ПРИЛОЖЕНИЕ № 1.10.
к Договору

ОПИСАНИЕ И УСЛОВИЯ ПРЕДОСТАВЛЕНИЯ УСЛУГИ «ML Space»

1. НАИМЕНОВАНИЕ УСЛУГИ

1.1. Наименование Услуги: «ML Space».

1.2. Настоящий документ содержит описание состава Услуги, ее базовой функциональности, возможных сопутствующих и дополнительных услуг, общего порядка подключения, изменения и отключения Услуги, условий предоставления и ограничений.

2. ИНФОРМАЦИЯ ОБ УСЛУГЕ

Услуга предоставляет собой доступ к платформе ML Space, которая обеспечивает полный цикл ML-разработки и совместную работу команд Data Scientist.

Состоит из следующих компонентов - сопутствующих услуг:

- ML Space Deployments
- ML Space Environments
- ML Space Data Catalog

2.1.1. Краткое описание услуги ML Space Deployments

Представляет собой веб-услугу для эффективной сборки Docker-образов на базе моделей Машинного и Глубокого обучения, а также для их дальнейшего разворачивания на ресурсах SberCloud в виде микросервисов со сгенерированным API.

Для предоставления Услуги Заказчику необходимым условием является наличие на его площадке подключения к сети интернет, достаточного для эффективной загрузки данных, моделей или их производных (например, чекпоинтов моделей или сериализованных моделей) на сервер.

С помощью Услуги Заказчик может развертывать модели искусственного интеллекта на базе инфраструктуры SberCloud и Услуги ML Space - Deployments для дальнейшего внедрения их в функции, бизнес-процессы или микросервисы.

Заказчику для успешной реализации вывода моделей искусственного интеллекта в виде микросервисов предоставляется возможность сборки образа с любым Программным обеспечением, python-библиотеками и способом взаимодействия с моделями искусственного интеллекта.

На Рисунке 1 приведена общая упрощенная схема взаимодействия с Услугами ML Space - Deployments с удаленной площадки Заказчика (с указанием зон ответственности):

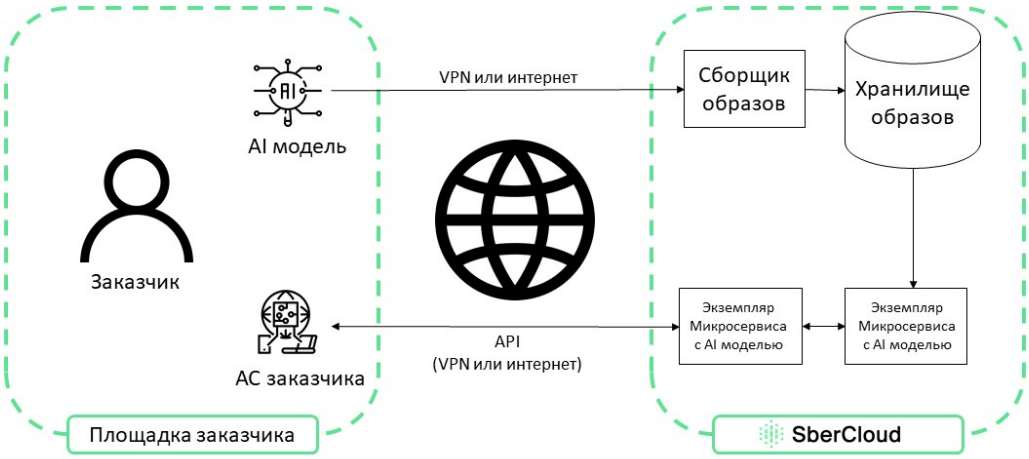

Рисунок 1: Схема взаимодействия Заказчика с Услугами разворачивания моделей машинного и глубокого обучения на мощностях кластера ML Space - Deployments.

В зоне ответственности SberCloud - функционирование Серверов с развернутыми моделями искусственного интеллекта, функционирование вычислительного кластера, Объектного хранилища S3, а также прочей инфраструктуры.
В рамках Услуги Заказчик может самостоятельно отслеживать и управлять состоянием развернутых моделей.

Для подключения к Услуге Заказчик может выбрать один или несколько типов подключения:

- Подключение через общий канал Интернет (shared) предполагает логическое подключение к общему для всех Заказчиков Услуги каналу передачи данных. Скорость сетевого соединения для каждого Заказчика не является гарантированной и зависит от загруженности общего канала передачи данных (Услуга предоставляется по умолчанию).
- Подключение через прямой канал связи. Данный способ подключения позволяет обеспечить взаимодействие сетей Заказчика с сетью в облаке с помощью выделенных каналов связи стороннего провайдера. Опционально, с помощью данного сценария, к Услуге Заказчика может быть подключен альтернативный канал в сеть Интернет. Для данного подключения могут быть использованы выделенные каналы Заказчика, организованные с использованием «темной оптики» (Услуга оплачивается отдельно).

Услуга предоставляется на базе защищенной инфраструктуры облачной платформы «ML Space», меры защиты которой приведены в описании Услуги ML Space - Environments.

2.1.2. Условия хранения данных в Объектном хранилище S3 (ML Space – Deployments)

Хранение, использование и тарификация хранения и использования данных в Объектном хранилище S3 осуществляется в рамках соответствующей Услуги по Договору.

2.1.3. Использование услуги ML Space - Deployments

Создание, конфигурация и разворачивание моделей искусственного интеллекта осуществляется напрямую Заказчиком.

2.1.4. Техническое описание решения ML Space - Deployments

2.1.4.1. Программная платформа

Услуга реализуется средствами веб интерфейса, внутреннего docker registry, сборщиком образов и комплексом KFServing/Knative/ISTIO/Kubernetes. Посредством них и программных библиотек пользователь имеет возможность собирать и разворачивать модели искусственного интеллекта в виде микросервисов.

2.1.4.2. Аппаратная платформа

Вычисления и обсчет задач осуществляется на предоставляемой Заказчику в рамках Услуги области кластера Кристофари, а также других вычислительных ресурсов и серверов SberCloud.

2.1.4.3 Технические особенности и ограничения:

Скорость загрузки данных на площадку Исполнителя ограничена пропускной способностью канала доступа в Интернет из инфраструктуры Заказчика до облака SberCloud, а также скоростью чтения данных с СХД Исполнителя.

Общие значения параметров Услуги «ML Space - Deployments»

Описание	Мин. значение	Макс. значение
Количество утилизируемых в рамках вычисления задачи GPU-секунд на кластере Christofari	1 GPU-секунда NVIDIA Tesla V100 в конфигурации DGX-2	В соответствии с количеством свободных GPU на кластере Christofari
Количество утилизируемых в рамках вычисления задачи GPU-секунд	1 GPU-секунда	В соответствии с количеством свободных GPU
Количество утилизируемых в рамках вычисления задачи CPU-секунд	1 CPU-секунда	В соответствии с количеством свободных CPU

2.2.1 Краткое описание услуги ML Space Environments

Услуга предоставляет среду разработки Jupyter Notebook, набор инструментов для хранения данных в Объектном хранилище S3, набор инструментов для предобработки данных, а также набор инструментов и библиотек для запуска задач по исполнению кода обучения моделей Машинного и глубокого обучения на ресурсах суперкомпьютера Кристофари (а также на прочих ресурсах, доступных в SberCloud – по усмотрению Заказчика) и мониторинга процесса обучения.

Для оказания Услуги Заказчику необходимым условием является наличие на его площадке подключения к сети интернет, достаточного для эффективной загрузки данных на сервер, а также наличия собственных данных для обучения модели.

С помощью Услуги Заказчик может вести разработку моделей и производить ускоренную подготовку данных и обучение моделей на больших объемах данных, благодаря мощностям суперкомпьютера и высокопроизводительным графическим ускорителям.

Заказчику для успешной реализации задачи обучения моделей на больших объемах данных предоставляется возможность загрузки и хранения данных в Объектное хранилище S3, а также возможность подключения к этому хранилищу как из Jupyter Notebook'а, так и из кластера, на котором будет вычисляться задача обучения модели.

На Рисунке 1 приведена общая упрощенная схема взаимодействия с Услугами ML Space Environments с удаленной площадки Заказчика (с указанием зон ответственности):

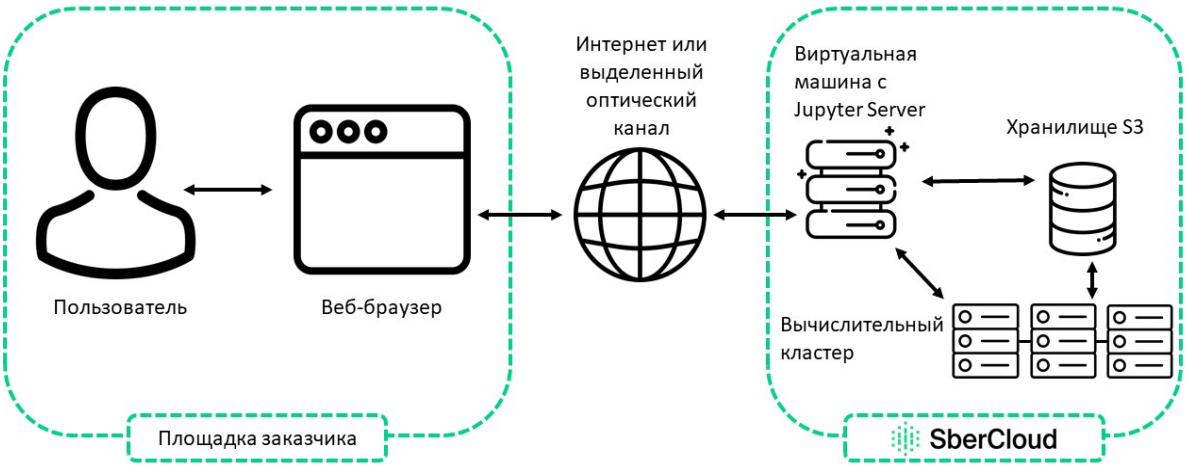

Рисунок 1: Схема взаимодействия Заказчика с Услугами обучения моделей машинного и глубокого обучения ML Space - Environments.

В зоне ответственности SberCloud - функционирование Серверов с развернутым Jupyter Server, функционирование вычислительного кластера и Объектного хранилища S3.

В рамках Услуги Заказчик может самостоятельно отслеживать состояние заданий обучения модели.

Для подключения к Услуге Заказчик может выбрать один или несколько типов подключения:

- Подключение через общий канал Интернет (shared) предполагает логическое подключение к общему для всех Заказчиков Услуги каналу передачи данных. Скорость сетевого соединения для каждого Заказчика не является гарантированной и зависит от загруженности общего канала передачи данных (Услуга предоставляется по умолчанию).
- Подключение через прямой канал связи. Данный способ подключения позволяет обеспечить взаимодействие сетей Заказчика с сетью в облаке с помощью выделенных каналов связи стороннего провайдера. Опционально, с помощью данного сценария, к Услуге Заказчика может быть подключен альтернативный канал в сеть Интернет. Для данного подключения могут быть использованы выделенные каналы Заказчика, организованные с использованием «темной оптики» (Услуга оплачивается отдельно).

При подключении через общий канал Интернет Заказчику предоставляется базовая защита информационных систем, размещаемых в инфраструктуре облачной платформы SberCloud, от DDoS-атак на канальном уровне.

2.2.2. Условия хранения данных в Объектном хранилище S3 (ML Space – Environments)

Хранение, использование и тарификация хранения и использования данных в Объектном хранилище S3 осуществляется в рамках соответствующей Услуги по Договору (Приложение № 1.8. к Договору).
Условия использования Заказчиком Объектного хранилища S3 для цели потребления Услуги:

- Заказчику для потребления Услуги ML Space - Environments предоставляется доступ к Объектному хранилищу S3 в размере, необходимом для хранения данных обучаемой модели.
- Для того, чтобы воспользоваться услугой доступа к Объектному хранилищу S3 Заказчику необходимо в Личном кабинете активировать доступ к ней посредством проставления «галочки» в соответствующем поле «Требуется S3 хранилище».
- Объём потреблённой в течение Отчётного периода Услуги Объектное хранилище S3 рассчитывается Исполнителем в соответствии с данными АСИ.
- Стоимость Услуги Объектное хранилище S3 определена в Приложении № 7 к Договору.
- Оплата Услуги Объектное хранилище S3 осуществляется Заказчиком в порядке постоплаты на основании выставленного Исполнителем счёта и при условии подписанного Сторонами Акта. Акт и счёт выставляются в порядке, установленном ст. 4 Договора.

2.2.3. Использование Услуги ML Space - Environments

Создание, конфигурация и запуск задач на обучение осуществляется Заказчиком через пользовательский интерфейс Услуги ML Space - Environments.

2.2.4. Техническое описание Услуги ML Space – Environments

2.2.4.1. Программная платформа

Услуга реализуется средствами Jupyter Server и Jupyter Notebook. Посредством него и программных библиотек пользователь имеет возможность запускать задачи на вычисление на кластере.

2.2.4.2. Аппаратная платформа

Вычисления и обсчет задач осуществляется на предоставляемой Заказчику в рамках Услуги области кластера Кристофари (а также на прочих ресурсах, доступных в SberCloud – по усмотрению Заказчика).

2.2.4.3 Технические особенности и ограничения

Скорость загрузки данных на площадку Исполнителя ограничена пропускной способностью канала доступа в Интернет из инфраструктуры Заказчика до облака SberCloud, а также скоростью чтения данных с СХД Исполнителя.

Описание	Мин. значение	Макс.значение
Количество утилизируемых в рамках вычисления задачи GPU	1 GPU	В соответствии с количеством доступных GPU на момент запуска задачи
Количество утилизируемых в рамках вычисления задачи CPU	1 CPU	В соответствии с количеством доступных CPU на момент запуска задачи

2.3.1. Краткое описание Услуги ML Space Data catalog

Услуга предоставляет возможность совместной работы, хранения, версионирования артефактов для машинного обучения (датасетов, моделей, докер-образов, кода и др.), а также доступ к Data marketplace.

Data catalog включает в себя:
- Файловый менеджер на основе Объектного хранилища (S3) с управлением правами доступа пользователей, тегированием файлов, версионированием и архивацией файлов;
- Data transfer service - коннекторы к файловым системам (HDFS, S3 (Amazon, Google Cloud Storage) и базам данных (PostgreSQL, MySQL, MS SQL, Hive, BigQuery и правила переноса, а также Data transfer между S3 и NFS с историей переносов;
- Docker registry - загрузка, организация, запуск, остановка, хранение, масштаби-рование, и иные способы управления контейнерами;
- Data marketplace – маркетплейс артефактов машинного обучения (датасетов, моделей, контейнеров, скриптов, пайплайнов и др.).

Основным хранилищем для данных Data catalog является Объектное хранилище (S3), подробное см. п. 2.2.2. Для использования данных Data catalog в Environments и Deployments, необходимые для использования функционала Услуги данные перекладываются на быстрое хранилище NFS.

Для оказания Услуги Заказчику необходимым условием является наличие на его площадке подключения к сети интернет, достаточного для эффективной загрузки данных на сервер, а также наличия собственных данных.

Для подключения к Услуге Заказчик может выбрать один или несколько типов подключения:

- Подключение через общий канал Интернет (shared) предполагает логическое подключение к общему для всех Заказчиков Услуги каналу передачи данных. Скорость сетевого соединения для каждого Заказчика не является гарантированной и зависит от загруженности общего канала передачи данных (Услуга предоставляется по умолчанию).
- Подключение через прямой канал связи. Данный способ подключения позволяет обеспечить взаимодействие сетей Заказчика с сетью в облаке с помощью выделенных каналов связи стороннего провайдера. Опционально, с помощью данного сценария, к Услуге Заказчика может быть подключен альтернативный канал в сеть Интернет. Для данного подключения могут быть использованы выделенные каналы Заказчика, организованные с использованием «темной оптики» (Услуга оплачивается отдельно).

Услуга предоставляется на базе защищенной инфраструктуры облачной платформы «ML Space», меры защиты которой приведены в описании Услуги ML Space - Environments.

2.2.5. Использование Услуги ML Space – Data catalog

Работа Услугами осуществляется Заказчиком через пользовательский интерфейс Услуги ML Space – Data catalog.

2.2.6. Техническое описание решения ML Space – Data catalog

2.2.4.1. Программная платформа

Услуга реализуется посредством файлового менеджера S3, перекладчика данных с S3 на NFS, Data transfer service и Docker registry. Посредством данных модулей реализуется возможность совместной работы, хранения,

версионирования артефактов для машинного обучения (датасетов, моделей, докер-образов, кода и др.), а также Data marketplace.

2.2.4.2. Аппаратная платформа

Данные из Data catalog используются в Услугих Environments и Deployments. Вычисления и обсчет задач осуществляется на предоставляемой Заказчику в рамках Услуги области кластера Кристофари (а также на прочих ресурсах, доступных в SberCloud – по усмотрению Заказчика).

2.2.4.3 Технические особенности и ограничения

Скорость загрузки данных на площадку Исполнителя ограничена пропускной способностью канала доступа в Интернет из инфраструктуры Заказчика до облака SberCloud, а также скоростью чтения данных с СХД Исполнителя.

Общие значения параметров Услуги ML Space – Data catalog

Описание	Мин. значение	Макс. значение
Количество утилизируемых GB S3	1 GB	В соответствии с количеством доступных GB на S3
Количество утилизируемых GB NFS	1 GB	В соответствии с количеством доступных GB на NFS

3. ТАРИФИКАЦИЯ УСЛУГИ

3.1.1. Возможные виды тарификации ML Space - Deployments:

- Динамическая тарификация (Pay as you go).

3.1.2. Стоимость Услуги формируется в зависимости от количества GPU/CPU, на которых происходило вычисление запросов к API микросервисов с моделями, а также самого времени, в течение которого вычислялись запросы к API микросервисов с моделями искусственного интеллекта и объема зарезервированного Заказчиком Объектного хранилища S3.

3.1.3. Момент начала списания денежных средств – с момента начала вычисления запроса к API микроуслуги (для каждого запроса).

3.1.4. Момент окончания списания денежных средств – с момента окончания вычисления запроса (для каждого запроса).

3.2.1. Возможные виды тарификации ML Space Environments:

- Динамическая тарификация (Pay as you go).

3.2.2. Стоимость Услуги формируется в зависимости от количества и конфигураций GPU/CPU, на которых происходило вычисление задачи, времени, в течение которого вычислялась задача, объема зарезервированного Заказчиком Объектного хранилища S3.

3.2.3. Момент начала списания денежных средств – с момента запуска обучения модели/с момента аллокации GPU/CPU под выбранное окружение (определяется Заказчиком через пользовательский интерфейс Услуги ML Space - Environments).

3.3.1. Возможные виды тарификации ML Space Data catalog:

- Динамическая тарификация (Pay as you go).

3.3.2. Стоимость Услуги формируется в зависимости от объема зарезервированного Заказчиком Объектного хранилища S3 и быстрого хранилища NFS (количества GB/мес).

3.3.1. Момент начала списания денежных средств – с аллокации на S3/NFS более чем 1 GB (определяется Заказчиком через пользовательский интерфейс Услуги ML Space – Data catalog).

4. ИНЫЕ УСЛОВИЯ, ПРИМЕНИМЫЕ К УСЛУГЕ

4.1. Возможные виды подключения / изменения / отключения Услуг:

4.1.1. Посредством подписания Заказа;

4.1.2. Посредством совершения действий на Портале и Консоли ML Space;

4.2. Возможный порядок расчётов по Услуге:

4.2.1. Предварительная оплата;

4.2.2. Постоплата (на основании отдельно заключенного письменного бланка Заказа).

4.3. Возможные способы оплаты / порядок пополнения баланса:

4.3.1. Оплата в безналичном порядке на основании выставленного Исполнителем счёта;

4.3.2. Оплата посредством электронных средств платежа.

4.4. В связи с характером потребления Услуги, а также объёмов, которых она может достигнуть в Отчётный период, Стороны установили, что в случае заключения с Заказчиком соглашения о применении постоплаты (пп. 4.2.2. настоящего Приложения) Заказчик выбирает лимит в пределах Отчётного периода, по достижении которого Услуги оказываются на основании соответствующего обращения уполномоченного лица в Контактный Центр и, по требованию Исполнителя, предоставления гарантийного письма.

4.5. Исполнитель обязуется не включать в состав Результатов работ программное обеспечение, используемое на основании открытой лицензии, условия которой требуют от пользователя раскрытия исходного кода модифицированного ПО, либо ограничивают право пользователя запрещать третьим лицам использование модифицированного ПО.