

# Будущее ИИ:

что для нас означает смена парадигмы обучения моделей

# Михаил Степнов

Исполнительный директор по  
исследованию данных, Sber AI

Курирует направления NLP, Graph ML в **Sber AI**

Создает среду для коммуникации и обмена  
опытом в области анализа данных и  
искусственного интеллекта с бизнесом и  
гос.органами



# Революция в мире ИИ

Главной новостью августа в мире ИИ стало заявление исследователей Центра исследований моделей фундамента при Университете Стэнфорда о радикальной смене парадигмы в обучении моделей



Stanford  
University

БУДУЩЕЕ ЗА

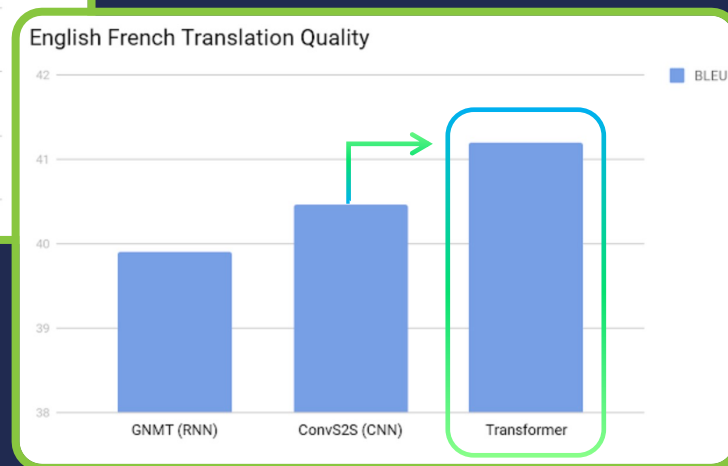
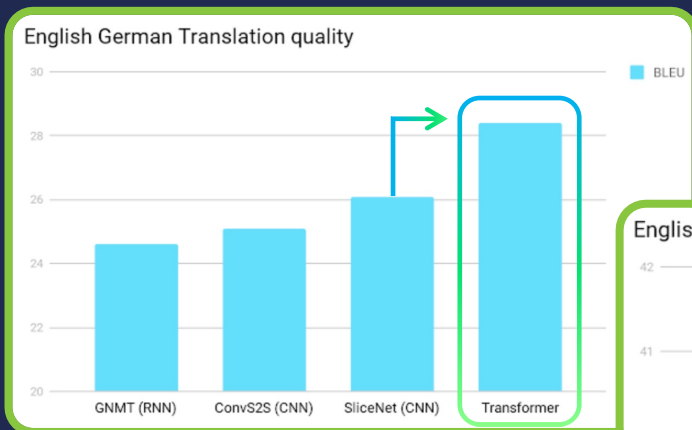
# FOUNDATION MODELS

# С чего все началось?

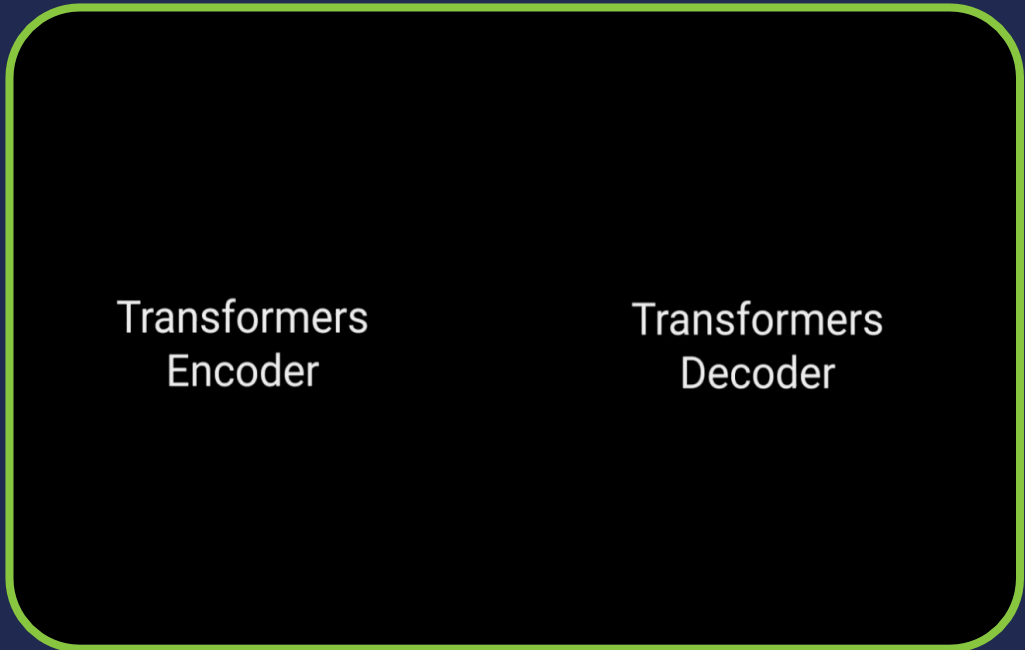
4 года назад вышла статья **”Attention is all you need”** – и перевернула мир Deep Learning

Первый прорыв трансформеров – задача машинного перевода

Принцип работы трансформеров

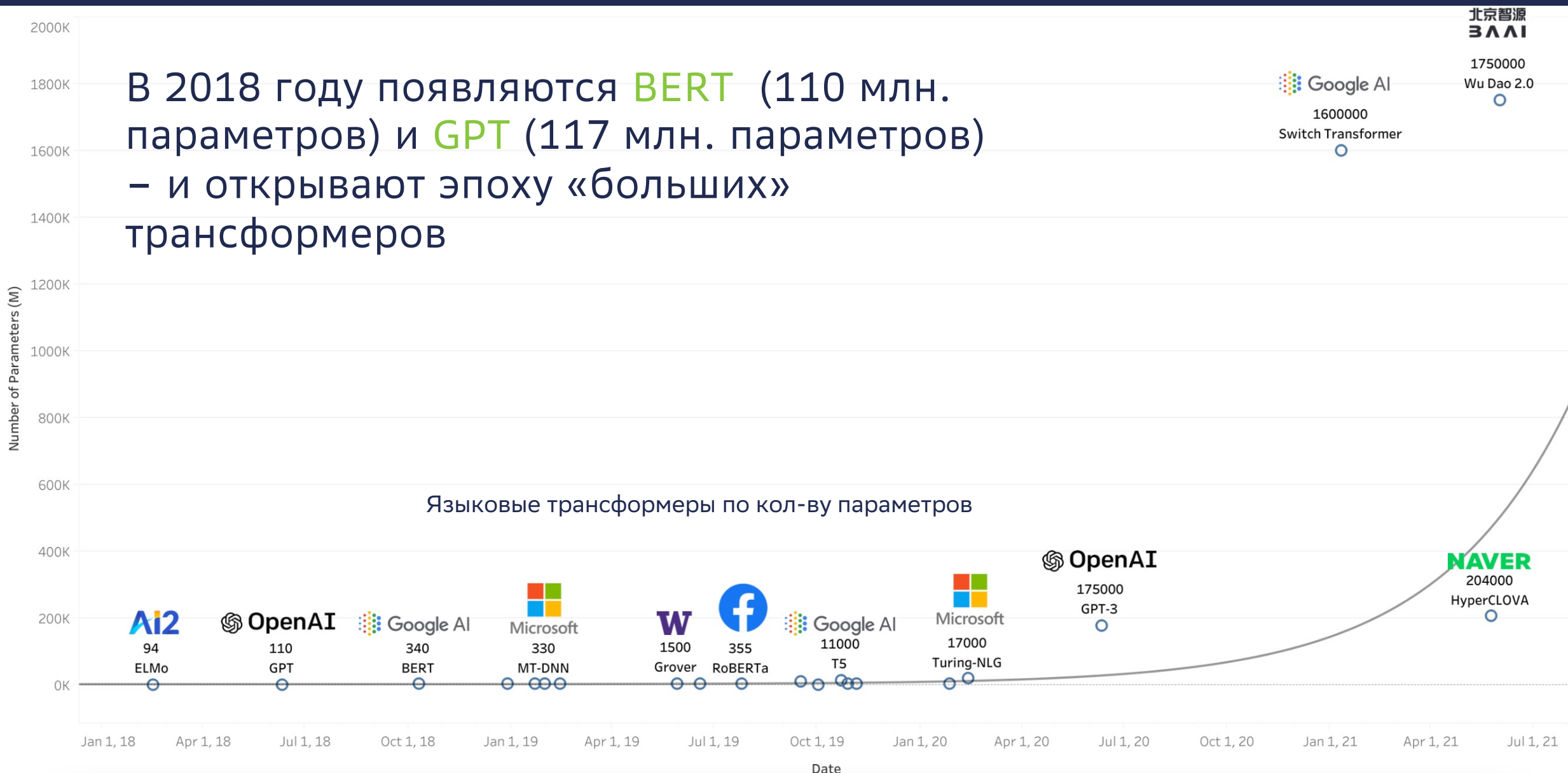


**BLEU (Bilingual Evaluation Understudy)** – это измерение различий между автоматическим переводом и одним или несколькими эталонными пользовательскими переводами одного исходного предложения.



# Большие модели и «стероидная гонка»

В 2018 году появляются **BERT** (110 млн. параметров) и **GPT** (117 млн. параметров) – и открывают эпоху «больших» трансформеров



# Каждому дата сайнтисту по трансформеру!

Появившись как архитектура для языковых моделей, трансформер доказывает свою эффективность для любой модальности

Vision Transformer –  
CV-трансформер



Jukebox – генерация  
музыки



Decision Transformer  
– RL-трансформер



Graphormer –  
графовый  
трансформер



TransGAN – GAN-  
трансформер,  
Creative AI



Transformer in  
Transformer – CV-  
трансформер



Temporal Fusion  
Transformer



И другие...

# One Transformer to rule them all

Совмещение модальностей в одной мета-модели – логичный следующий шаг, агрегирующий результаты разрозненных моделей

## Facebook Unified Transformer

- 2 модальности (CV, NLP)
- 8 downstream tasks
- Архитектура с единым декодером

## CLIP и DALL-E



- 2 модальности (CV, NLP)
- Модель для генерации и классификации
- Разработка OpenAI
- SberDevices и SberAI обучают ruCLIP и ruDALL-E

## Shared Workspace

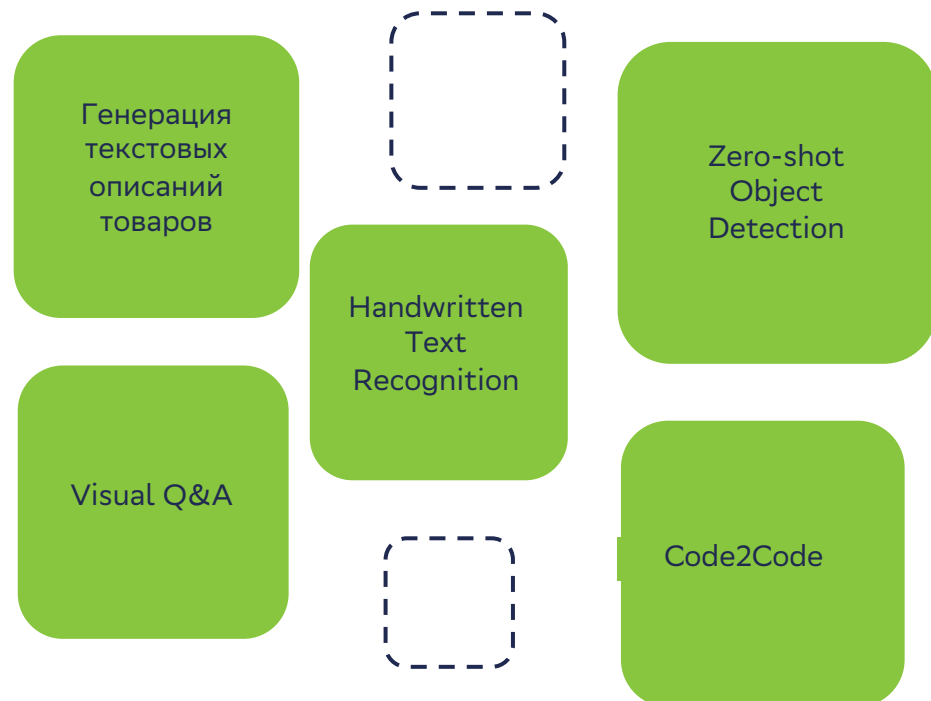
- Подход группы Иошуа Бенджио
- Единый слой нейросети для оркестрации нескольких workers
- Вдохновлен AI-исследователями 70х-80х

## Fusion Brain



- 4 модальности (CV, NLP, Time Series, Graphs)
- 16 downstream tasks
- Архитектура с единым декодером
- Разработка AIR Institute и Sber AI

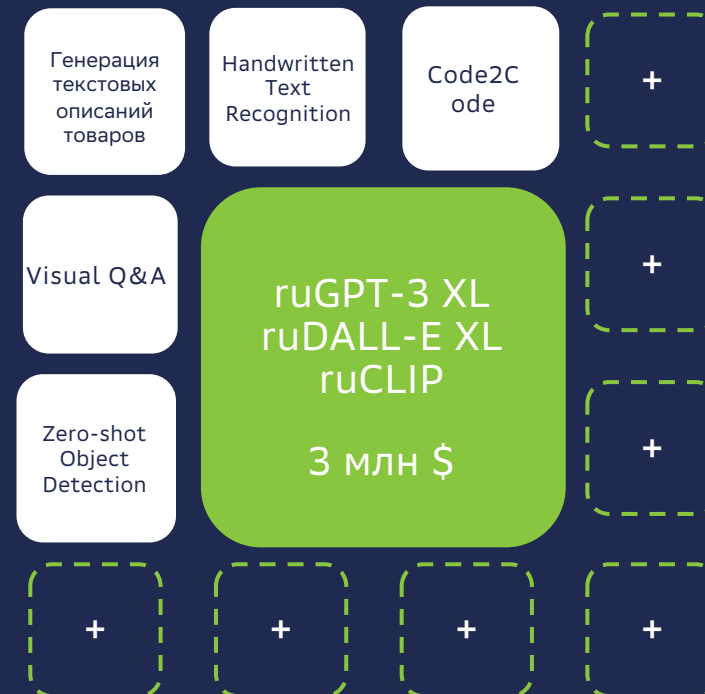
# Почему Foundation Models это хорошо?



Решение каждой задачи с нуля:  
**МНОГО** мощностей и данных на решение каждой

≈ **2300 млн GPU minutes**

**VS**



Использование предобученных моделей: **небольшое**  
 количество вычислительных ресурсов и данных для дообучения

≈ **100 млн GPU minutes**

# Классическая парадигма обучения

100 %

% of Training Parameters

## Task

Handwritten Text Recognition

Распознавание эмоций на аудио

Классификация эмоциональной окраски отзыва

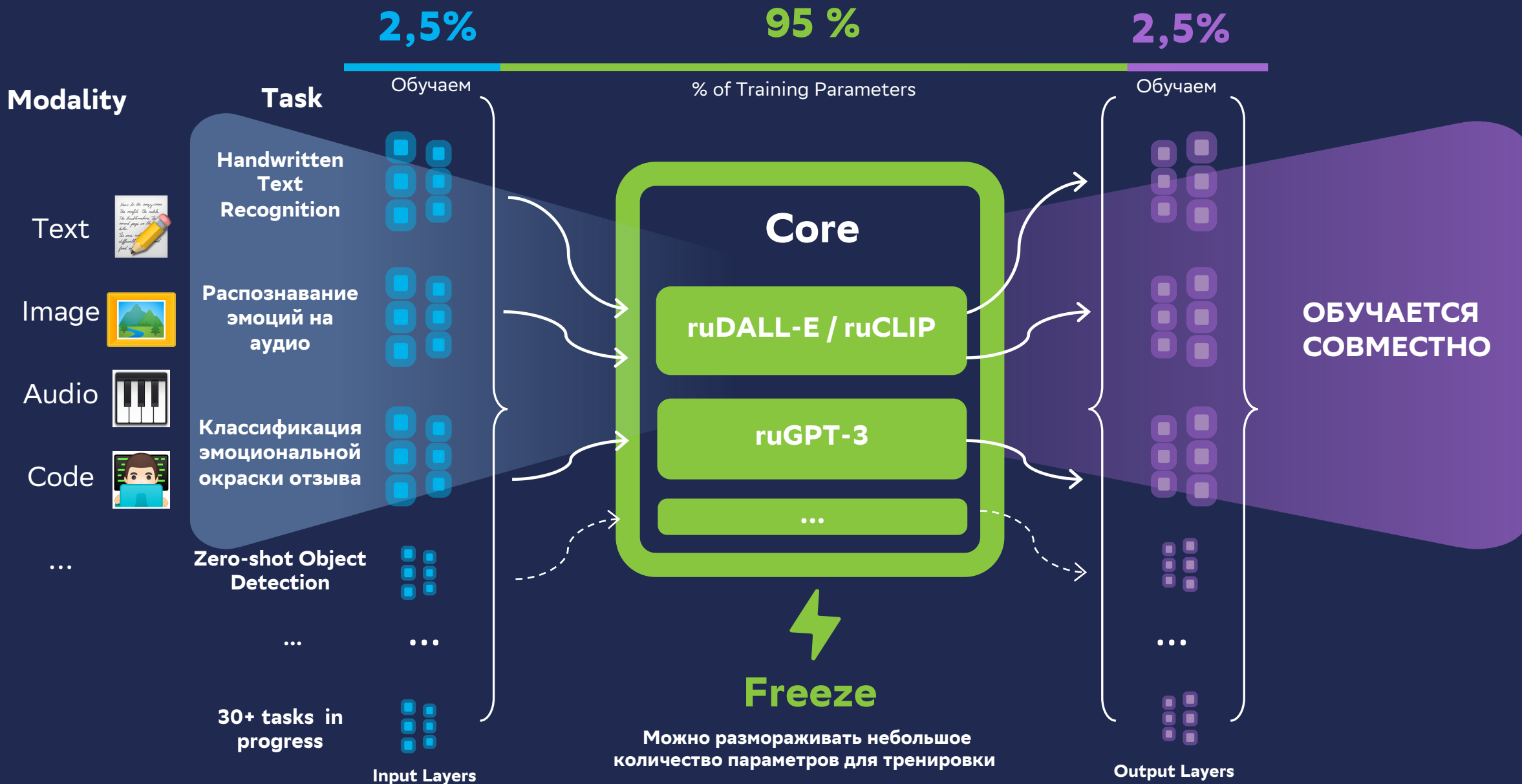
Zero-shot Object Detection

30+ tasks in progress

Все задачи обучаются отдельно



# Мультимодальная мультитаск архитектура



**Больше не нужно выбирать  
между ценой и качеством модели**