

Тестирование проводилось при $max_length = 100$

Текст: “Он на первый взгляд, самый обычный среднестатистический американец”

Наблюдения:

1. При температуре = 0.001 (top_k - любое) - много повторов
2. При $top_k = 1$ - повторы текста
3. При темпер. < 0.3 и значении $top_k = 100$ - повторы при большом тексте

В маленьком все ок:

Темп. = 0.1 + $top_k = 100$

«Он на первый взгляд, самый обычный среднестатистический американец (это было видно даже по внешности). А как он подошел к нам! Он был очень собран, говорил на хорошем английском. Посмотрев на него внимательно, я и не предполагал, что это – посол США.»

4. При увеличении температуры и top_k - текст становится интереснее, но не всегда смысловой
5. Чем больше температура тем меньше нужно брать top_k

Темп. = 1.0 + $top_k = 10$

«Он на первый взгляд, самый обычный среднестатистический американец, который родился где – то в Техасе и живет в маленьком провинциальном городке. Но при более тщательном рассмотрении он оказывается настоящим «неотесанным мужланом», с очень непростым характером. Он может быть очень добрым и мягким, а может быть и жестоким и грубым. В конце 80х годов прошлого века один из главных героев фильма «Остров проклятых», который по сюжету является одним из главных героев фильма»

Темп. = 2.0 + $top_k = 3$

«Он на первый взгляд, самый обычный среднестатистический американец. Но это лишь на первый. А на второй - это не американец, и не просто среднестатистический, а самый настоящий американец-интеллектуал. Он не только читает книги, но еще и пишет статьи.

И, что интересно, не в интернете. Он пишет статьи в солидных изданиях. Например в журнале "Ньюс".

И вот, что интересно - в этом журнале он тоже не просто пишет статьи...»

$$\text{Темп.} = 0.7 + \text{top_k} = 10$$

«Он на первый взгляд, самый обычный среднестатистический американец. Но если присмотреться, то можно увидеть в нем что-то особенное. А именно - он очень умный, умен, образованный и интеллигентный мужчина.

В этом видео он рассказывает о том, почему он хочет жить в России»

6. $\text{top_k} = 100$ - тест не связан по смыслу (как и в случае большая температура + большой top_k)

$$\text{Пример плохого текста: Темп.} = 5.0 + \text{top_k} = 30$$

«Он на первый взгляд, самый обычный среднестатистический американец средних сиквидей: он живёт в обычной, благополучнейшей обстановке со средневозрастная школа или сэндвич. Но за этим, внешним лоск, скрыт мощный боевой потенциал Америки тех дней». (Лурдес Оганесиян. История Второй Ингер, стр. 361.) По мнению ряда историков: 1 ноября 1941. Асаф Галемчик-Джелайли, бывший премьерминистр Ливана — генерал»

7. Чем > температура, тем < нужно брать top_k

$$\text{Темп.} = 0.7 + \text{top_k} = 10$$

«Он на первый взгляд, самый обычный среднестатистический американец, но он умеет делать все, что нужно, чтобы стать великим. И это не просто слова, а реальные факты. В этом фильме вы найдете много информации о том, как стать великим актером и как стать великим режиссером.

В основе сюжета фильма лежит история о двух знаменитых людях в Голливуде.

В главных ролях: Джордж Клуни, Сьюзан Сарандон...»

8. При темпер. = 0.5 лучше выбирать $\text{top_k} > 40$

$$\text{Темп.} = 0.5 + \text{top_k} = 50$$

«Он на первый взгляд, самый обычный среднестатистический американец. Но это только на первый взгляд. На самом деле, он очень близок к восточным религиям. Он верит в бога, как и большинство жителей США.

По словам политика, в США очень много «неправильных» религий. В США можно встретить приверженцев различных культов.

Многие люди, живущие в США, верят в бога, но они не знают, что он существует», — сказал политик.»