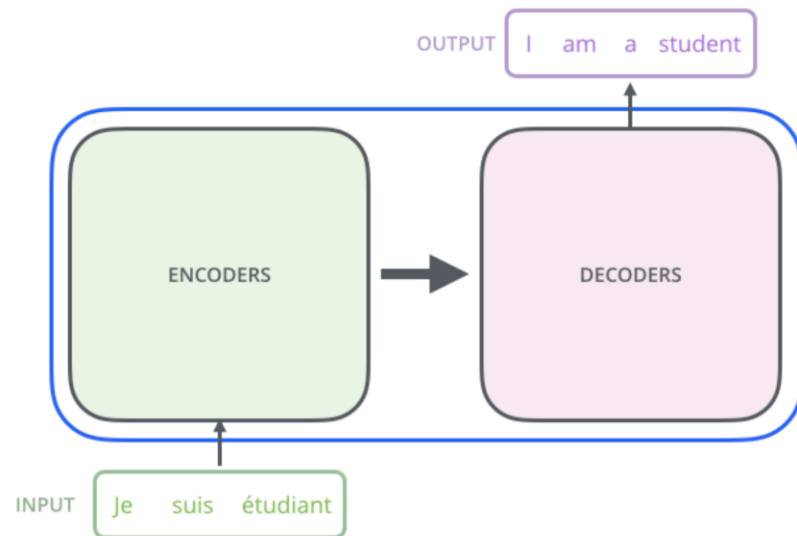


Gpt –3.5

(Generative Pretrained Transformer)

Нейросеть на базе архитектуры трансформер

Составляющие трансформера



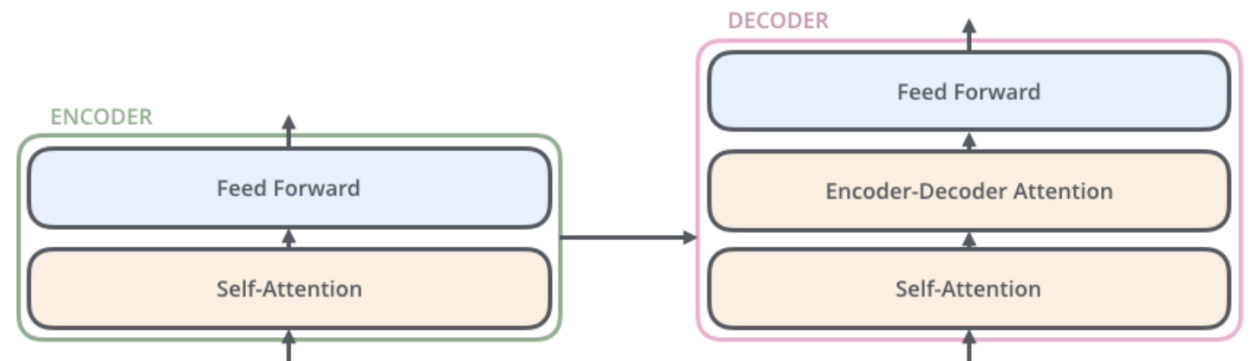
Кодирующий компонент – это стек энкодеров

Декодирующий компонент – это стек декодеров

Все энкодеры идентичны по структуре, хотя и имеют разные веса. Каждый можно разделить на два подслоя:

- слой внутреннего внимания (self-attention)
- нейронная сеть прямого распространения (feed-forward neural network)

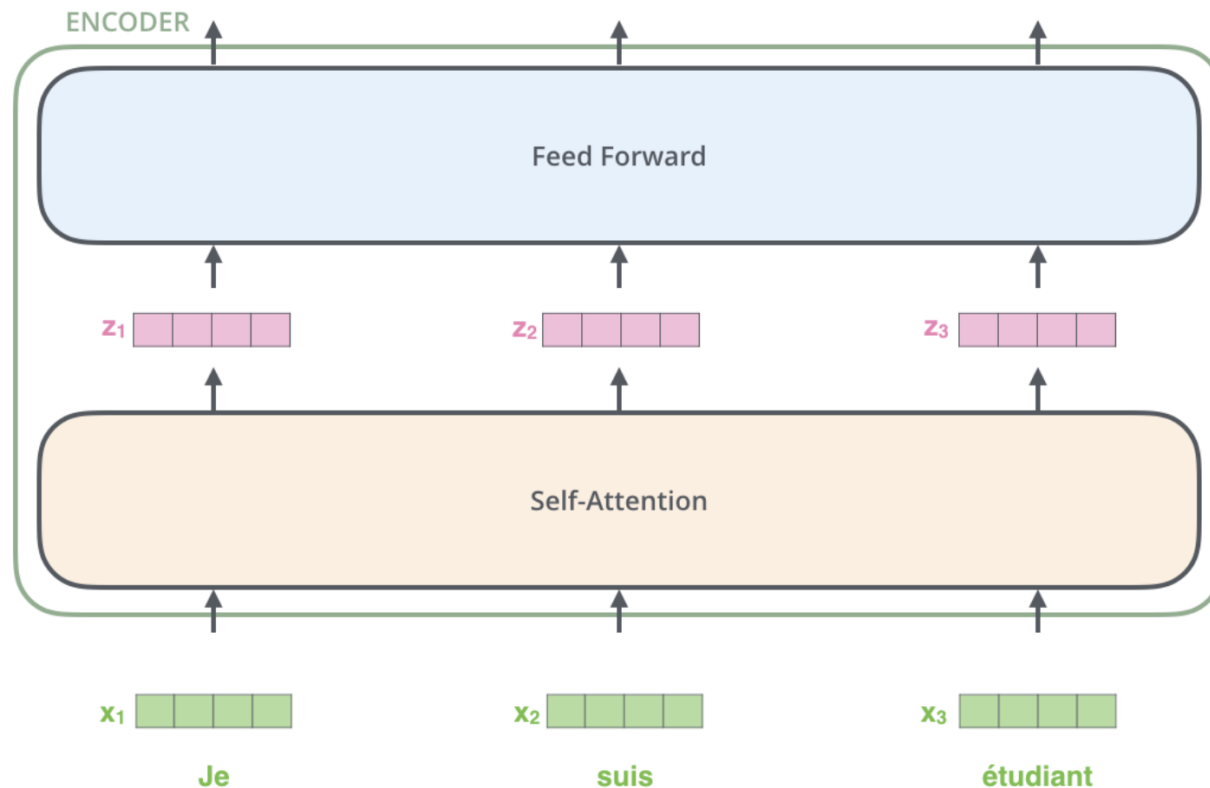
Декодер также содержит два этих слоя, но между ними есть слой внимания



Энкодер

После того как слова входящего предложения преобразовались в эмбединги, каждый из них в отдельности проходит через два слоя энкодера.

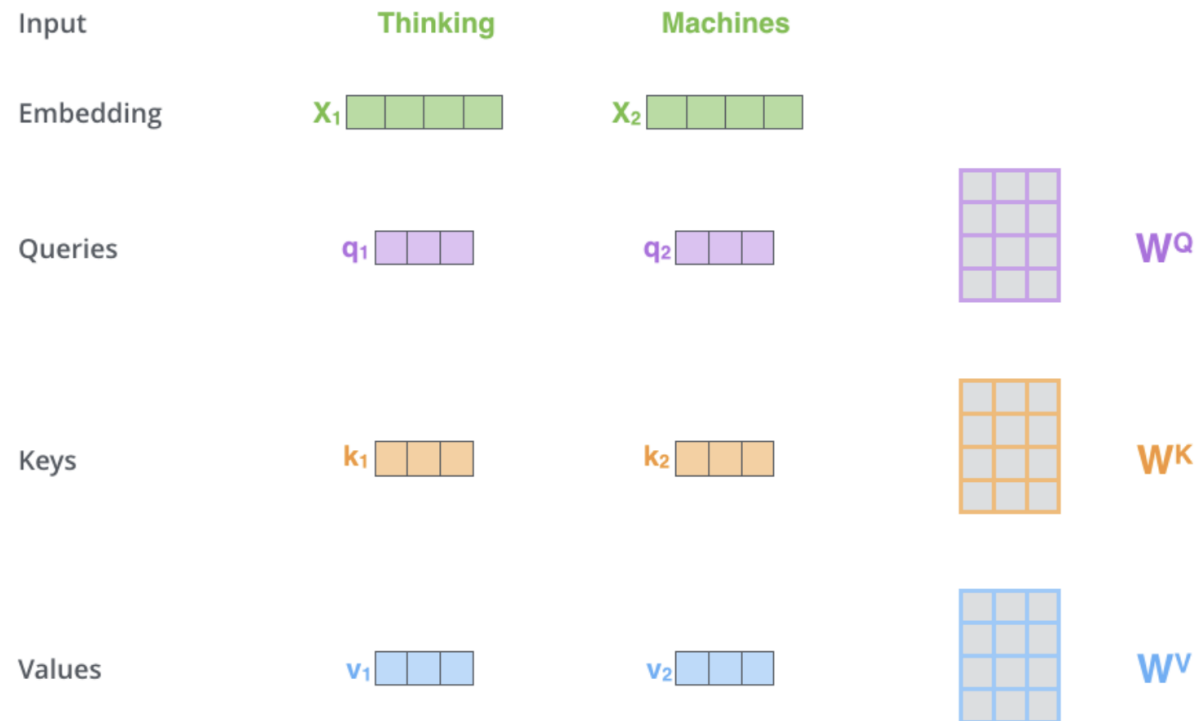
По мере того как модель обрабатывает каждое слово, внутреннее внимание позволяет модели взглянуть на другие позиции входной последовательности и найти подсказку, помогающую лучше закодировать данное слово.



Механизм внутреннего внимания

Первый этап – создание трех векторов из каждого входящего вектора : вектор запроса (Query vector), вектор ключа (Key vector) и вектор значения (Value vector). Эти векторы создаются с помощью перемножения эмбединга на три матрицы, которые мы обучили во время процесса обучения.

Второй этап – получение коэффициента. Допустим, мы подсчитываем внутреннее внимание для первого слова в нашем примере – «Thinking». Нам нужно оценить каждое слово во входящем предложении по отношению к данному слову. Коэффициент определяет, насколько нужно сфокусироваться на других частях входящего предложения во время кодирования слова в конкретной позиции.

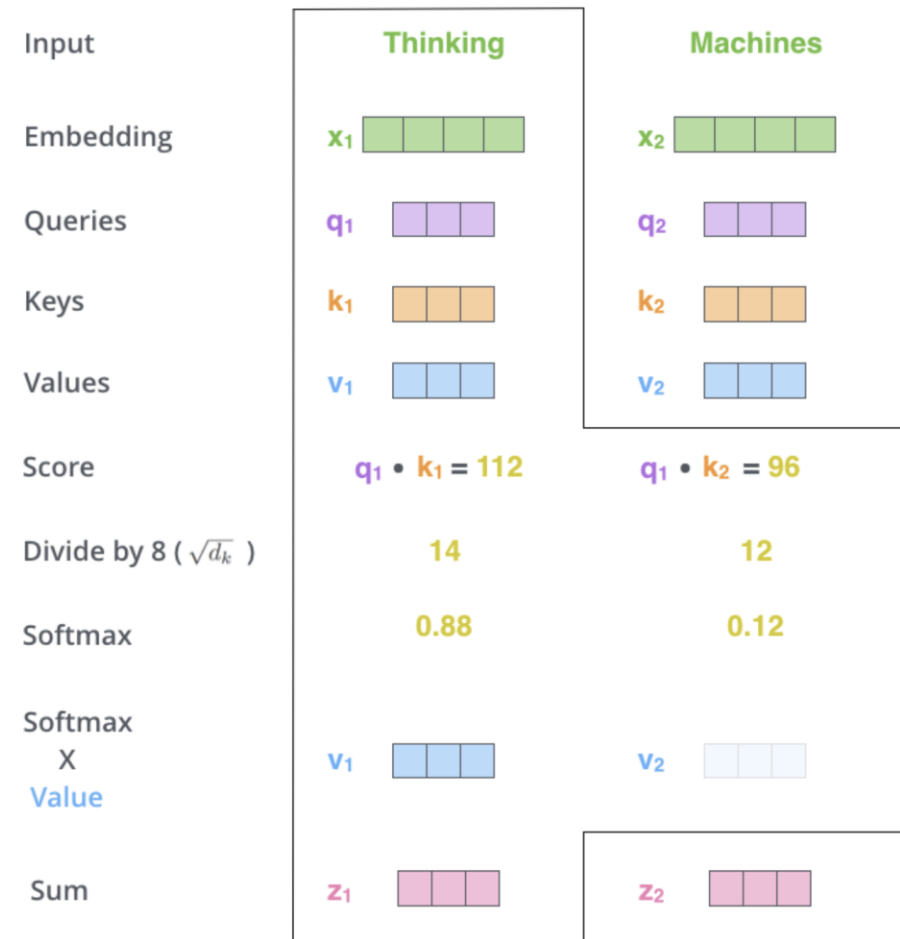


Механизм внутреннего внимания

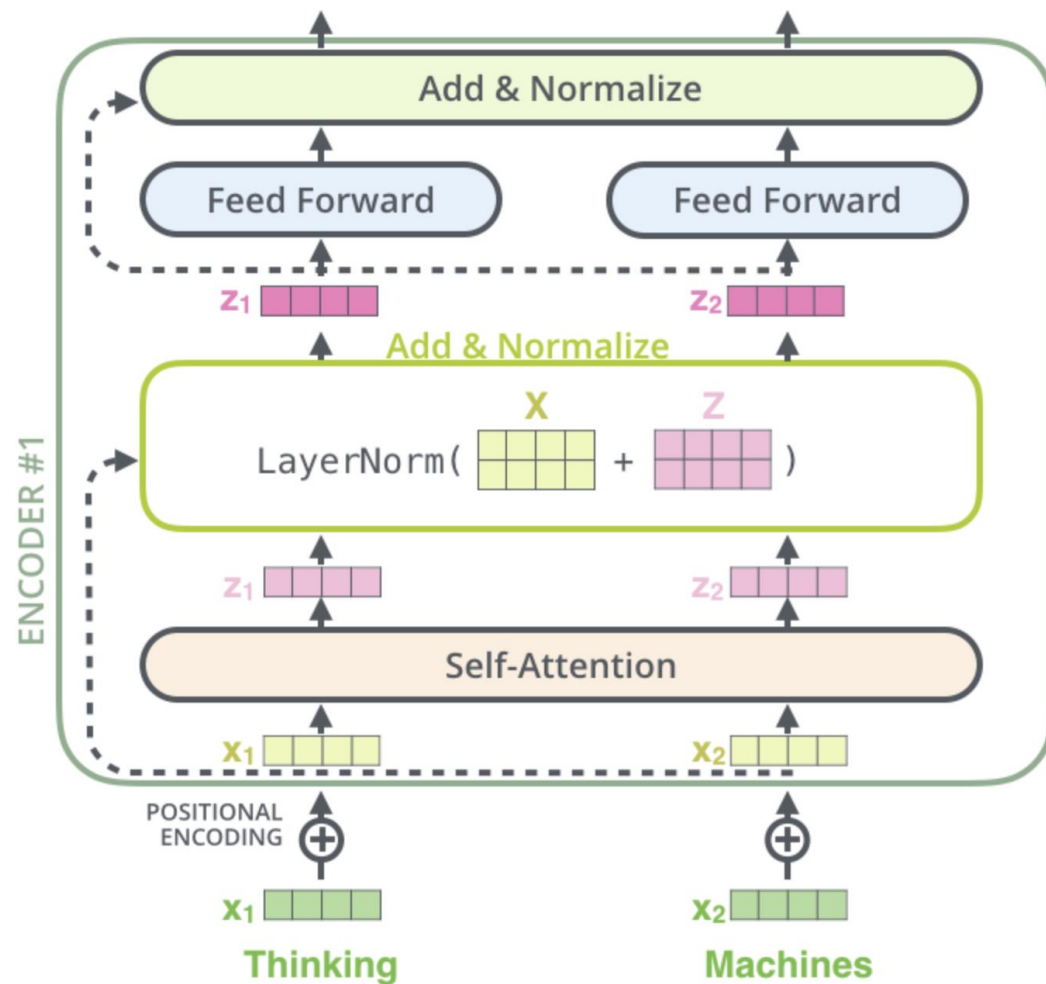
Третий и четвертый этапы – разделить эти коэффициенты на 8 , а затем пропустить результат через функцию softmax. Данная функция нормализует коэффициенты так, чтобы они были положительными и в сумме давали 1.

Пятый этап – умножить каждый вектор значения на softmax-коэффициент.

Шестой этап – сложить взвешенные векторы значения. Это и будет представлять собой выход слоя внутреннего внимания в данной позиции (для первого слова).



Этап нормализация



Функция потерь

Поскольку параметры модели (веса) инициализированы случайным образом, необученная модель генерирует распределение вероятностей с произвольными значениями для каждой ячейки/слова. Мы можем сравнить его с реальным выходом, затем изменить все веса, используя метод обратного распространения ошибки для того, чтобы приблизить выход к желаемому.