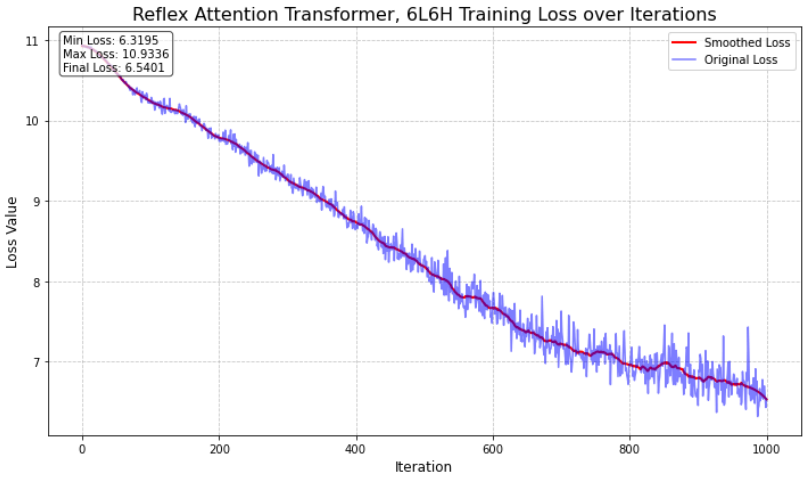
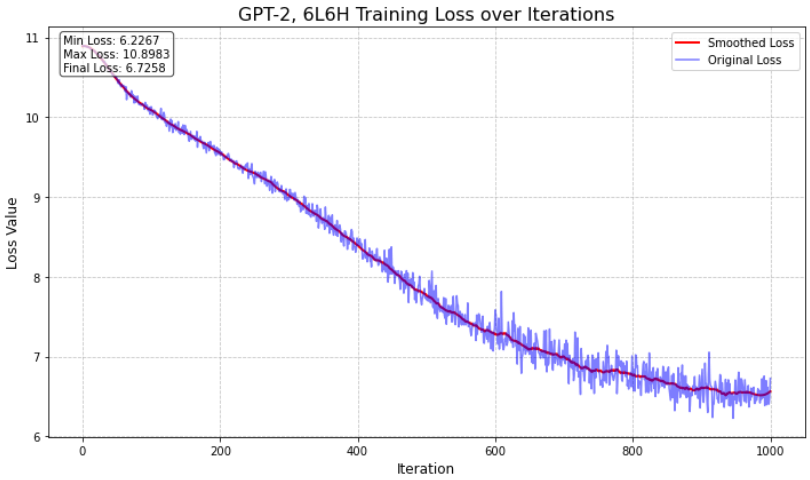
Отчет об экспериментах

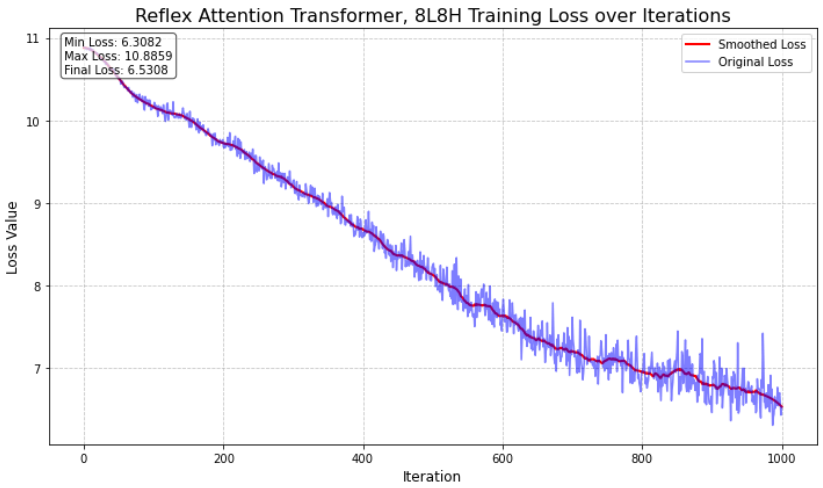
- (Обучал в kaggle, для использования нужно будет поменять пути к файлам и скачать скрипты отдельно)
- Что получилось
    - Создал nn.Module на базе Reflex attention
    - Обучил пару моделей: Reflex Transformer 8L8H, Reflex Transformer 6L6H, nanoGPT 6L.



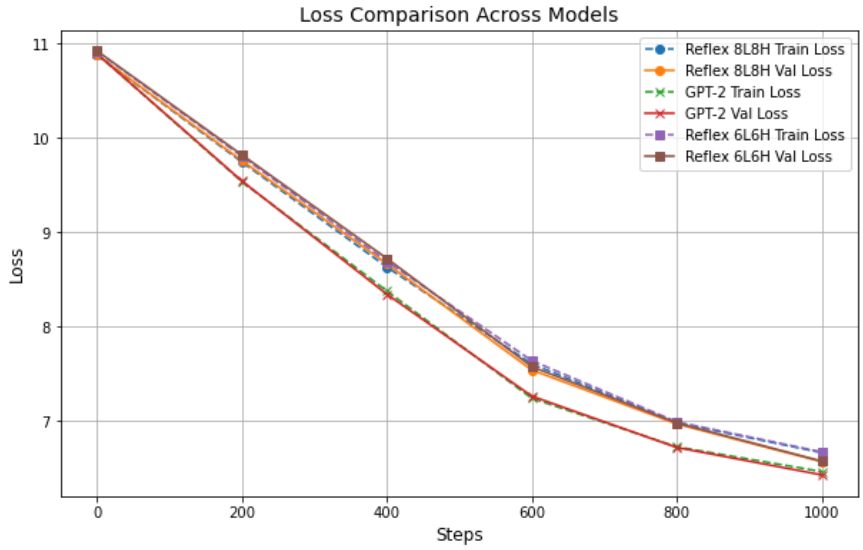
-



-



-



-

    - Выбор гиперпараметров был на основе nanogpt, возможно поэтому другие модели набрали меньше.
- Что не получилось
    - Я не знаю как лучше всего проверять качество моделей на бенчмарках для таких маленьких моделей, из статей нашел пару бенчмарков типо GLUE, но я не был до конца уверен насколько они нужны так как основная мотивация для архитектуры изучение длинных контекстов. Можно было использовать датасеты которые используются для исследования маленьких моделей (IOI, sorting), но 6 слойный трансформер и так должен их решать очень хорошо.
    - Для kaggle создать модель больших слоев скорее всего было бы тяжело, поэтому я не экспериментировал с линейной комбинацией и выбором top_k.
        - Я бы использовал дополнительной слой self.linear_combination = nn.Linear(k, 1, bias=False) и top_k или отдельно для нужных слоев линейная комбинация из всех предыдущих.
- Я оценил модель на задаче сортировки длинной последовательности и ARC easy (fine-tuning на трейн части и оценка качества на тесте). Если бы было бы больше

времени я бы использовал LAMBADA, где длины последовательностей больше, и датасеты где нужно отвечать на вопросы по контексту.

    - На ARC easy модель Reflex attention 6L6H набрала accuracy 25, то есть эквивалентно выбору случайного ответа, возможно модель научилась именно этому. Я использовал max_len = 5, что могло повлиять на результаты.
    - Так как модель (6L6H) очень маленькая, стоило ожидать такого результата для таких оценок.
    - Модель (6L6H) показала неплохие результаты на сортировке длинных последовательностей из 100 чисел, но все равно встречались проблемы с вытягиванием всех повторений разных чисел.

- AGI has not been achieved.