

# Huggin face

## 1. Text - bert-base-NER

Модель bert-base-NER основана на архитектуре BERT - механизм трансформеров с многоголовым вниманием. Принцип её работы в том, что каждое слово в предложении анализируется в контексте всех остальных слов, а не только соседних.

В задаче Named Entity Recognition (NER) BERT последовательно классифицирует каждый токен как принадлежащий к одному из заранее заданных классов: O (вне сущности), B-XXX (начало сущности), I-XXX (продолжение сущности), где XXX — тип сущности (LOC, PER, ORG, MISC).

### Метки классов:

LOC - географический объект

PER - имя человека

ORG - название организации

MISC - сущность

### Датасет

Модель обучалась на **CoNLL-2003 NER**, который собран на новостях агентства *Reuters*. Данные размечены по схеме BIO — Begin, Inside, Outside.

В датасете 946 статей, почти 15 тысяч предложений и более 200 тысяч токенов.

Наборы выглядят так:

Dataset	LOC	MISC	ORG	PER
Train	7140	3438	6321	6600
Dev	1837	922	1341	1842
Test	1668	702	1661	1617

### Метрики качества

Metric	Dev	Test
f1	95.1	91.3
precision	95.0	90.7
recall	95.3	91.9

## **Результат использования**

### **Входные данные:**

Arkady Volozh founded Yandex in Moscow, Russia, in 1997, building it into one of the leading technology companies in Eastern Europe.

### **Выходные данные:**

Ark - B-PER - 0.9976  
##ady - B-PER - 0.8354  
Vol - I-PER - 0.9972  
##oz - I-PER - 0.9962  
##h - I-PER - 0.9677  
Yan - B-ORG - 0.9980  
##de - I-ORG - 0.9927  
##x - I-ORG - 0.9972  
Moscow - B-LOC - 0.9997  
Russia - B-LOC - 0.9998  
Eastern - B-LOC - 0.9980  
Europe - I-LOC - 0.9959

## **2. Video - videomae-base-finetuned-kinetics**

Модель VideoMAE (Video Masked Autoencoder) основана на архитектуре Vision Transformer и расширяет идею Masked Autoencoders для видео. Основная идея заключается в том, что модель обучается восстанавливать замаскированные части видеокадров, тем самым осваивая внутреннее представление структуры и динамики видео.

Видео разбивается на последовательность патчей фиксированного размера, который преобразуется в вектор и подаётся в энкодер. К началу последовательности добавляется специальный токен [CLS], который агрегирует информацию обо всём видео и используется для классификации.

## Датасет

Модель была обучена на датасете **Kinetics-400**, который содержит 400 классов человеческих действий, например: "playing piano", "drinking coffee", "climbing stairs", "hugging" и т.д.

## Метрики качества

Результаты на тестовой выборке:

**Top-1 Accuracy:** 80.9%

**Top-5 Accuracy:** 94.7%

Топ-1 показывает, насколько часто модель угадывает точный класс, а Топ-5 — насколько часто правильный класс входит в пятёрку наиболее вероятных.

## Результат использования

**Входные данные:** dance.mp4

**Выходные данные:**

```
[{'score': 0.3257863521575928, 'label': 'cartwheeling'},  
 {'score': 0.13009761273860931, 'label': 'belly dancing'},  
 {'score': 0.0891110822558403, 'label': 'dancing ballet'},  
 {'score': 0.05970514938235283, 'label': 'somersaulting'},  
 {'score': 0.02212045155465603, 'label': 'bending back'}]
```

# PyTorch

## 3. Audio - Silero VAD

Модель Silero VAD (Voice Activity Detector) предназначена для обнаружения голосовой активности в аудио — определения участков, где присутствует речь, и отделения их от тишины или шумов.

Модель принимает аудиосигнал, разбивает его на короткие фреймы, и для каждого фрейма вычисляет вероятность того, что в нём присутствует речь. На основе этих вероятностей формируются интервалы, где обнаружена голосовая активность.

## **Метрики качества**

Metric	Test
ROC-AUC	0.94
Accuracy (speech datasets)	0.91
Accuracy (noise datasets)	0.59

## **Результат использования**

**Входные данные:** audio.mp3

**Выходные данные:** [{'start': 32, 'end': 11663}]

# **TensorFlow**

## **4. Image - Frame Interpolation for Large Motion (FILM)**

Модель FILM (Frame Interpolation for Large Motion) предназначена для интерполяции кадров видео — то есть генерации промежуточных кадров между двумя существующими.

Модель работает в два шага: сначала оценивает оптический поток — то есть движение между первым и вторым кадром, а затем нейросеть восстанавливает промежуточный кадр, устранивая разрывы и добавляя детали.

## **Датасет**

Модель обучалась на **Vimeo-90K** — большом наборе видеоданных. Он содержит 89 800 видеоклипов, загруженных с vimeo.com, которые охватывают широкий спектр сцен и действий.

## **Метрики качества**

Чтобы понять, насколько хорошо FILM восстанавливает промежуточные кадры, я посмотрела на три основных метрики, которые сравнивают результат модели с ground truth — то есть с настоящим промежуточным кадром.

**PSNR (Peak Signal-to-Noise Ratio)** — измеряет точность восстановления пикселей: чем выше значение, тем меньше ошибка.

**SSIM (Structural Similarity Index)** — оценивает структурное сходство, обращая внимание на яркость, контраст и текстуры. Значения, близкие к 1, говорят о хорошем совпадении.

**LPIPS (Learned Perceptual Image Patch Similarity)** — отражает перцептивное качество с точки зрения человеческого зрения: чем меньше значение, тем более реалистичным кажется кадр.

Metric	Test
PSNR	25.3
SSIM	0.87
LPIPS	0.13

## LLM

В качестве LLM я выбрала **gemma3:1b** в связке с **Ollama** — инструментом, который автоматизирует процесс развертывания LLM. Он самостоятельно загружает веса, инициализирует модель и предоставляет интерфейс для взаимодействия через API, консоль или само приложение.

The screenshot shows the Ollama interface with a dark theme. At the top, there are window control buttons (red, yellow, green, close, minimize, maximize). A button labeled "подскажи рецепт блинов" (Tell me a pancake recipe) is visible. Below it, a message says: "Конечно! Вот простой и вкусный рецепт блинов, который легко приготовить дома:" (Of course! Here is a simple and delicious pancake recipe that is easy to make at home:). The recipe details are as follows:

**Ингредиенты:**

- 100 г муки
- 250 мл молока (можно использовать любое)
- 1 яйцо
- 1/2 чайной ложки разрыхлителя (или 1/4 чайной ложки соды)
- 1/4 чайной ложки соли
- 2 столовые ложки растительного масла (или сливочного)

**Приготовление:**

Send a message

gemma3:1b ↑