

ОПИСАНИЕ И УСЛОВИЯ ПРЕДОСТАВЛЕНИЯ УСЛУГИ «EVOLUTION ML FINETUNING»

1. ПРЕДМЕТ СОГЛАШЕНИЯ

- 1.1. Evolution ML Finetuning (далее – Услуга) – это облачное решение, которое предоставляет возможность дообучить ML-модель Машинного обучения, размещенную вне Облака Cloud.ru, посредством API и пользовательского интерфейса в Личном кабинете.
- 1.2. Услуга реализована на оборудовании, принадлежащем Исполнителю, и средствами системы виртуализации собственной разработки (в т. ч. на базе компонентов с открытым исходным кодом). Ресурсом Услуги является среда определенной конфигурации (Инстанс-тип¹) с доступом к GPU, которая предоставляется на время дообучения модели Машинного обучения. На этих Ресурсах происходит процесс дополнительного обучения, в результате которого сохраняется дообученная модель.
- 1.3. Функциональные возможности:
- дообучение модели Машинного обучения (например, Большой языковой моделям (БЯМ)) на данных пользователя, размещенных вне Облака Cloud.ru, для получения новой копии модели или Loga-адаптера².
- 1.4. Услуга состоит из следующих компонентов:
- сервис-контроллер для управления Услугой из Личного кабинета
 - компонент, отвечающий за сбор тарификационных данных
- 1.5. Услуга оказывается Заказчику с учётом следующих ограничений:
- 1.5.1. Ограничений, описанных в разделе 5 Приложения;
- 1.5.2. Ограничений, установленных правообладателем (разработчиком) БЯМ в лицензионном соглашении³. Запуская процесс дообучения Пользователь принимает установленное правообладателем лицензионное соглашение, а Исполнитель не несет ответственности за его нарушение.

2. БАЗОВАЯ ФУНКЦИОНАЛЬНОСТЬ И ПАРАМЕТРЫ РЕСУРСОВ УСЛУГИ

- 2.1. Параметры предоставляемых Ресурсов Услуги:

Таблица 1. Параметры предоставляемых Ресурсов Услуги

Услуга	Тарифицируемые Ресурсы	Метрика	Кратность
Дообучение модели Машинного обучения на вычислительных ресурсах	Время дообучения модели Машинного обучения на 1xH100 80 Gb	Время работы (часы)	1 минута

3. ТАРИФИКАЦИЯ УСЛУГИ

- 3.1. Для данной Услуги используется Динамическая тарификация (Pay as you go).
- 3.2. Динамическая тарификация предполагает оплату Ресурсов (см. п. 2.1. Приложения) по факту их потребления Заказчиков в течение Отчетного периода.
- 3.3. Окончательная стоимость Услуги в Отчётном периоде формируется в соответствии с тарифами, установленными в Приложении №7.EVO.25.
- 3.4. Специфика тарификации Услуги: Заказчик начинает платить за обучение модели после перехода ее в статус «Запущено» в ЛК, плата начисляется за время потребления соответствующего Инстанс-типа.

¹ Инстанс-тип — предопределённый набор вычислительных ресурсов, включая количество виртуальных процессорных ядер, объем оперативной памяти, параметры дискового пространства и пропускную способность сети, предоставляемых Заказчику в рамках облачного сервиса.

² LoRA (Low-Rank Adaptation) — это метод дообучения больших нейронных сетей, при котором вместо изменения всех параметров модели обучаются лишь небольшая часть модели. Loga-адаптер — это небольшая добавочная подсеть, которые при суммировании с оригинальными весами модели реализуют «тонкую» настройку без изменения основных весов модели.

³ Каждая конкретная БЯМ, используемая для оказания Услуги, имеет в своём описании в ЛК текст лицензионного соглашения правообладателя (разработчика), ограничения которого распространяются на Заказчика.

3.5. Величина ежемесячного платежа за пользование Услугой определяется в соответствии с фактическим потреблением вычислительных ресурсов. Доступные Ресурсы и методика расчета перечислены в примере ниже:

3.6. Формула расчета

$$\text{Стоимость} = A * \frac{t}{60}$$

Где:

A - стоимость Инстанса — фиксированная стоимость выбранной конфигурации вычислительных ресурсов (инстанс-типа), указанной в Тарифах Услуги;

t - время в минутах — продолжительность процесса дообучения в минутах.

Пример расчета: (Стоимость инстанса: 840 ₽/час; Продолжительность процесса дообучения – 20 минут):

$$\text{Стоимость} = 840 * \frac{20}{60} = 280 \text{ рублей}$$

4. ИНЫЕ УСЛОВИЯ, ПРИМЕНИМЫЕ К УСЛУГЕ

- 4.1. Возможные виды подключения / изменения / отключения Услуги:
- посредством совершения действий в Личном кабинете.
- 4.2. Возможный порядок расчётов по Услуге:
- предоплата⁴;
 - постоплата⁵;
- 4.3. Возможные способы оплаты / порядок пополнения баланса:
- в безналичном порядке на основании выставленного Исполнителем счёта⁶;
 - оплата посредством электронных средств платежа⁷.
- 4.4. Требования к инфраструктуре Заказчика:
- наличие доступа в Интернет.
- 4.5. Исполнитель не сохраняет копии датасета, исходной модели, дообученной модели после оказания Услуги.

5. ОСОБЕННОСТИ УРОВНЯ ПРЕДОСТАВЛЕНИЯ УСЛУГИ

- 5.1. В соответствии с пп. 1.1.4. Приложения №2.EVO.0. к Договору вносятся следующие уточнения в уровень предоставления Услуги, действующий в отношении услуг Платформы Cloud.ru Evolution по умолчанию:
- 5.2. Для Услуги устанавливаются следующие особенности определения уровня Доступности:
- К Доступности Услуги не относится доступ к конкретной модели в хранилище моделей и датасету в хранилище датасетов в конкретный момент времени.
 - К Доступности Услуги не относятся проблемы с обучением, связанные с некорректным набором параметров дообучения заданным пользователем при запуске.
 - К Доступности Услуги не относятся проблемы связанные с некорректным формата датасета.
- 5.2.2. Недоступностью Услуги является ситуация, при которой обучение модели было прервано по причинам, зависящим от Cloud.ru;
- 5.2.3. Компенсация выплачивается пропорционально объёму недоступных Ресурсов Услуги, т.е. Компенсация за нарушение целевых показателей Доступности Услуги рассчитывается согласно количеству недоступных Ресурсов.
- 5.3. Во всём остальном в части уровня предоставления Услуги применимы положения Приложения №2.EVO.0. к Договору.
- 5.4. Заказчик понимает и соглашается с тем, что оказание Услуги осуществляется с использованием технологий искусственного интеллекта на основании информации, изображений и т.п. объектов (далее – Объекты), загружаемых Заказчиком в целях получения Услуги. Указанное означает исключение возможности какого-либо контроля в части проведения проверки таких Объектов и полученных на основании их использования результатов (далее – Результаты) со стороны Исполнителя, в связи с чем Заказчик обязуется в процессе оказания Услуги, помимо общих положений настоящего Приложения:
- не использовать (не загружать, не создавать, не распространять) Объекты, нарушающие требования законодательства Российской Федерации, права третьих лиц, общепринятые нормы морали и этики, а также влекущие нарушение указанных требований, прав и норм;
 - самостоятельно оценивать риски использования того или иного Объекта до начала пользования Услугой;
 - проверять соответствие Объекта законодательству, нормам морали и этики;

4 Является способом по умолчанию для Заказчиков-физических лиц

5 Является способом по умолчанию для Заказчиков-юридических лиц и ИП

6 Является способом по умолчанию для Заказчиков-юридических лиц и ИП

7 Является способом по умолчанию для Заказчиков-физических лиц

- соблюдать авторские права, а также права правообладателей Объекта;
- нести ответственность перед контролирующими органами, авторами, правообладателями, иными лицами за использование Объектов и Результатов, в т.ч. хранение, распространение, воспроизведение и т.п. действия;
- сопровождать Результаты информацией о создании с помощью Услуги.

- 5.5. Ни при каких условиях Исполнитель не несет ответственности за действия Заказчика, совершенные им в рамках указанной Услуги и не соответствующие либо противоречащие настоящему Приложению.
- 5.6. Исполнитель не несет ответственности за точность, актуальность или полноту Результатов. Все данные и выводы, полученные с помощью Услуги, предназначены исключительно для информационных целей и не должны рассматриваться как окончательные или полностью надежные.