

# Исследование архитектуры Transformers

## Описание выбранной архитектуры

Архитектура Transformers была впервые представлена в статье "Attention is All You Need" (2017) от группы исследователей из Google. В основе Transformers лежит механизм внимания (attention), который позволяет модели сосредотачиваться на определенных частях входной последовательности при создании выхода. Ключевыми компонентами этой архитектуры являются:

### 1. Механизм внимания

- **Self-attention**

Позволяет каждому элементу последовательности учитывать все остальные элементы, выделяя важные для текущей задачи.

- **Dot-product attention**

Для вычисления внимания используется скалярное произведение, что позволяет эффективно обрабатывать длинные последовательности.

### 2. Позиционные кодировки

Так как модель не обладает внутренним представлением порядка, используются позиционные кодировки, которые добавляются к входным данным.

### 3. Многоуровневый подход

Transformers применяют механизм внимания многократно с помощью слоев внимания и feed-forward сетей, что позволяет захватывать сложные зависимости.

### 4. Отказ от рекурсии

В отличие от RNNs, Transformers обрабатывают данные параллельно, что улучшает производительность на длинных последовательностях за счет уменьшения времени обучения.

# Области применения

## 1. Обработка естественного языка (NLP)

- **Машинный перевод**

Модели, такие как BERT и GPT, добились значительных успехов в переводе с одного языка на другой без явного привлечения традиционных методик.

- **Обработка текста и вопросно-ответные системы**

Архитектура Transformers в основе BERT и других моделей позволяет эффективно обрабатывать и интерпретировать текстовые данные.

## 2. Компьютерное зрение

- **Image Transformer**

Использование энкодеров для обработки изображений с целью захвата длинных зависимостей в пиксельных данных.

- **Vision Transformers (ViT)**

Использует квадратные блоки изображения как "слова", что значительно улучшило результаты в задачах классификации изображений.

## 3. Акустические данные

- **Расшифровка речи**

Применяются для улучшения точности распознавания речи за счет улучшенной обработки последовательностей звуковых данных.

# **Актуальность**

## **1. NLP**

Transformers остаются основой для большинства современных NLP моделей и продолжают быть актуальными. Развитие идет в сторону оптимизации и уменьшения моделей, например, через дистилляцию.

## **2. Компьютерное зрение**

Архитектура Vision Transformers активно применяется и развивается, и большинство исследований подтверждают её эффективность наравне или даже лучше CNN в ряде задач.

## **3. Распознавание речи**

В этой области Transformers внедряются совместно с другими методами, такими как CTC и RNN-T, показывая высокие результаты и актуальность.

## **Заключение**

На сегодняшний день архитектура Transformers продолжает оставаться актуальной благодаря своей гибкости и способности обучаться на больших объемах данных. Однако, с ростом интереса к уменьшению вычислительных затрат, ведутся разработки по улучшению эффективности и уменьшению объема моделей (например, TinyBERT, DistilBERT). Эти разработки позволяют сохранять актуальность Transformers, адаптируя их к новым требованиям масштабируемости.