

# Simplex LoRA: Expand important adapters

Давыденко Григорий, Шалыгин Игорь

Научный руководитель: Безносиков А. Н.

Научный консультант: Веприков А. С.

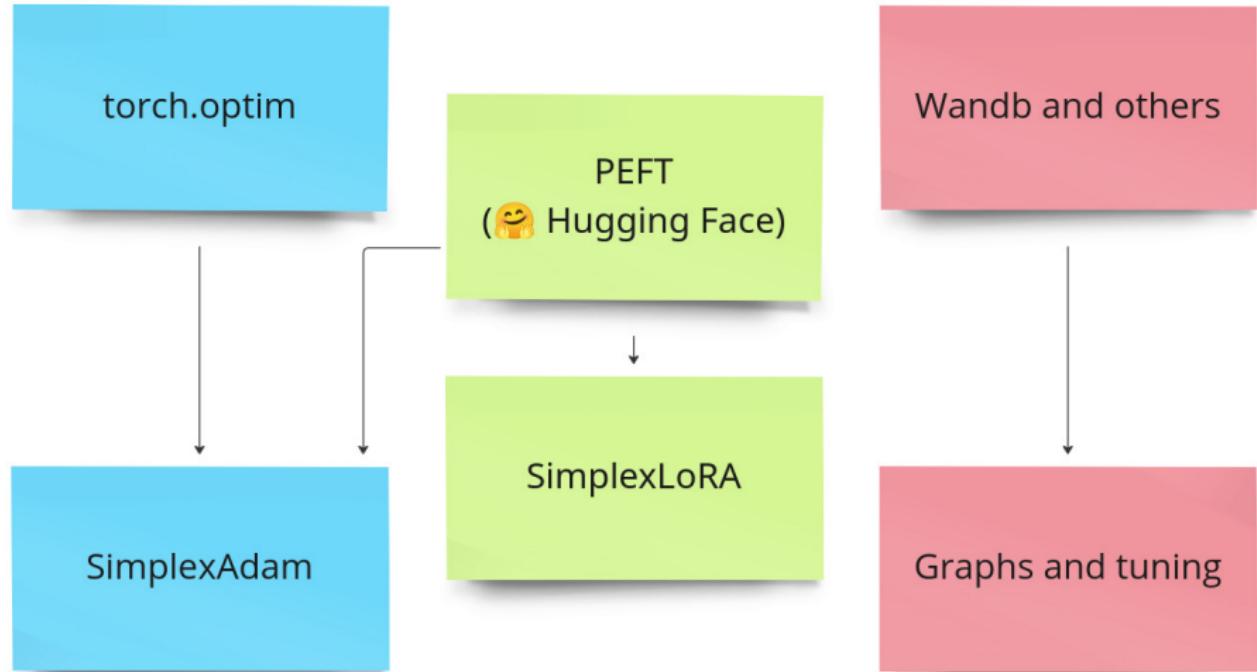
Московский физико-технический институт

6 мая 2025 г.

# Содержание доклада

- ① Напоминание: структура репозитория
- ② Оптимизатор, тьюнер, модели, бенчмарки
- ③ Эксперименты: подбор шага
- ④ Эксперименты: статья

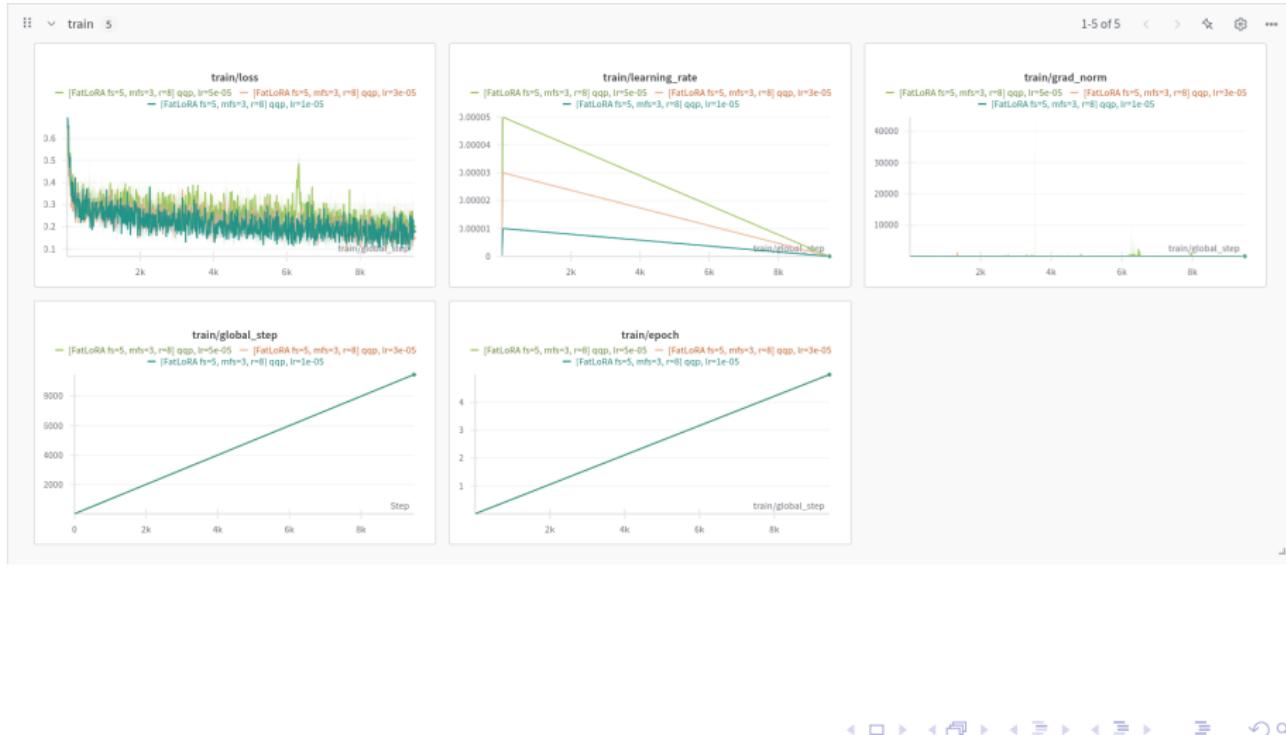
# Напоминание: структура репозитория



# Оптимизатор, тьюнер, модели, бенчмарки

- SimplexAdam
  - ① Шаг с проекцией
  - ② Сброс параметров скольжения
  - ③ Работа с методами тьюнера
- SimplexLoRA
  - ① Внедрение методов в имеющийся тьюнер
  - ② 2 вида изменения рангов
  - ③ Работа с «пустыми» адаптерами
- deberta-v3-base + GLUE
  - ① Препроцессинг датасетов: mnli, mrpc, cola и др.
  - ② Импорт модели и метрик
- Llama-3.1-8B + GLUE
  - ① Подготовка к запускам

# Эксперименты: подбор шага



# Эксперименты: подбор шага

GLUE MRPC	accuracy / f1-score			
	r=1	r=2	r=4	r=8
WLoRA [k=20] warmup_step=0	0.88725/0.91986 3e-5	0.8948/0.9174 5e-5	0.89461/0.92196 5e-5	0.89235/0.91549 1e-4
FatLoRA [mfs=2, fs=10] warmup_step=50	0.88725/0.91901 5e-5	0.8971/0.90619 8e-5	0.89661/0.92596 3e-5	0.9021/0.90815 3e-5
LoRA warmup_step=100	0.89052/0.92125 1e-3	0.89379/0.92444 1e-3	0.89424/0.92479 1e-3	0.90196/0.93004 1e-3
SimplexLoRA warmup_step=30	5e-5	3e-5	5e-5	3e-5

- Перебор lr по сетке
- Выбор наиболее результативного шага

# Эксперименты: статья

microsoft/deberta-v3-base ; # Params 185M ; GLUE Benchmark											
Method	# Params Train memory	MNLI Acc Train Time, m	SST-2 Acc Train Time, m	CoLA Mcc Train Time, m	QQP Acc / F1 Train Time, m	QNLI Acc Train Time, m	RTE Acc Train Time, m	MRPC Acc / F1 Train Time, m	STS-B Corr Train Time, m	ALL Avg Train Time, m	
LoRA (r = 8)	442368 (0,24%)	0,8797 13	0,9450 11	0,6913 10	0,8802 62	0,8437 60	0,9301 10	0,8448 18	0,8897 11	0,9220 24,38	0,8717 0,8665
WLoRA (r = 8, k=10)	122916 (0,07%)	0,8706 15	0,9473 11	0,6703 11	0,8728 27	0,8362 13	0,9167 11	0,5343 11	0,8775 11	0,9101 11	0,8897 0,8224
SimplexLoRA (r = 8) avg 3 rndm seeds		0,8984 это 20 эпох, а не 50	0,9587 это 20 эпох, а не 50	0,6621 это 20 эпох, а не 50		0,9381 это 20 эпох, а не 50	0,8099 это 20 эпох, а не 50	0,8987 это 20 эпох, а не 50	0,9289 это 20 эпох, а не 50		

- Логгирование результатов для занесения в статью
- Запуски на 3 различных сидах с усреднением
- Измерение показателей запуска: метрики, время, количество параметров