

Шаблон брифа для training-конфигурации

Заполните вместе с ML-командой, передайте в Инком интер для предварительной оценки конфигурации GPU-сервера под обучение или дообучение моделей.

Как пользоваться шаблоном. Пройдите пять блоков ниже, отметьте применимые варианты и заполните поля своими данными. Что пока неизвестно — оставьте пустым, обсудим на этапе оценки. Готовый бриф даёт базовое ТЗ, из которого получается конфигурация под конкретный сценарий, а не под усреднённый проект.

1 Задача обучения

СТРАТЕГИЯ

- ☐ Обучение с нуля (from scratch) ☐ Full fine-tuning (полная переобучение)
☐ LoRA / parameter-efficient fine-tuning ☐ Continuous learning (регулярное дообучение)
☐ Другое (укажите ниже)

КЛАСС МОДЕЛИ ПО ЧИСЛУ ПАРАМЕТРОВ

- ☐ Малая (сотни млн – единицы млрд) ☐ Средняя (десятки млрд) ☐ Большая (сотни млрд и более)
☐ Не определено

ТИП ДАННЫХ

- ☐ Текст ☐ Изображения ☐ Видео ☐ Мультимодальные (текст + картинки / видео) ☐ Другое

СВОБОДНОЕ ОПИСАНИЕ ЗАДАЧИ И ЕЁ КОНТЕКСТА В БИЗНЕСЕ

2 Датасет и цикл обучения

ОРИЕНТИРОВОЧНЫЙ ОБЪЁМ ДАТАСЕТА (ГБ / ТБ)

ДЛИНА КОНТЕКСТА / РАЗМЕР ВХОДНОГО ПРИМЕРА

ЦЕЛЕВОЕ ВРЕМЯ ОДНОГО ЦИКЛА ОБУЧЕНИЯ

☐ Часы ☐ Дни ☐ Недели ☐ Не критично

ПЛАНИРУЕМАЯ ЧАСТОТА ЦИКЛОВ

☐ Разовое обучение ☐ Периодически (раз в месяц/квартал) ☐ Регулярно (continuous)
☐ Не определено

3 Требования к GPU и памяти

ЦЕЛЕВАЯ МОДЕЛЬ / ПОКОЛЕНИЕ GPU (ЕСЛИ ЕСТЬ ПРЕДПОЧТЕНИЕ)

Например: NVIDIA H100 / H200 / B200, AMD MI300X. Или оставьте пустым — подберём.

ПЛАНИРУЕМОЕ КОЛИЧЕСТВО GPU НА ОДИН СЕРВЕР

☐ 1–2 ☐ 4 ☐ 8 ☐ Больше 8 (multi-node) ☐ Не определено

ПЛАНИРУЕМАЯ MULTI-NODE КОНФИГУРАЦИЯ

☐ Один сервер ☐ Два-три сервера ☐ Кластер (5+ узлов) ☐ Не определено

ТРЕБОВАНИЯ К VRAM (ЕСЛИ ИЗВЕСТНЫ)

Например: 80 Гб на карту, суммарно на кластер.

4 CPU, RAM, storage и интерконнект

ТРЕБОВАНИЯ К CPU (ЯДРА / ПОКОЛЕНИЕ)

ТРЕБОВАНИЯ К RAM СЕРВЕРА

STORAGE ПОД ДАТАСЕТ И ЧЕКПОИНТЫ

☐ NVMe (высокоскоростной) ☐ SAS/SATA SSD ☐ Сетевой storage (parallel FS) ☐ Не определено

ТРЕБОВАНИЯ К ИНТЕРКОННЕКТУ МЕЖДУ GPU И МЕЖДУ СЕРВЕРАМИ

☐ NVLink (внутрихостовой) ☐ InfiniBand ☐ Ethernet 100/200/400G ☐ Не определено

5 Инженерка, бюджет, сроки

УСЛОВИЯ РАЗМЕЩЕНИЯ

☐ Собственная серверная / стойка ☐ Партнёрский ЦОД (colocation) ☐ Не определено

ДОСТУПНАЯ МОЩНОСТЬ НА СТОЙКУ

☐ До 10 кВт ☐ 10–20 кВт ☐ 20–40 кВт ☐ Больше 40 кВт ☐ Не определено

ТИП ОХЛАЖДЕНИЯ

☐ Воздушное ☐ Rear Door Heat Exchanger ☐ Direct-to-chip liquid ☐ Immersion ☐ Не определено

ОРИЕНТИРОВОЧНЫЙ БЮДЖЕТ ПРОЕКТА

ЦЕЛЕВЫЕ СРОКИ ПОСТАВКИ И ЗАПУСКА

ДОПОЛНИТЕЛЬНЫЕ КОММЕНТАРИИ

Готовый бриф отправьте в Инком интер

minsk@incom.world · +375 (17) 304-35-45

Инком интер

Инком интер · incom.by · Минск, ул. Тимирязева, 72