
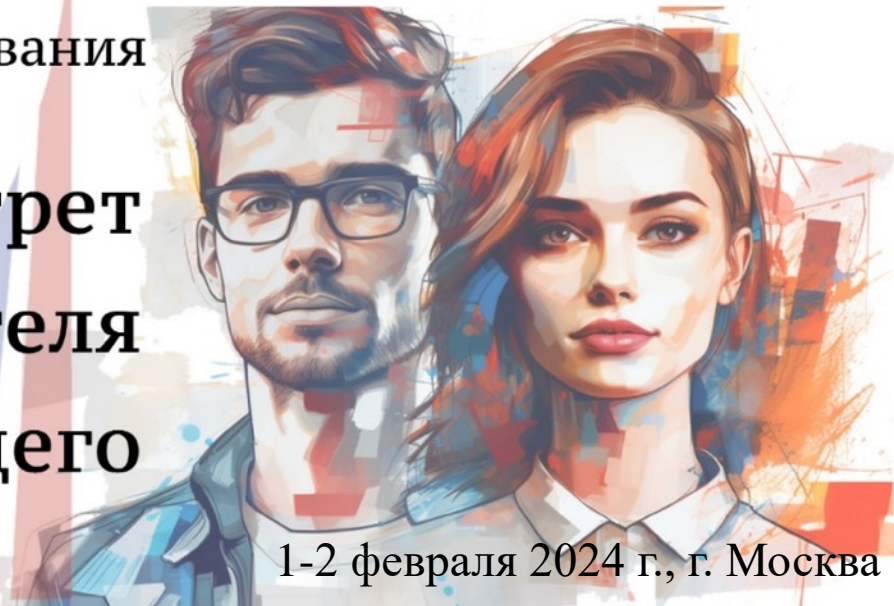


**МЕЖДУНАРОДНАЯ  
НАУЧНО-МЕТОДИЧЕСКАЯ  
КОНФЕРЕНЦИЯ**

 **Форсайт образования**

**Портрет  
преподавателя  
будущего**

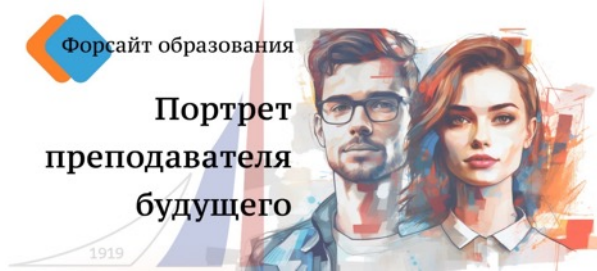


1919

1-2 февраля 2024 г., г. Москва

**Пицца для ума:  
полнотекстовые базы (ЭБС)  
как средство обучения  
Искусственного интеллекта**

Костюк Константин Николаевич  
канд. полит. наук, ген. директор  
ИД «Директ-Медиа»  
Февраль 2024

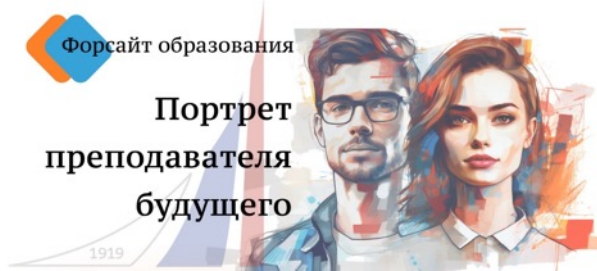


# Текст как пространство смысла

## Полнотекстовые библиотеки – старая технология?

- Текст – базовый формат хранения знаний (появились базы данных, видеоролики)
- Текст – избыточный формат по сравнению с повседневным опытом
- Текст – пространство смысла, выходящее за пределы повседневной коммуникации
- В текстах и через тексты рождается смысловая сложность, общество, культура, цивилизация (писанная история = история = общество)

Задача ИИ – достичь «сознания уровня человека» – поставлена неверно. Не нужно постигать и искать аналога «природного сознания» – нужен аналог «текстового интеллекта» (GPT)



# ИИ и Трансформеры как навигаторы текстов

Прорыв в ИИ произошел благодаря обращению к текстам и языку как основе и ткани смысловой реальности.

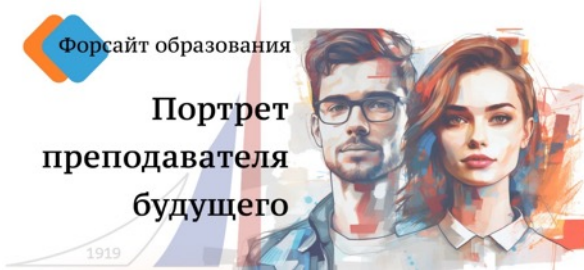
Трансформер (GPT) задуман как дешифровщик текстов – с целью создать генератор/конструктор смыслов. Основа GPT – датасет из текстов, для обучения нейросети.

Ключевое условие – создание Большой языковой модели (LLM) – кода, описывающего всю языковую коммуникацию человека.

**Речь идет о способности реконструировать смысловую коммуникацию через синтаксическую данность языка.**

Если представить коммуникацию как массив текстов, то язык – матрица генерации текстов. Постструктуралисты: язык – потенция, в языке заложена вся коммуникация, все тексты как возможность.

GPT – так же реализует возможности языка, отталкиваясь от реконструируемой им системы языка.



# Как ИИ осваивает язык

GPT делает это не как человеческое сознание, а моделируя работу языка и создавая аналог.

В онтогенезе человек осваивает язык, исходя из контакта с предметной реальностью. Это способ адаптации человека к природе. Надо помнить, что сознание и мозг являются органами животного, и решают проблемы животного тела. Решение этих проблем создает опыт. Эмоции, потребности, инстинкты – то, чего никогда не может иметь ИИ.

Язык и знаки при этом служат экономикой смысловых связей внешнего мира. При этом ребенок не изобретает язык, а опирается на существующую языковую коммуникацию.

Мы также осваиваем «иностраный» язык как новое языковое пространство. Но при этом мы уже опираемся на наш собственный язык, и изучение языка является лишь переводом нашего собственного языка.

**Трансформер GPT осваивает язык совершенно иначе, без связи с предметной реальностью. Для него не существует реальности иной, кроме как языковой, смысловой, текстовой реальности.**

Он реконструирует язык из его употребления. Этому служат модели векторизации и токенизации языка. Их цель – построить такие траектории языковой генерации (на основе статистически-вероятностных взаимосвязей), которые аналогичны нашей человеческой коммуникации.

ИИ строит языковые конструкции и «мыслит» совершенно иначе. Он «тупо» выстраивает аналогичные траектории языка, не понимая их предметного смысла и связи. Отталкиваясь (обучаясь) от обратной связи, от одобрения человеком результата. Но когда это получается, когда есть (с нашей точки зрения) осмысленность, **возникает «параллельный» язык**, обладающий абсолютно иной систематикой – язык ИИ. Мыслит ли ИИ, когда получает полностью аналогичные результаты?

# Траектории обучения языку

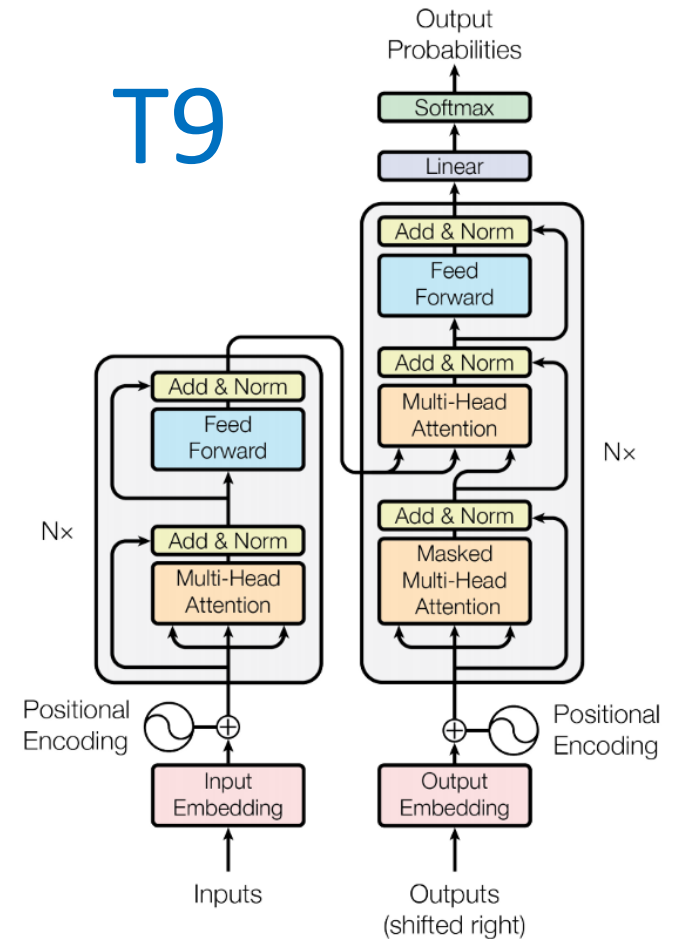
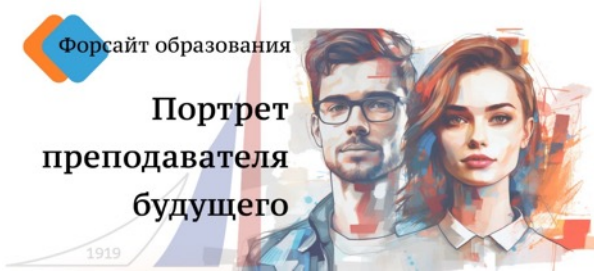


Figure 1: The Transformer - model architecture.

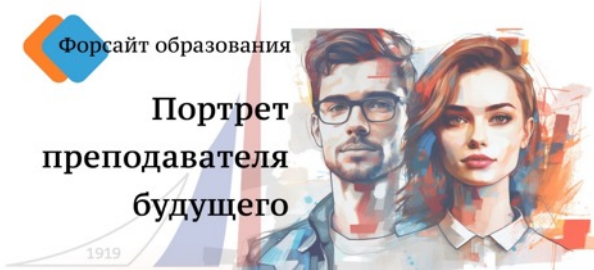


# «Человеческая размерность» текстов

- GPT «обучается» языку, основываясь на обратной связи с человеком. Учится выстраивать текстовые конструкции, которые бы «устроили» человека
- **Обучение = разметке**; человек дает сигнал, что он хочет получить. Разметка – ключ!
- Эволюция GPT-4:
  - Трансформер – 2017
  - Предобучение моделей архитектуры Трансформера - 2018
  - Подготовка Датасета и наращивание размера модели для предобучения – 2019, 2020
  - Инструкции - 2021
  - Инструкции + обратная связь от человека (чатбот RLHF) – 2022
- GPT в форме Чат-бота – возможность получать запросы и оценки от внешних пользователей

Актуальные направления развития GPT:

- Поиск моделей
- Подготовка датасетов
- Интеграция с иными технологиями



# ЭБС как база Знаний

База данных датасета – ключевое сырье для языковых моделей

- BERT – опирается на БД «Википедии»
- GPT – опирается на Reddit и два датасета с книгами художественной литературы 8 млн текстов

Разметка – т.е. получение «целей» обучения нейросети – еще один ключевой показатель. Обратная связь с человеком

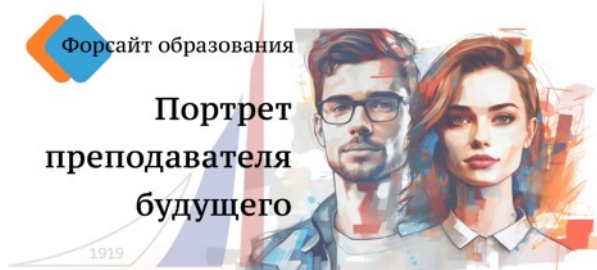
- Датасет – образ языка, из которого вырастает модель ИИ
- Разметка – образ смысловой коммуникации, на которой обучается ИИ

Моя гипотеза - дальнейшая эволюция свойств GPT во многом будет определяться тем, какое «сырье», какой язык ложится в основу ИИ

Как и среди людей, возможна различная специализация ИИ: научный, справочный, диалоговый...

Для ИИ, генерирующего знания, необходима основа в виде базы знаний (Сегодня GPT – болтун, а не ученый)

1. ЭБС – электронная библиотека научных и учебных текстов – основа для воспитания «интеллигентного», «образованного» GPT. Это основа для генерации знаний, а не просто текстов!
2. GPT не имеет знаний, он может только воспроизводить речь. Интеграция с иными системами ИИ – **Экспертными системами** – превращает его в действительно интеллектуальный ресурс. Например, интеграция со словарями и справочниками.



# Что мы можем предложить

Наша база текстов обладает следующими свойствами:


- Это база книг – учебников и монографий – содержащих универсум знаний
- На этом массиве обучается сам человек!
- Он является проверенным источником знаний, авторизованным не случайными людьми, а научным сообществом
- Объем нашего датасета составляет около 150 000 томов
- Этот датасет размечен по ключевым библиографическим параметрам – жанр, формат, авторы, дисциплины, тематики
- Этот датасет содержит полные тексты – является уже подготовленным, размеченным для обучения ИИ
- Экономия работы такого объема составляет колоссальную величину!
- В GPT только доли процента составляют тексты на русском языке. Русскоязычные тексты - самостоятельная ветка развития GPT (Yandex-Sber).



# Библиографическая информация

В библиографической информации заложена навигация по смысловому миру:

- Классификаторы тематик
- Жанровая специфика
- Возрастная разметка
- Ключевые слова
- Автор, источник
- Статистика обращений
- Оценка читателями
- Популярность



**Безопасность жизнедеятельности: учебное пособие** </>

Дисциплина: [Безопасность жизнедеятельности](#)

Жанр: [Учебники и учебные пособия для вузов](#)

Красноярск: [Сибирский федеральный университет \(СФУ\)](#), 2022

Объем: 320 стр.

Составитель: Горбунова Л.Н., Чурбакова О.В.

УДК: 658.382.3(07)  
ББК: 68.903я73

Книга доступна в мобильном приложении

Выберите действие ⌵ Читать онлайн

[Библиографическое описание](#) / [Аннотация](#) / [Содержание](#) /

Безопасность жизнедеятельности : учебное пособие : [16+] / сост. Л. Н. Горбунова, О. В. Чурбакова ; Сибирский федеральный университет. – Красноярск : Сибирский федеральный университет (СФУ), 2022. – 320 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=706640> (дата обращения: 27.01.2024). – Библиогр. в кн. – Текст : электронный.

# Элементы внутренней разметки

mb4.ru



## Содержание

|  |    |
|--|----|
| Введение   | 5  |
| Лабораторная работа №1. Исследование запыленности и загазованности воздушной среды в производственных помещениях | 6  |
| 1. Краткие теоретические сведения  | 6  |
| 2. Методы измерения концентрации пыли в воздухе  | 11 |
| 3. Порядок выполнения лабораторной работы  | 12 |
| 4. Контрольные вопросы и задания   | 19 |
| Лабораторная работа №2. Исследование искусственного освещения  | 21 |
| 1. Краткие теоретические сведения  | 21 |
| 2. Порядок выполнения лабораторной работы  | 44 |
| 3. Контрольные вопросы и задания   | 49 |
| Лабораторная работа №3. Исследование параметров микроклимата в производственных помещениях                       | 50 |
| 1. Краткие теоретические сведения  | 50 |
| 2. Порядок выполнения лабораторной работы  | 71 |
| 3. Контрольные вопросы и задания   | 73 |
| 4. Защитных экранов  | 75 |

```
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <meta charset="utf-8">
5     <title>Язык разметки</title>
6   </head>
7   <body>
8     <h1>Язык разметки</h1>
9     <p><strong>Язык разметки</strong> в компьютерной терминологии – набор
10    символов или последовательностей, вставляемых в текст для передачи
11    информации о его выводе или строении. Принадлежит классу
12    компьютерных языков.</p>
13    <h2>Структура языка</h2>
14    <p>Текстовый документ, написанный с использованием языка разметки,
15    содержит не только сам текст но и дополнительную информацию:</p>
16    <ul>
17      <li>указание на заголовки</li>
18      <li><mark>выделения</mark></li>
19      <li>списки</li>
20      <li>и т. д.</li>
21    </ul>
22    <p>В более сложных случаях язык разметки позволяет вставлять в
23    документ <i>интерактивные элементы</i> и содержание других
24    документов.</p>
25  </body>
26 </html>
```

### Язык разметки — HTML Academy

## Структура языка

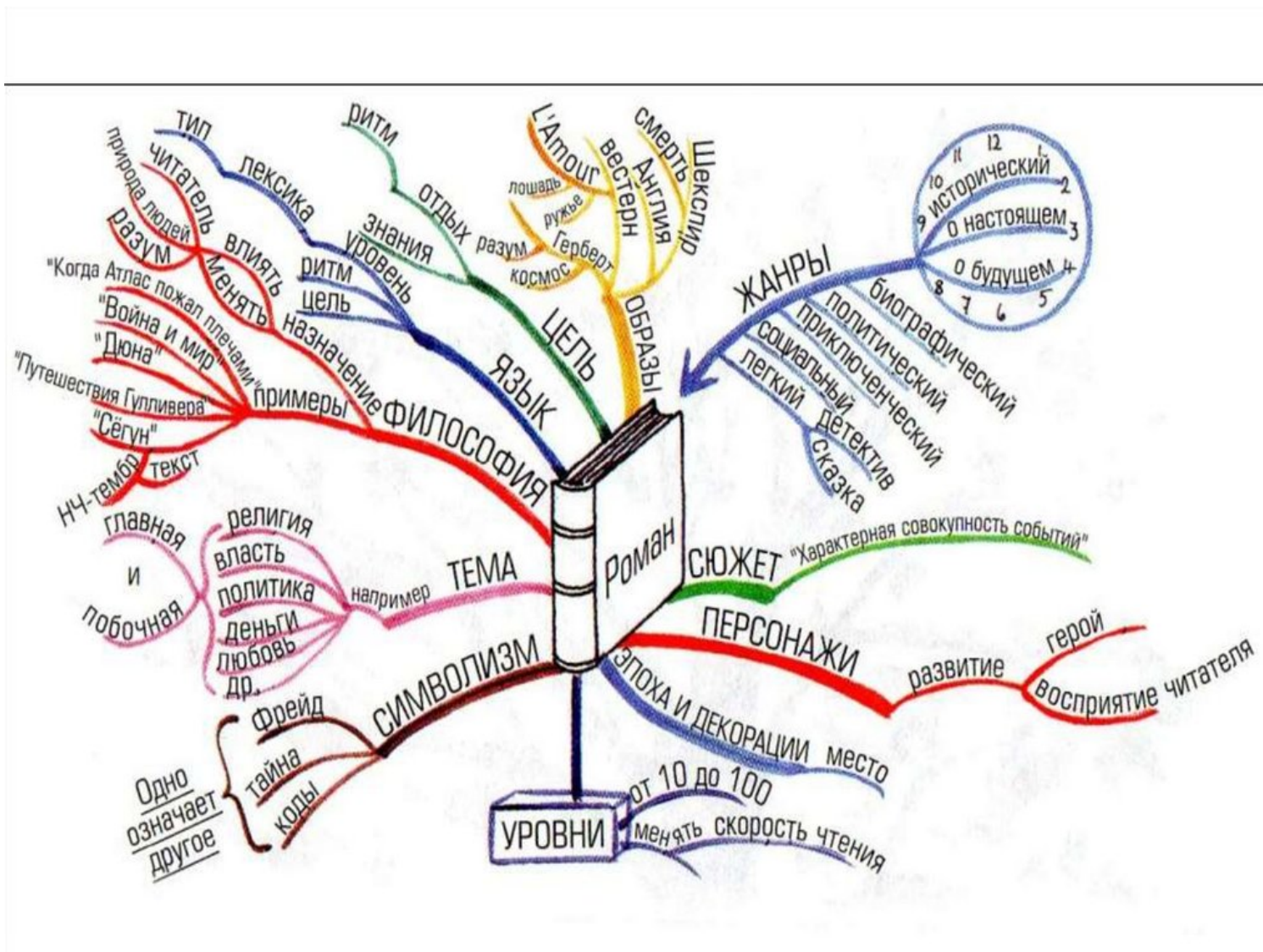
Текстовый документ, написанный с использованием языка разметки, содержит не только сам текст но и дополнительную информацию:

- указание на заголовки
- выделения**
- списки
- и т. д.

В более сложных случаях язык разметки позволяет вставлять в документ *интерактивные элементы* и содержание других документов.

# Тесты

# Внутритекстовые связи



Каждый Текст имеет свой тексто-культурный контекст. Но и сам служит контекстом других текстов. Реконструкция внутритекстовых связей превращает базу текстов в текстовый массив –

- Единый
  - Взаимосвязанный
  - Многовекторный
  - Политематический
- Элементы связей:

- Глоссарии
- Терминология
- Расшифровки имен, географических названий

# ЭБС как Экспертная система

ЭБС сама может служить как носитель ИИ – когда будет использована не только как датасет, но как платформа интегрированных решений.

Экспертная система – комплекс, способный к эффективному решению сложных задач.

- интеграция со словарями и справочниками;
  - интеграция с контрольно-тестовыми инструментами;
  - внутритекстовая, гипертекстовая связанность;
  - интеграция с учебными программами, контингентом, отчетами, -
- позволили бы выступить ЭБС в качестве обучающей платформы, экспертной системы, способной решать комплекс задач, связанных с базой Знаний!

# БЛАГОДАРЮ ЗА ВНИМАНИЕ!



Костюк К.Н.

«Информационный ландшафт  
цифровой цивилизации»

М., Директмедиа Паблишинг, 2023

Костюк К.Н.

Может ли искусственный интеллект  
мыслить?

М.: Директмедиа Паблишинг, 2024



ООО «ДИРЕКТ-МЕДИА», МОСКВА

Контактные данные:

[manager@directmedia.ru](mailto:manager@directmedia.ru)