



ІІТМО

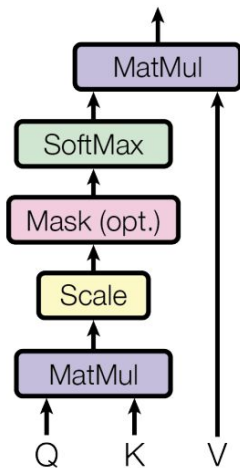
**Современные архитектуры
нейронных сетей**

Эффективное внимание

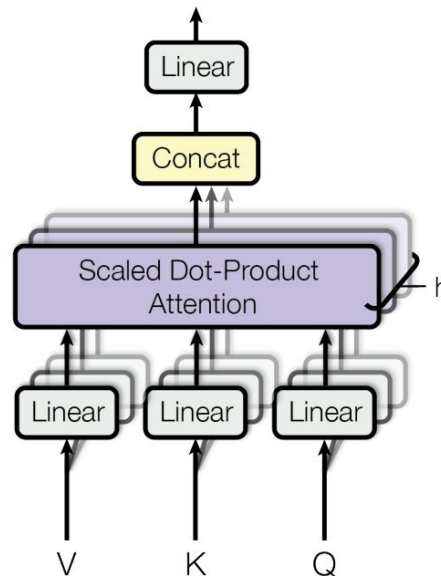
Проблемы qkv-внимания



Scaled Dot-Product Attention



Multi-Head Attention

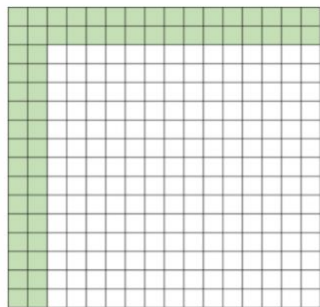
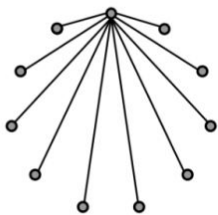


Проблемы qkv-внимания



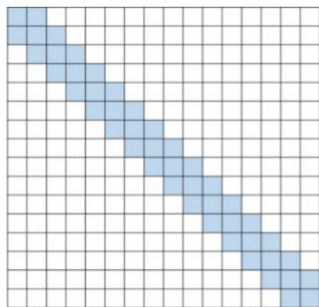
- Квадратичная сложность
- Необходимо много памяти
- Много вычислений

Sparse Attention



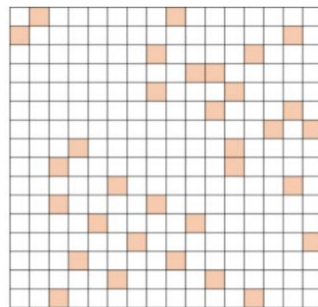
Global

+



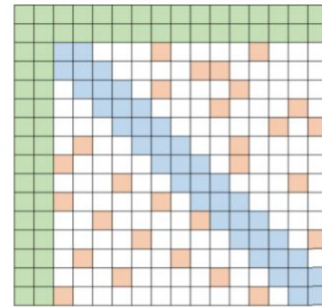
Local

+



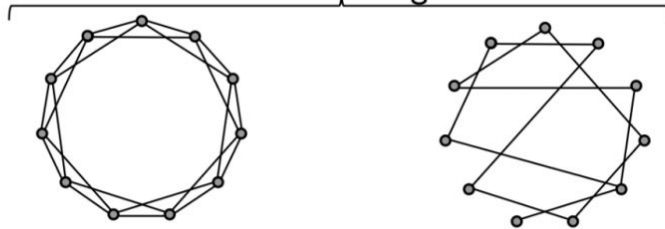
Random

=

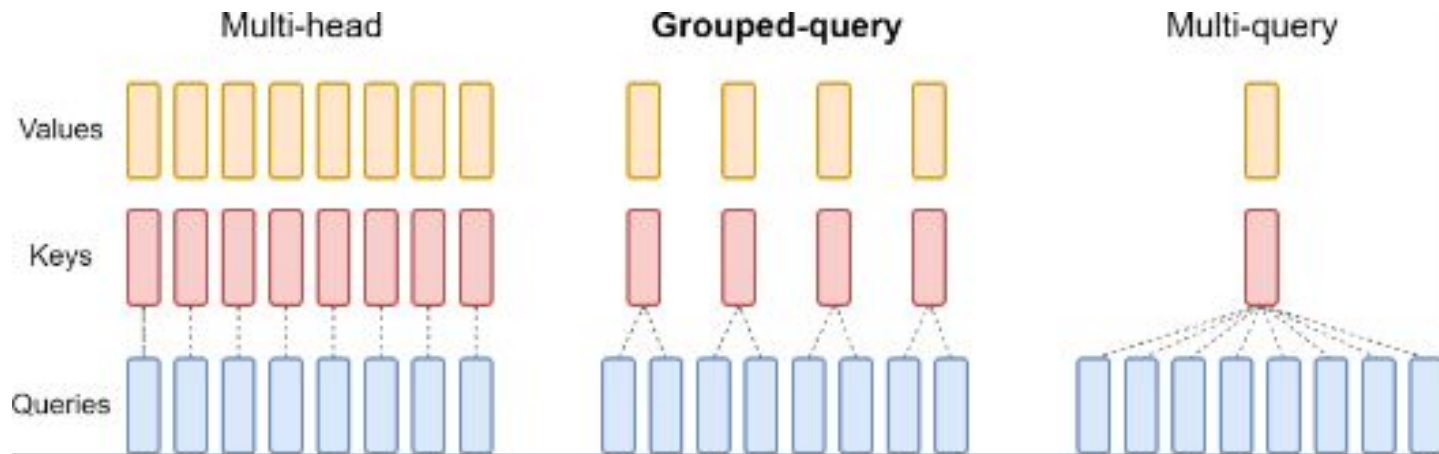


BigBird

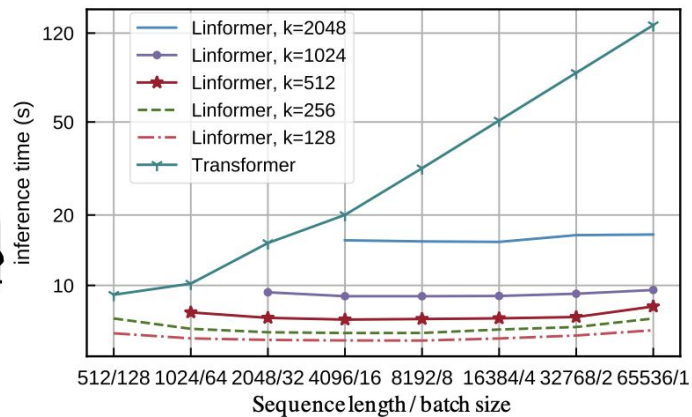
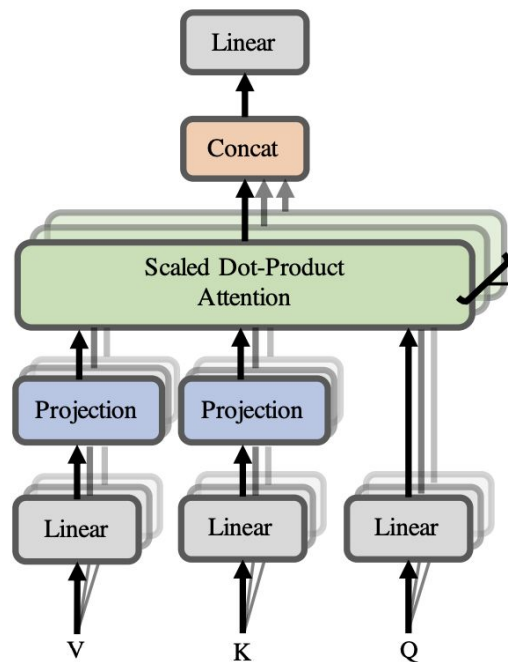
Watts-Strogatz



Multi-Query Attention



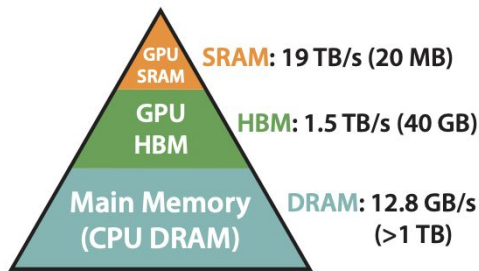
LinFormer



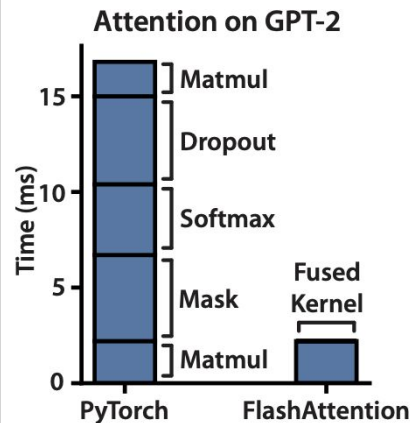
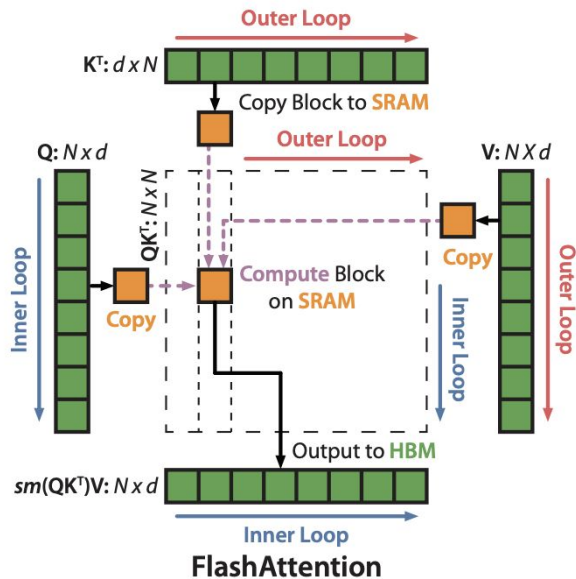
Linformer

$$\begin{matrix} k \times n \\ \text{Grid} \end{matrix} \times \begin{matrix} K (V) \\ n \times d_m \\ \text{Grid} \end{matrix} \times \begin{matrix} W^K (W^V) \\ d_m \times d_k \\ \text{Grid} \end{matrix} = \begin{matrix} k \times d_k \\ \text{Grid} \end{matrix}$$

FlashAttention



Memory Hierarchy with Bandwidth & Memory Size



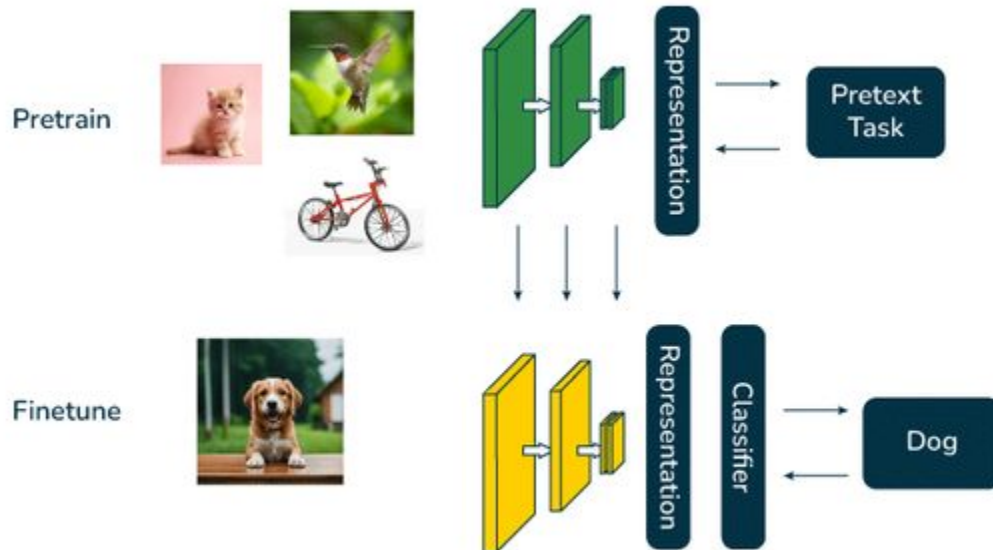


ІІТМО

**Современные архитектуры
нейронных сетей**

Self-supervised learning

Self-supervised learning

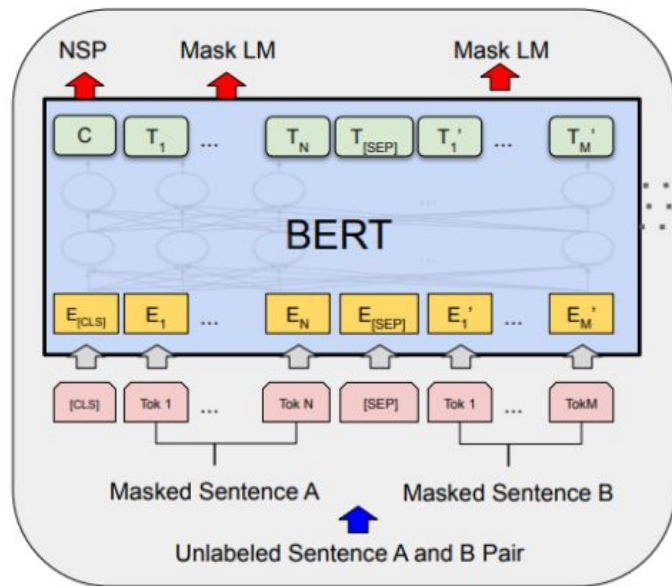


Типы архитектур трансформеров

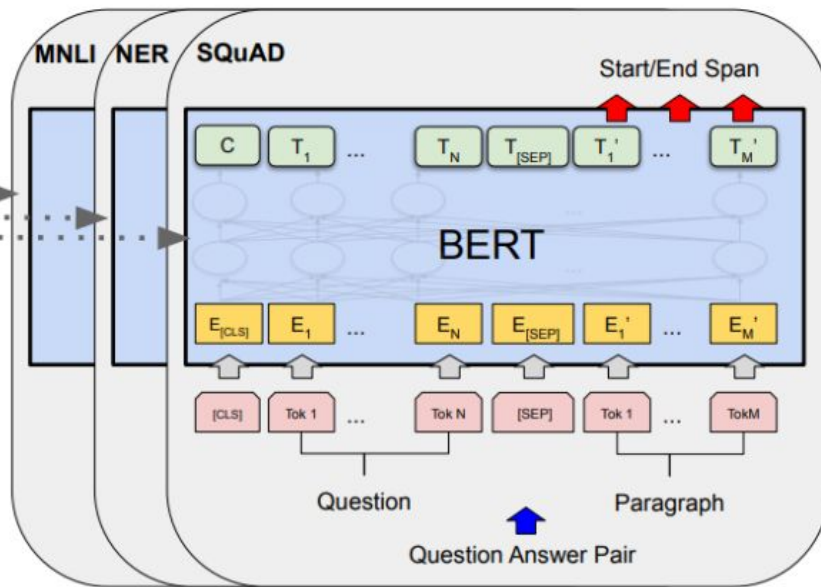


- **Encoder-Decoder** (Оригинальная архитектура трансформера)
- **Encoder** (ViT)
- **Decoder** (Пока не встречали таких)

BERT



Pre-training



Fine-Tuning

Обучение BERT

Pretraining

MLM

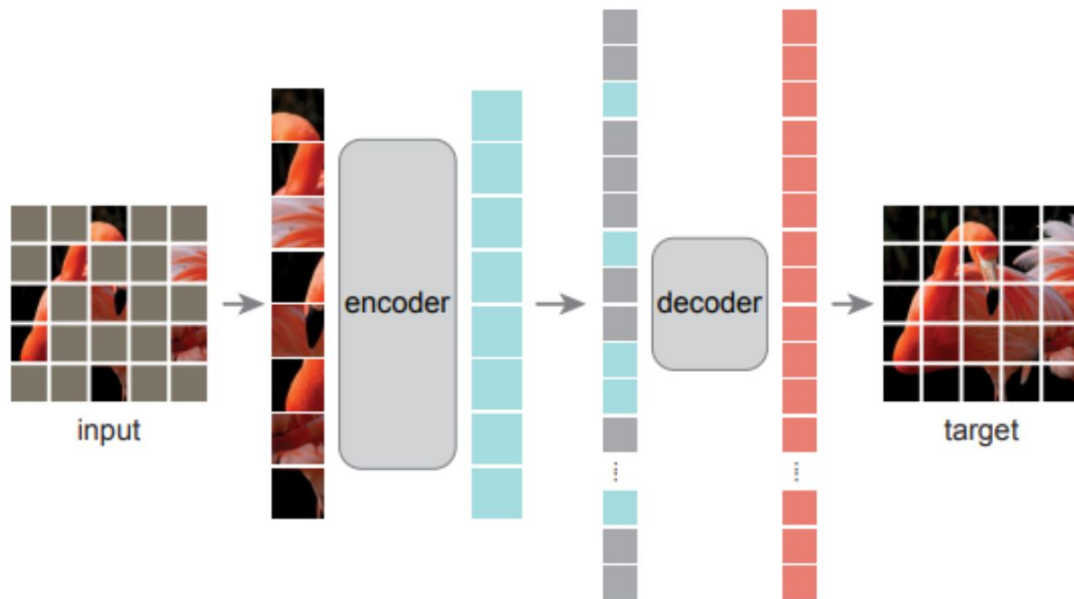
NSP

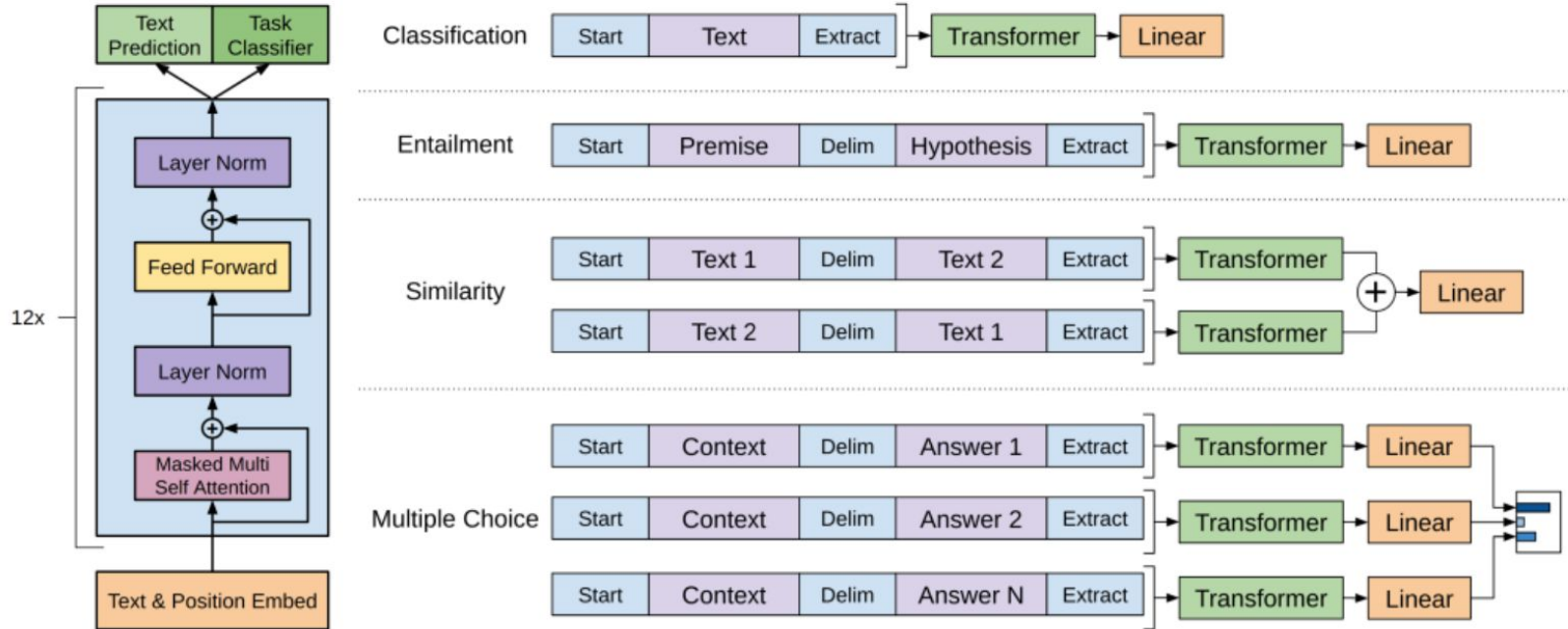
Fine-tune



ViTMAE

ViTMO





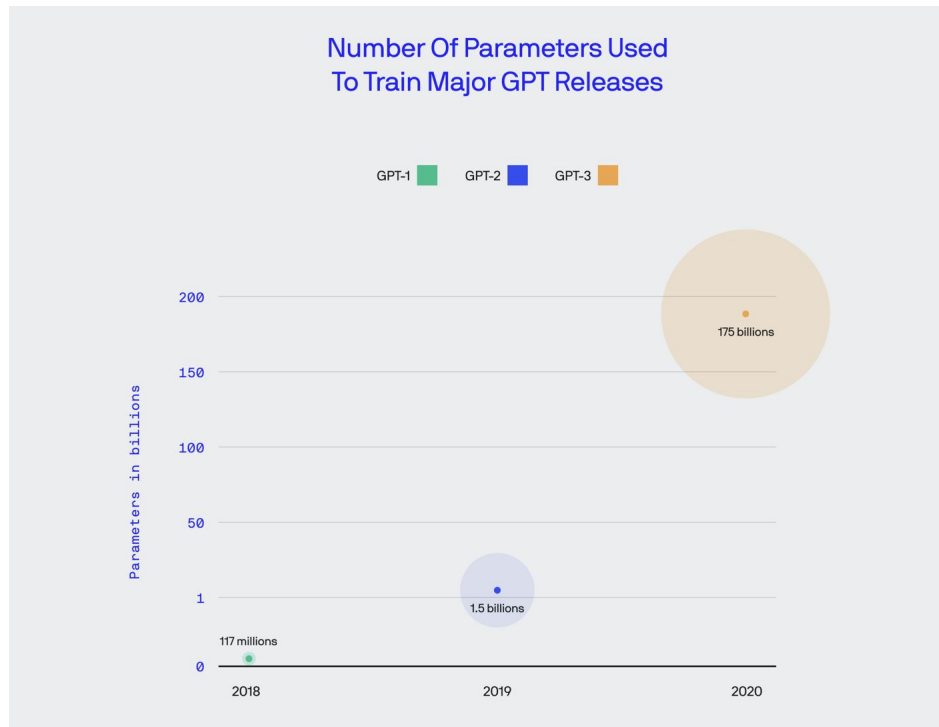


ІТМО

**Современные архитектуры
нейронных сетей**

Большие языковые модели

Эволюция моделей GPT

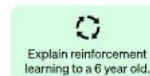


RL по обратной связи от человека

Step 1

Collect demonstration data and train a supervised policy.

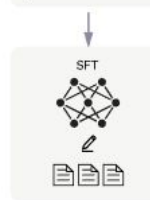
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2

Collect comparison data and train a reward model.

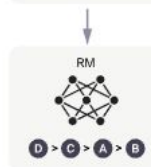
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.



The PPO model is initialized from the supervised policy.



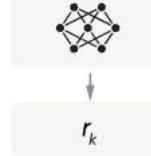
The policy generates an output.



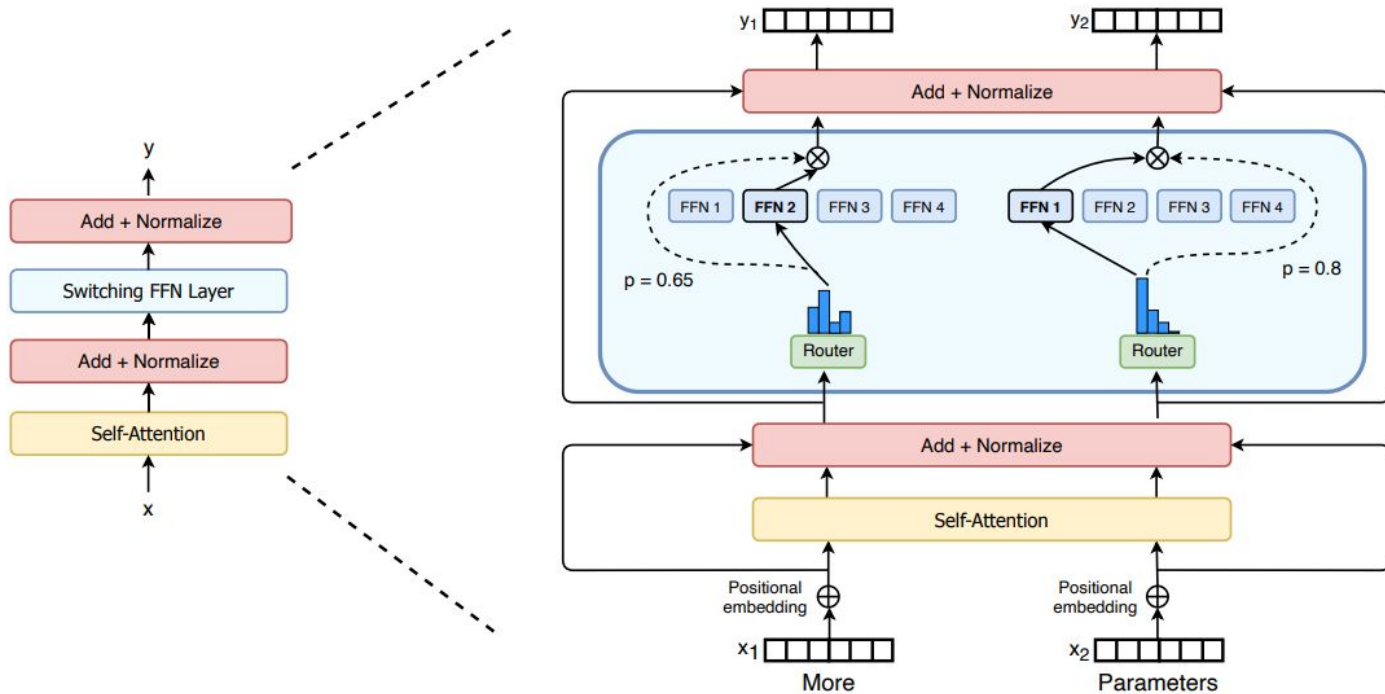
The reward model calculates a reward for the output.



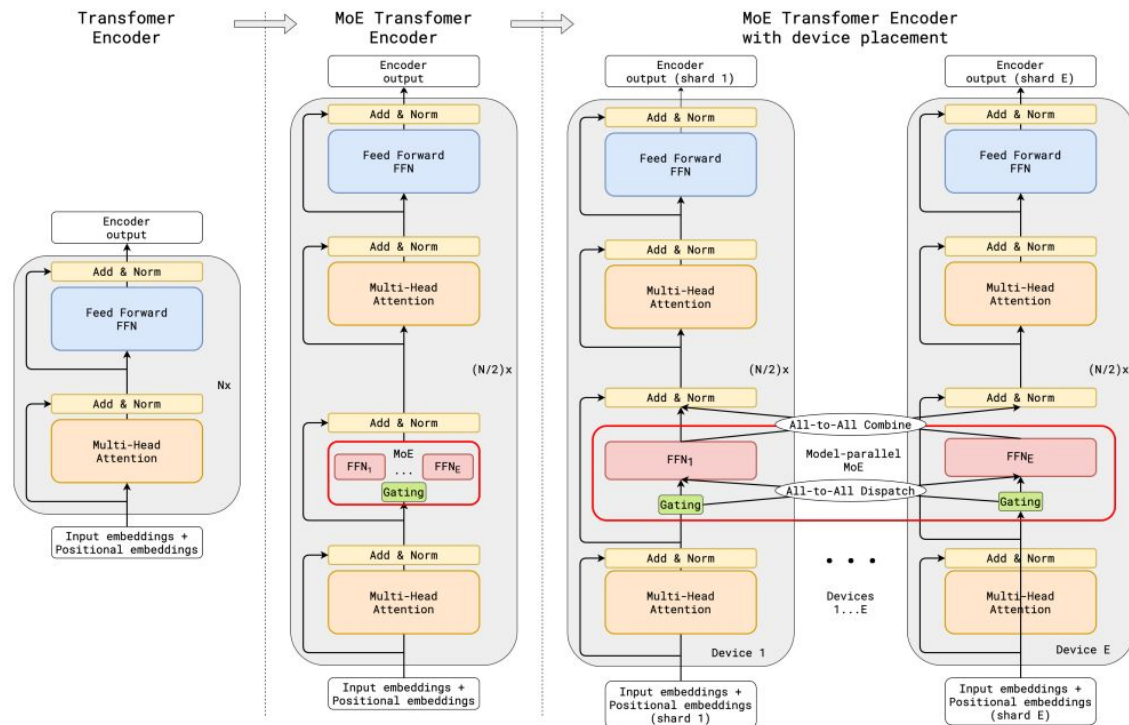
The reward is used to update the policy using PPO.



Mixture of Experts (MoE)

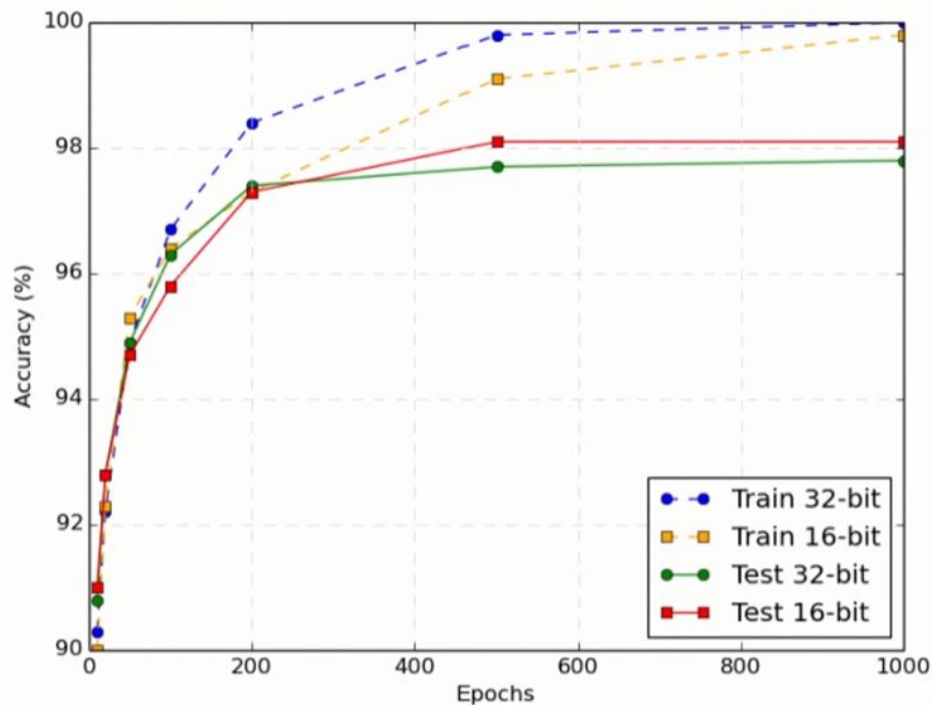


Mixture of Experts (MoE)

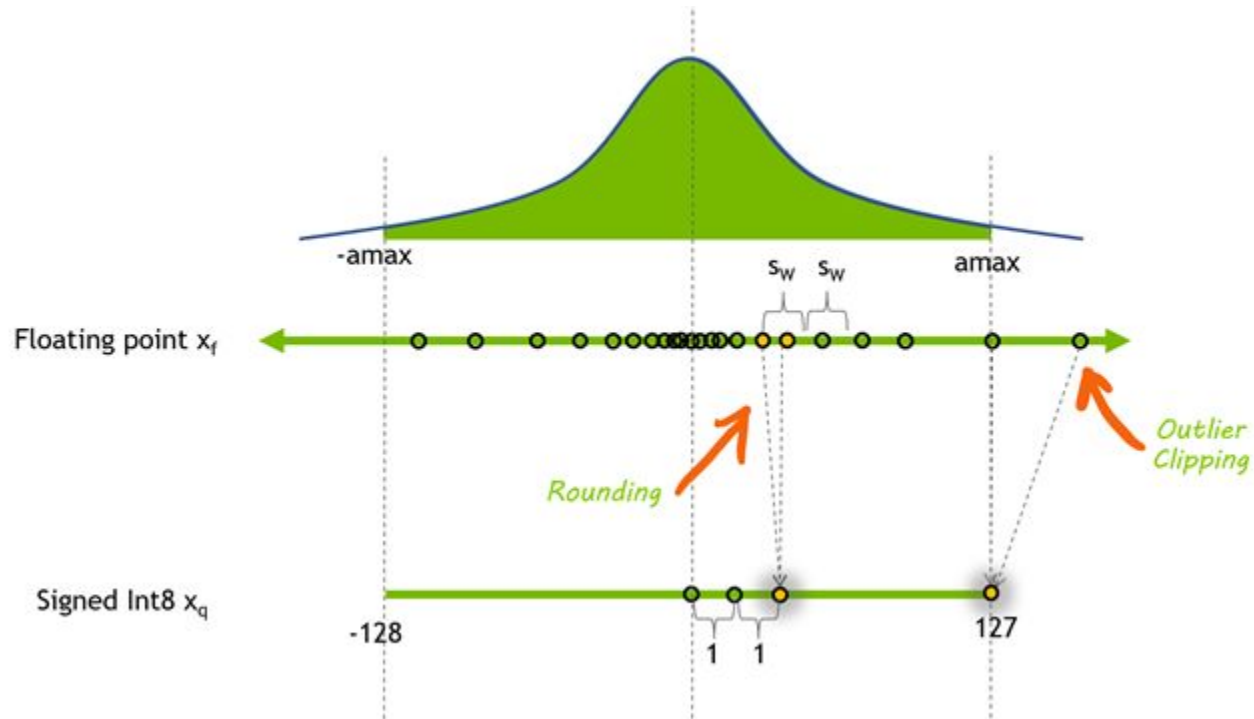


Половинная точность

ИТМО



Квантизация



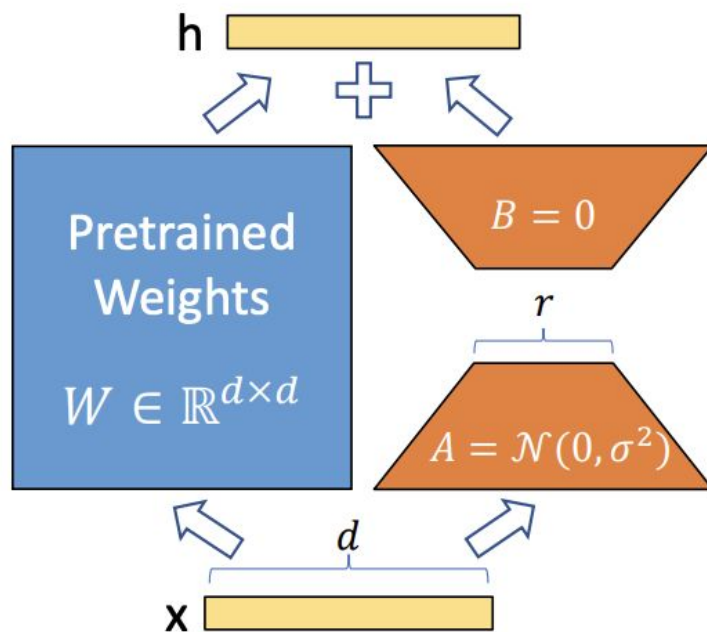


Схема работы LLM+RAG

