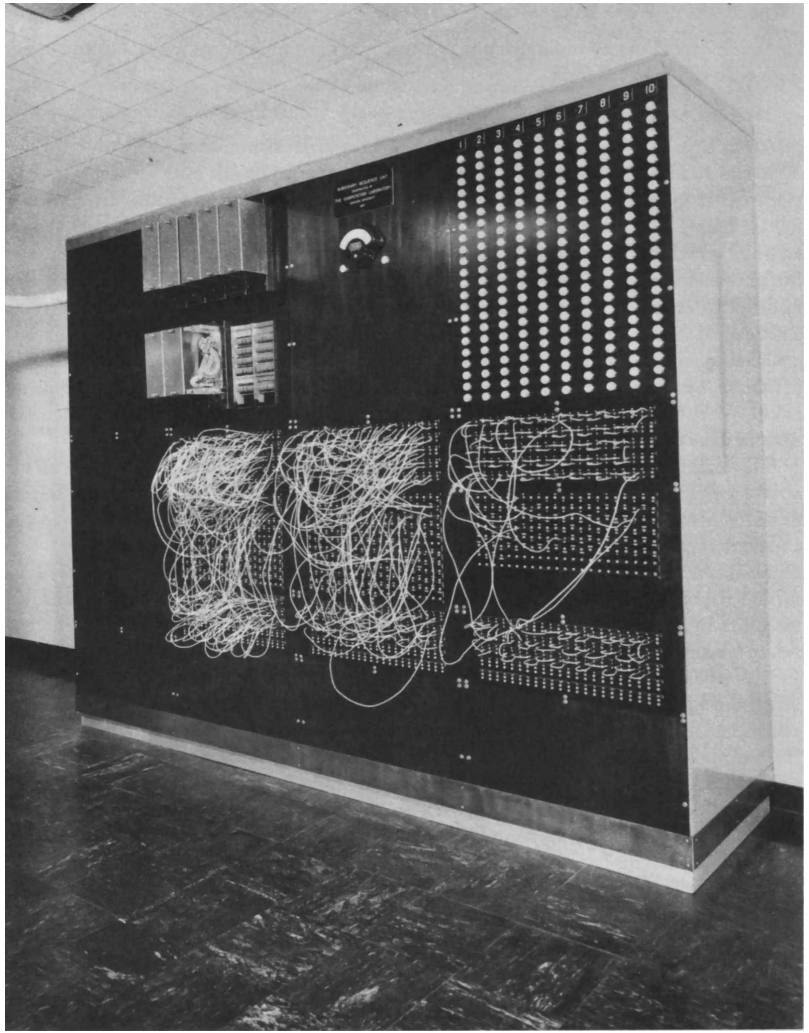
$P(A \mid B)$ — вероятность гипотезы A при наступлении события B (апостериорная вероятность);

$P(B \mid A)$ — вероятность наступления события B при истинности гипотезы A;

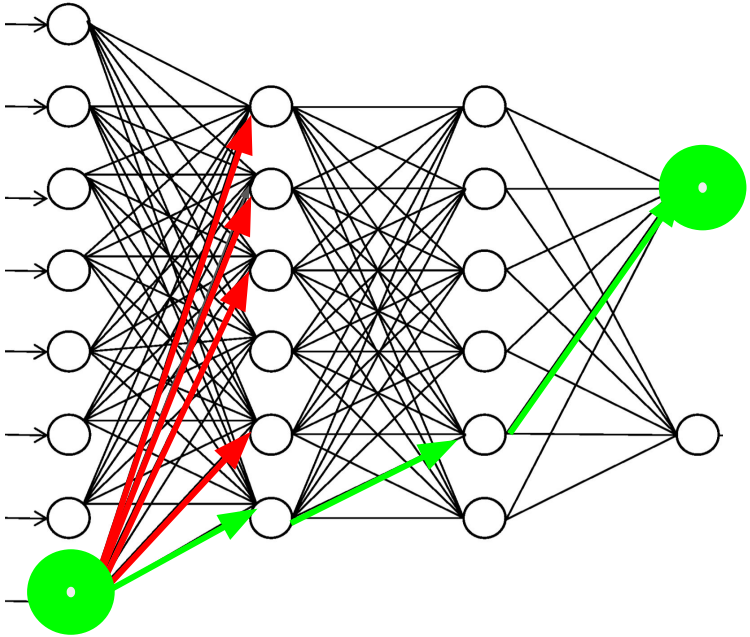
$P(B)$ — полная вероятность наступления события B.

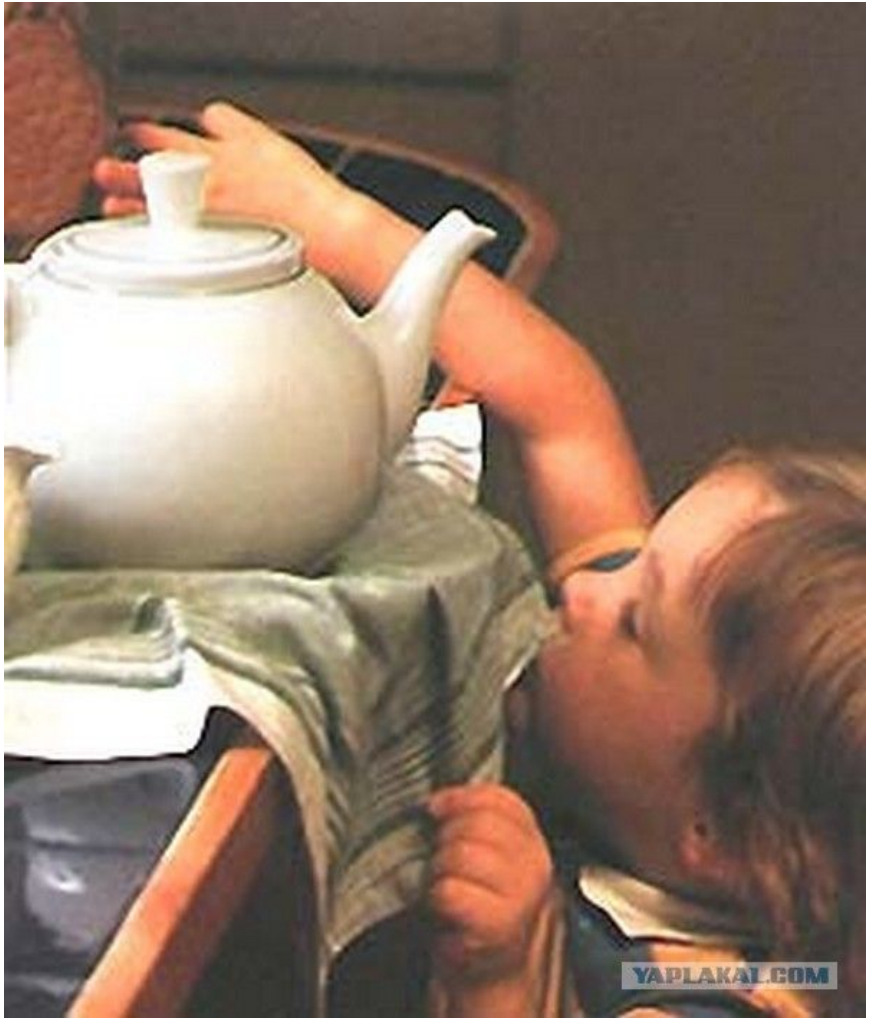
В 1958 году, это ***Марк 1***, работает на алгоритме персептрон, Его “мозг” состоит из слоя рецепторов, нейронов, и классификаторов.

1)Рецепторы принимают сигнал, как это делает сетчатка нашего глаза и передает дальше нейрону.
2)Каждый рецептор связан с каждым нейроном, отсюда такое количество проводов.
3)Нейроны суммируют все сигналы и передают сигнал в классификаторы, они то и
распознают изображение, каждому образу, например цифре, букве или квадрату соответствует свой классификатор.



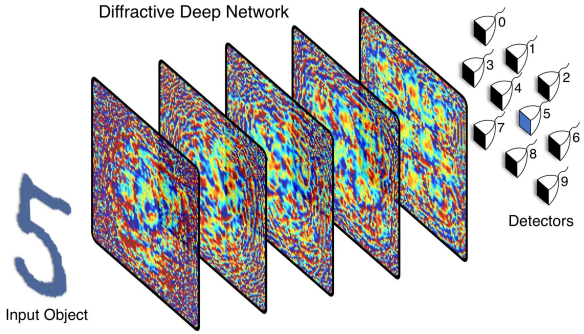
Сила связей между рецепторами и нейронами в процессе обучения меняется, связи приводящие к правильному распознаванию усиливаются, к неправильному наоборот ослабляются.
Как ребенок понимает со временем, что чайник горячий, так и марк 1 со временем учится распознавать цифры буквы и простые геометрические фигуры.

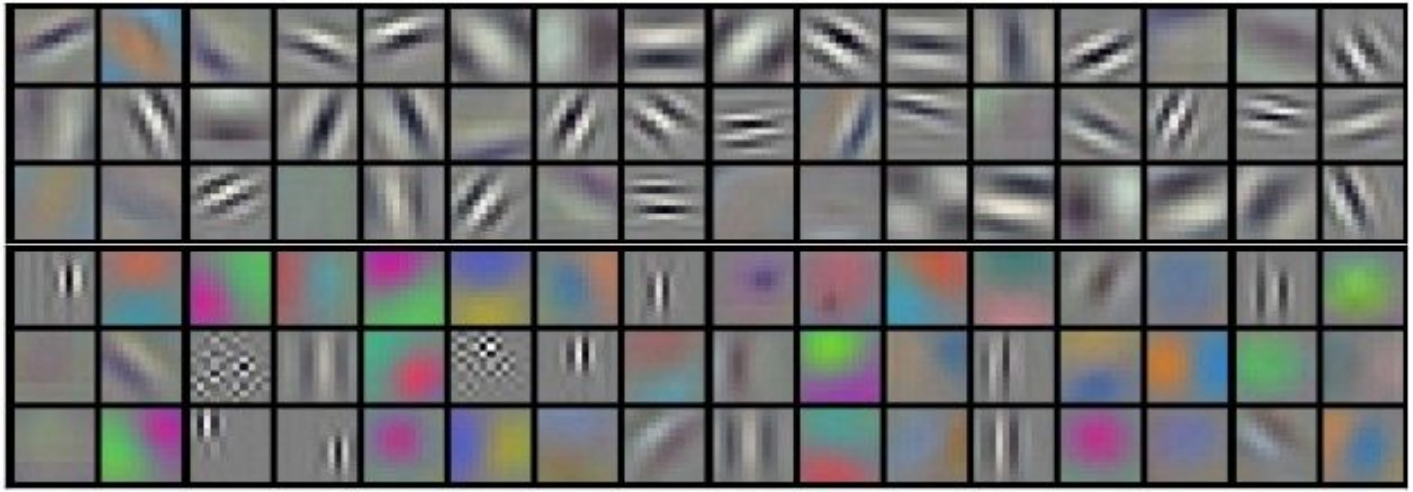




Нейросети

Сейчас нейросеть состоит из множества слоёв:
первый слой как младенец , видит только свет и цвет, расплывчатые силуэты
второй слой различает элементы и текстуры
третий слой механизмы части тела людей и животных
четвертый слой классифицирует объекты, понимает, что ему показывают
Нейросеть это обычная программа, но главное отличие в том что она учится решать задачи сама





Живые организмы быстро приспосабливаются, при этом иногда полностью меняя привычный для себя алгоритм действий. Эти действия могут происходить неосознанно, пчела, которая миллион раз собирала нектар с цветов, полетит в сторону цветочного луга, чуя знакомый запах. При этом влияние сторонних факторов не отражается на жизненном цикле пчелы.

И тут мы понимаем, что ум обычной пчелы развит больше, чем ум самого совершенного компьютера на данный момент.