

Фамилия, имя, номер группы:

.....

Внесите сюда ответы на тест:

Вопрос	1	2	3	4	5	6	7	8	9	10
Ответ										

Горам все равно.

*Надпись на входе в национальный парк "Скалистые горы"*

Работа состоит из трёх частей: тестовая, задачи и ответы на открытые вопросы, ML-design. Списывание карается обнулением работы. Ко всем одинаковым формулировкам я буду очень сильно придираться. Удачи!

## Часть первая: тестовая

Дайте ответ на 11 тестовых вопросов. Каждый вопрос стоит 3 балла. Последний вопрос стоит 2 балла. Никакие дополнительные пояснений в этой части работы от вас не требуются.

**Вопрос 1.** При обучении рекуррентных нейронных сетей возникает проблема взрыва градиентов. Какие из способов ниже помогают решить эту проблему?

- ☐ A Использовать продвинутые виды ячеек (LSTM, GRU)
 ☐ C Обрезать градиенты (Gradient clipping)
 ☐ E Использовать ReLU в качестве функции активации
- ☐ B Использовать Dropout
 ☐ D Инициализировать веса ортогональными матрицами
 ☐ F Нет верного ответа.

**Вопрос 2.** Какая из функций потерь является функцией потерь генератора при обучении GAN?

- ☐ A  $\frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z_i)))$ 
☐ C  $\frac{1}{m} \sum_{i=1}^m \log(1 - G(D(z_i)))$ 
☐ E  $\frac{1}{m} \sum_{i=1}^m \log G(z_i)$
- ☐ B  $-\frac{1}{m} \sum_{i=1}^m \log D(G(z_i))$ 
☐ D  $-\frac{1}{m} \sum_{i=1}^m \log G(D(z_i))$ 
☐ F Нет верного ответа.

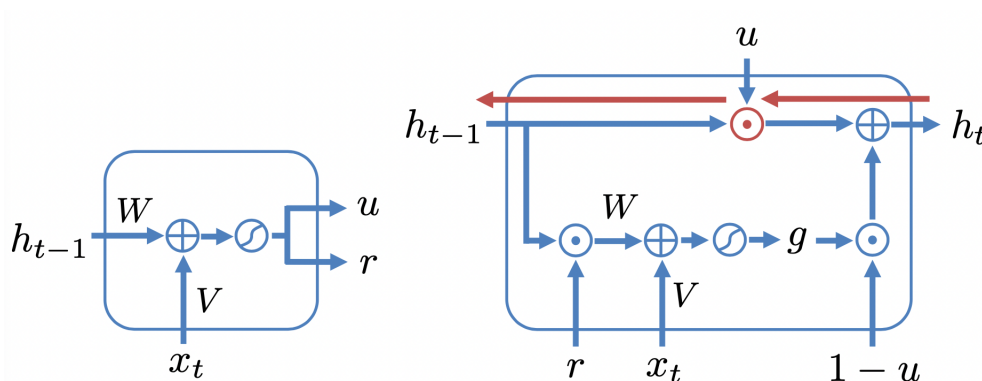
**Вопрос 3.** Какие из моделей обычно используют для классификации текстов?

- ☐ A Логистическая регрессия поверх tf-idf
 ☐ C Случайный лес поверх tf-idf
 ☐ E Свёрточные нейронные сети
- ☐ B Рекуррентные нейронные сети
 ☐ D Градиентный бустинг поверх tf-idf
 ☐ F Нет верного ответа.

**Вопрос 4.** Какие из архитектур, перечисленных ниже, используют рекуррентные ячейки?

- ☐ A Fasttext
 ☐ C ELMO
 ☐ E BERT
- ☐ B Node2Vec
 ☐ D GPT-3
 ☐ F Нет верного ответа.

**Вопрос 5.** GRU-ячейка выглядит следующим образом.



Пусть гейт  $u$  оказался нулевым. Что это означает для нашей сетки?

- |  |   |  |
|--|---|--|
| <input type="checkbox"/> <b>A</b> Предыдущее скрытое состояние будет проигнорировано | <input type="checkbox"/> <b>C</b> Текущее скрытое состояние никак не повлияет на формирование $g$ | <input type="checkbox"/> <b>E</b> Кандидат в скрытое состояние не передаст на следующий шаг новой информации |
| <input type="checkbox"/> <b>B</b> Значение $x_t$ никак не повлияет на прогноз $y_t$  | <input type="checkbox"/> <b>D</b> Гейт $g$ окажется равен единице                                 | <input type="checkbox"/> <b>F</b> Нет верного ответа.  |

**Вопрос 6.** Выберите все верные утверждения про предобработку текстов

- |  |   |  |
|--|---|--|
| <input type="checkbox"/> <b>A</b> Очень часто встречающиеся слова (стоп-слова) обычно выкидывают из данных | <input type="checkbox"/> <b>C</b> Очень редко встречающиеся слова обычно оставляют в данных | <input type="checkbox"/> <b>E</b> Подсчёт $tf$ занижает важность стоп-слов |
| <input type="checkbox"/> <b>B</b> Сделать лемматизацию намного быстрее, чем стеминг                        | <input type="checkbox"/> <b>D</b> Подсчёт $idf$ занижает важность стоп-слов                 | <input type="checkbox"/> <b>F</b> Нет верного ответа.                      |

**Вопрос 7.** Мы обучаем word2vec. В словаре  $n$  слов. Мы учим вектора размерности  $d$  с помощью skip-gram. Сколько параметров мы будем обучать в рамках такого подхода?

- |                                       |   |   |
|---------------------------------------|---|---|
| <input type="checkbox"/> <b>A</b> $n$ | <input type="checkbox"/> <b>C</b> $n \cdot d$         | <input type="checkbox"/> <b>E</b> $n \cdot d + 1$     |
| <input type="checkbox"/> <b>B</b> $d$ | <input type="checkbox"/> <b>D</b> $2 \cdot n \cdot d$ | <input type="checkbox"/> <b>F</b> Нет верного ответа. |

**Вопрос 8.** Одно из слабых мест классического механизма внимания – время его работы. На вход в трансформер приходит текст длины  $n$ . Какова сложность по времени работы механизма self-attention?

- |  |  |   |
|--|--|---|
| <input type="checkbox"/> <b>A</b> $O(n)$   | <input type="checkbox"/> <b>C</b> $O(n^3)$ | <input type="checkbox"/> <b>E</b> $O(n!)$             |
| <input type="checkbox"/> <b>B</b> $O(n^2)$ | <input type="checkbox"/> <b>D</b> $O(2^n)$ | <input type="checkbox"/> <b>F</b> Нет верного ответа. |

**Вопрос 9.** Выберите все верные утверждения про трансформеры

- ☐ **A** Для каждого токена с помощью полносвязного слоя считаются три сущности: ключ (key), значение (value) и запрос (query)
- ☐ **B** Внутри каждого слоя энкодера есть остаточная связь (residual connection)
- ☐ **C** Обучение трансформера можно распараллелить, именно для этого его делают многоголовым (multihead self-attention)
- ☐ **D** Внутри трансформера также как и в RNN используются рекуррентные ячейки
- ☐ **E** В декодере кроме механизма самовнимания (self-attention) есть ещё и механизм внимания, на вход в который идёт выход энкодера
- ☐ **F** Нет верного ответа.

**Вопрос 10.** Представим, что у вас есть генеративная нейронная сеть (GAN), которая генерирует картинки с яблоками. Какое утверждение из представленных ниже неверное:

- ☐ **A** Основная цель генератора — выучить распределение картинок с яблоками.
- ☐ **B** После того, как мы обучим GAN, мы можем использовать дискриминатор, чтобы отличать яблоки от ананасов.
- ☐ **C** После обучения GAN ошибка дискриминатора стабилизируются около какой-то константы.
- ☐ **D** Генератор может создавать изображения с яблоками, которых никогда не существовало в природе.
- ☐ **E** Генератор и Дискриминатор нужно обучать последовательно. Сначала несколько шагов обучения генератора, потом дискриминатора.
- ☐ **F** Нет верного ответа.

**Вопрос 11.** Думал(а) когда-нибудь о том, чтобы сбежать из дома и жить в пещере?

- ☐ **A** Да, ведь я сторонник анархо-примитивизма
- ☐ **B** Только если роботы захватят мир
- ☐ **C** Да, обожаю ходить по дому голым, в пещере никто не сможет заставить меня одеться
- ☐ **D** Нет, там холодно и сыро
- ☐ **E** Да, потому что мама заставляет меня мыть посуду, а я больше не могу мыть посуду
- ☐ **F** Нет верного ответа.

## Часть вторая: вопросы и задачи

Все ответы должны быть обоснованы. Решения должны быть прописаны для каждого пункта. Рисунки должны быть чёткими и понятными. Все линии должны быть подписаны. За решение каждой задачи можно получить 8 баллов.

**Вопрос 12.** Илон Маск учит FastText для словаря из  $n$  слов. Все слова состоят из букв латинского алфавита (26 букв). Другие символы в корпусе не встречаются. Сколько параметров нужно оценить Илону Маску? Как можно будет применить модель к новым словам, которых не было в словаре?

**Вопрос 13.** Выпишите уравнения, описывающие LSTM-ячейку с забыванием и GRU-ячейку. Какие последовательности и веса нужно занулить, чтобы эти ячейки превратились в простую RNN-ячейку?

**Вопрос 14.** Объясните в чём заключается смысл механизма self-attention в трансформерах. Выпишите основные формулы.

**Вопрос 15.** Объясните в чём заключается главный недостаток Greedy Decoding. Как Beam Search пытается решить этот недостаток?

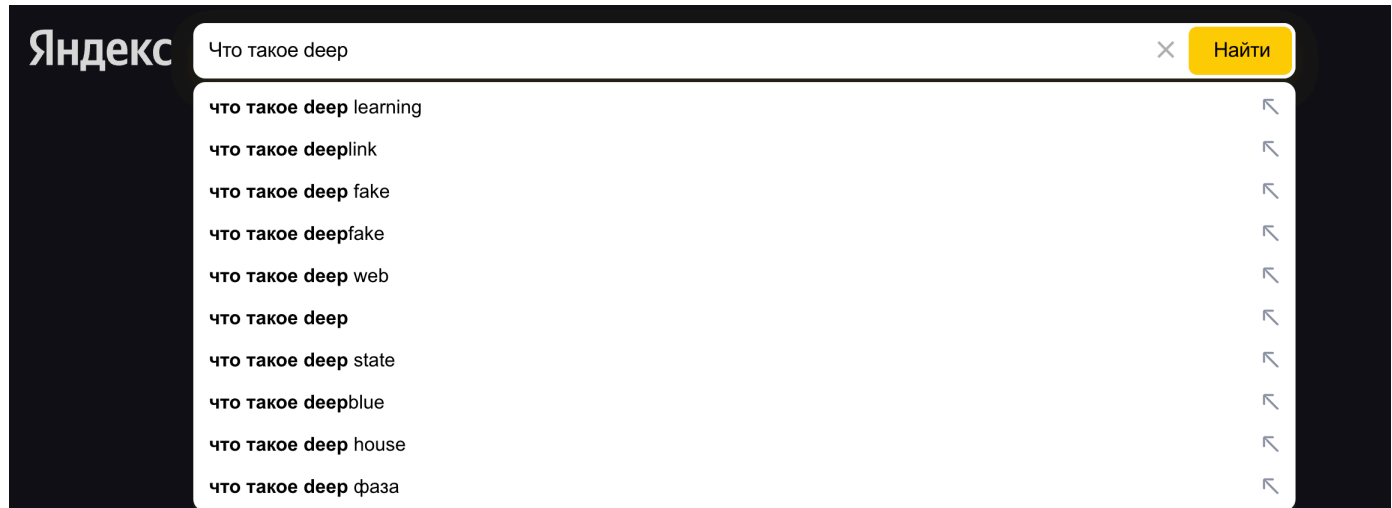
**Вопрос 16.** Объясните чем GPT отличается от BERT. Как это все связано с архитектурой трансформера? Почему в расшифровке BERT (Bidirectional Encoder Representations from Transformers) встречается слово Bidirectional?

**Вопрос 17.** Предположим, что вы обучили два w2v на разных корпусах текстов. Как бы вы выяснили, у какой из моделей более хорошее качество?

## Часть третья: ML-design

За эту задачу вы можете получить максимально 20 баллов. Попробуйте как можно чётче структурировать свои мысли. Опишите какие данные вам

**Вопрос 18.** Когда пользователь в поисковике печатает свой запрос, под строкой ввода ему обычно показываются разные варианты продолжения его запроса и другие варианты запроса (саджесты).



Вы устроились в крупную технологическую компанию ML-инженером и вам нужно реализовать такой механизм. Как бы вы решали такую задачу?

- Как бы вы сформулировали задачу ML?
- Как бы вы собирали выборку, факторы, разметку?
- Как бы вы сформулировали таргет?
- Что бы вы использовали в качестве бейзлайна (изначальное очень простое решение)?
- Какую модель вы бы использовали? Как бы вы её обучали?
- Как бы вы проверили качество внедрения?