

ML Space

Платформа ML Space для ML-разработки
полного цикла и совместной работы
DS-команд

Материалы для встречи с НБМЗ
Декабрь 2022 года



AI

Процесс самостоятельного ML



Как выглядит верхнеуровневый процесс

Взять файл с данными (excel, csv) - Написать в Jupyter Notebook код модели
- Обучить модель (CPU/GPU) - Интегрировать к модели API для ее вызова -
Запустить модель и API (Docker, Kubernetes, просто как сервис)

Как на самом деле

Автоматизировать доставку данных из источников

Написать препроцессинг данных для обучения

Обучить модель

Разработать решение, как связать препроцессинг с обучением

Разработать процесс переобучения модели

Сервинг модели: написать собственный Фреймворк или использовать готовый

Собрать и обучить команду data scientist-ов

Организовать мониторинг

Создать хранилище данных (с поддержкой историчности)

Требования к инфраструктуре

Отказоустойчивость Автоскелинг ресурсов

Хранилище для модели, артефакты, датасеты

ML платформы сокращают time to market до 30%

➤ **Упрощенный E2E цикл ML-разработки:**
Продуктом могут пользоваться не только Data Scientists, а также аналитики, администраторы и devops инженеры.

➤ **Скорость и быстрое развертывание**
Среда совместного управления артефактами машинного обучения, экспериментами и деплоям. Отправка модели на развертывание 1-м кликом

➤ **Сокращение затрат на команду**
Не требуется дорогостоящая команда из 5-6 чел для обучения модели

➤ **Быстрое обучение и предсказание**
Auto ML - поддерживает обучение «без учителя», автоматизированная платформа по выполнению ML, не требует экспертизы DS

➤ **Единый инструмент**
для коллаборации, визуализации и версионирования. Позволяет обмениваться данными с другими пользователями и вместе выполнять эксперименты

ML Space Особенности



Совместная работа
DS команд



Полный цикл
ML-разработки



Pay as you go



+1700 GPU Tesla
V100 & A100



Быстрая генерация API
к модели и деплой любых
контейнеров



Работа через
UI / API / Pipelines



Суперкомпьютеры
и мультикластерность



Распределенное обучение



152 Ф3, У3-1
для sensitive data

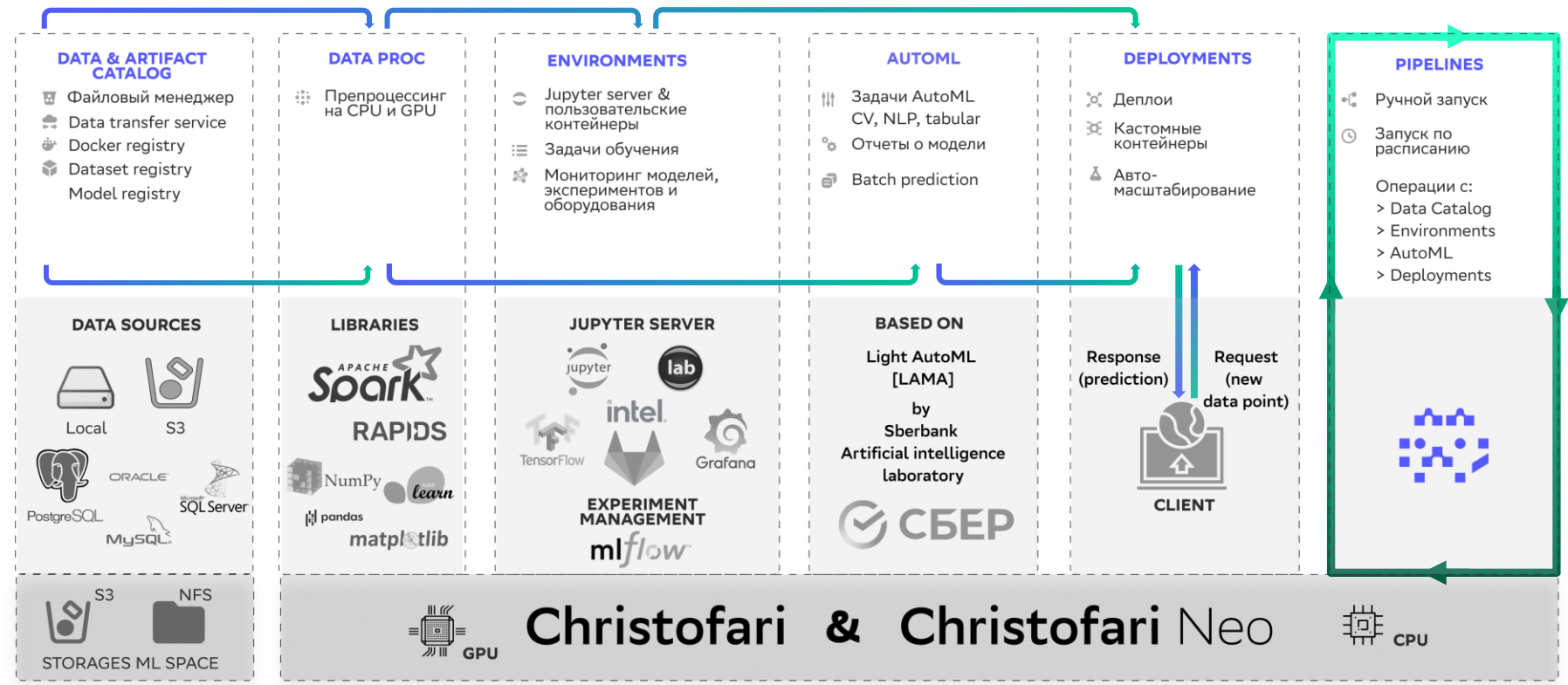


Готовые окружения с популярными
библиотеками и фреймворками



Маркетплейс датасетов, моделей,
инструментов и сервисов

ML Space Компоненты платформы



DataHub



AI Services

Мультикластерность

Модели платформы в 4-х регионах на 7 кластерах

Private Cloud

Локальная установка согласно политикам вашей компании

ML Space DataHub & AI Services

Модели

Все Текст Изображения

**ruDALL-E XL 1.3B**

Русская text-to-image модель, 1.3 млрд параметров

[Подробнее](#)

**ruBERT base**

Самая известная модель-encoder для русского языка, 178 млн параметров

[Подробнее](#)

**ruCLIP XL 216M**

Модель-ранжировщик текстов и изображений, 216 млн параметров

[Подробнее](#)

**ruGPT-3 XL**

Генерация текста на любые темы, 1.3 млрд параметров

[Подробнее](#)

**ruT5 base**

Русская Sequence2Sequence Model, 222 млн параметров

[Подробнее](#)

**ruT5 large**

Русская Sequence2Sequence Model, 737 млн параметров

[Подробнее](#)

**ruBERT large**

Самая известная модель-encoder для русского языка, 427 млн параметров

[Подробнее](#)

**ruRoBERTa large**

Русская encoder модель, 355 млн параметров

[Подробнее](#)

**ruGPT-3 large**

Генерация текста на любые темы, 760 млн параметров

[Подробнее](#)

**ruGPT-3 13B**

Генерация текста на любые темы, 13 млрд параметров

[Подробнее](#)

**ruCLIP BIG**

Будет доступно с 25-го декабря

[Подробнее](#)

**ruDALL-E XXL 12B**

Русская text-to-image модель, 12 млрд параметров

[Подробнее](#)

Каталог сервисов

Все Текст Речь Изображения

**Smart Speech**

Распознавание и синтез речи

[Подробнее](#) Бесплатно

**Demo ruGPT-3 XL**

Генерация текста на любые темы, 1.3 млрд параметров

[Подробнее](#) Бесплатно

**ruGPT-3 13B**

Генерация текста на любые темы, 13 млрд параметров

[Подробнее](#)

**Demo Болталка**

Диалоговая модель-собеседник, chit-chat на основе ruGPT-3

[Подробнее](#) Бесплатно

**Demo Рерайт-сервис**

Модель рерайтинга для текстов разной длины и доменов

[Подробнее](#) Бесплатно

**ruDALLE 12B**

Русская text-to-image модель, генерирующая изображения по тексту

[Подробнее](#)

**Demo ruDALLE**

Демо text-to-image модели, 1.3 млрд параметров

[Подробнее](#) Бесплатно

**ruCLIP Base [vit-base-patch16-384] exclusive**

Модель-ранжировщик текстов и изображений, 150 млн параметров

[Подробнее](#)

**ruCLIP Large [vit-large-patch14-336] exclusive**

Модель-ранжировщик текстов и изображений, 430 млн параметров

[Подробнее](#)

Семейство больших трансформерных моделей для русского языка:

- обработка текста, изображений,
- решения для AI-сервисов и стартапов
- готовые окружения и деплои

ML Space DataHub & AI Services | Набор артефактов для задач физической химии от AIRI

Описание

- Набор артефактов включает в себя один из крупнейших датасетов молекул и их метаданных. Кроме того, в набор входит 4 модели для предсказания энергии молекулярных конформаций и 2 модели для предсказания DFT-гамильтониана.
- Комбинация перечисленных артефактов позволяет не только создавать ML-модели для решения задач физической химии, но и оценивать их эффективность.

Возможности

- Датасет содержит результаты численного моделирования уравнения Шредингера для большого количества разнообразных молекул и их конформаций. Это позволяет разрабатывать и исследовать модели машинного обучения направленные на решение задач физической химии.
- Фокус на медицинской химии позволяет применять предобученные модели в классических пайплайнах разработки новых лекарственных препаратов.

Преимущества:

- Самый большой объем из открытых датасетов: более 1 млн. драг-лайк молекул и свыше 5 млн. конформаций.
- Широкий спектр доступных для предсказания свойств: энергия конформации, гамильтониан теории функционала плотности, матрица электронной плотности, дипольный момент, частичные заряды на атомах, HOMO-LUMO gap и др.
- В набор входят современные предобученные модели предсказания энергии конформации и электронных свойств молекул с высокой степенью точности.
- Различные типы тестовых подвыборок позволяют оценить генерализацию моделей как между различными конформациями, так и между различными хемотипами молекул.

Кейсы ML Space | **Аватар машина** | Чат-бот психологической поддержки

Описание

- С помощью ML Space и Christorafi была **впервые** обучена модель генерации текста для русского языка на основе архитектуры GPT-3 от OpenAI.
- Модель ruGPT-3 XL содержит 1,3 миллиарда параметров и доступна для скачивания и до обучения в DataHub в ML Space.
- Для целей проекта по созданию чат-бота психологической поддержки в качестве дополнительного датасета для дообучения использовали парсинг сообщений на сайтах и форумах про знакомства и отношения.

Результат

Получили: свободный и естественный разговор с чат-ботом дружеской «психологической поддержки» в мобильном приложении для людей, которые оказались в сложных жизненных ситуациях.

Преимущества: контроль интенгов и реализация личности – отдельная легкая модель-классификатор (KNN, CNN для эмбедингов) подсказывает ruGPT-3 как сгенерировать уникальный ответ без шаблонных фраз.
Достигаемый эффект – комбинация лучших «умений» двух архитектур нейросетей: вежливый, тематический разговор на узкую тему с эмуляцией личности(знаний, характера) определенного человека(персонажа).

Бизнес-эффект:

Показатель вовлеченности. Увеличение времени нахождения пользователя в мобильном приложении на 50%.

Кейсы ML Space | **Запределье** | Синтетический человек Meta Human

Описание

- С помощью ML Space и Christorafi была **впервые** обучена модель генерации текста для русского языка на основе архитектуры GPT-3 от OpenAI.
- Модель gpt-3 XL содержит 1,3 миллиарда параметров и доступна для скачивания и дообучения в DataHub в ML Space.
- Для целей проекта в качестве дополнительного датасета для дообучения использовали бизнес-проекты Фонда «Сколково».

Результат

Получили:

- ✓ VR голограмму бизнес-консультанта в фонде «Сколково» на основе 3D аватара metahuman;
- ✓ Для имитации голоса использовался Yandex speech to text;
- ✓ Проект был продемонстрирован на New Digital Life в Сочи в июне 2021 года.

Преимущества: Визуальное восприятие гораздо мощнее, чем просто чат-бот.

Планы по развитию:

- ✓ проецирование базовых 7 эмоций человека;
- ✓ имитация невербальных жестов, макро и микро мимики;
- ✓ обучение движению на основе костюмов с датчиками для motion capture.

Перспективы применения:

- ✓ замена операторов колл-центров,
- ✓ виртуальных бариста.

Кейсы ML Space | СберЗдоровье | Распознавание КТ снимков

Описание

Сервис СберЗдоровье и облачная платформа SberCloud, входящие в экосистему Сбербанка, запустили совместный проект по **распознаванию снимков компьютерной томографии легких**.

Любой человек может загрузить на сайт свой снимок КТ и получить его расшифровку в течение 5-10 минут на почту.



Результат

Система использует алгоритмы искусственного интеллекта, разработанные в Лаборатории по искусственному интеллекту Сбербанка и работает на вычислительных мощностях суперкомпьютера «Кристофари», компании SberCloud.

На сегодняшний день она умеет показывать в процентах пострадавшую часть легких, **детектировать** пораженные участки легких (матовое стекло и консолидация), и **рассчитывать их общий объем**. Размеченные DICOM-серии файлов можно скачать и посмотреть в любой программе для просмотра DICOM снимков.

- Такой анализ позволяет пользователю самостоятельно **оценить динамику** при повторном исследовании КТ, чтобы следить за развитием заболевания, не выходя из дома.
- Врачам сервис поможет быстро и просто **расшифровывать** снимки пациентов. На основе результатов автоматического анализа возможно реализовать приоритизацию рабочего списка рентгенолога и категоризировать исследования.

Кейсы ML Space | ВЕСТТРЭЙД | Эмбриология

Описание

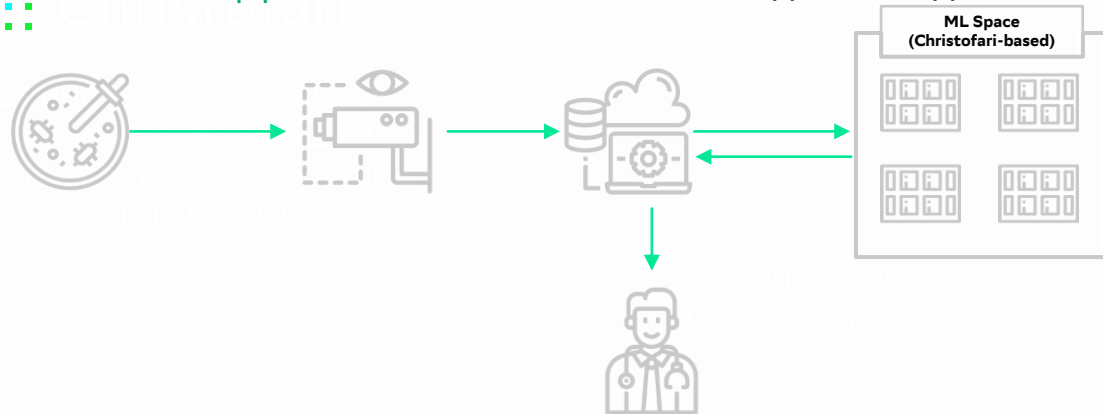
- В ходе ЭКО оплодотворение и рост эмбрионов происходит в искусственных условиях с использованием специальных чашек
- Этапы развития эмбрионов фиксируются специальными камерами с малой периодичностью
- Все снимки специалисту-эмбриологу нужно изучать вручную

А также

- Большой объем снимков влечет большие трудозатраты
- Возможна человеческая ошибка

Результат

- Модель компьютерного зрения **автоматически** изучает все снимки
- Нейронная сеть **эффективно и быстро** обучается и дообучается на суперкомпьютере Christofari на сервисе ML Space Environments
- Эмбриологу поступают **только требующие внимания** снимки, **снижая рутинный труд** и риск человеческой ошибки.
- **Повышается эффективность лечения** бесплодия методами ЭКО





Тестовый доступ к ML Space до 30 дней *



**Иван
Тараскин**

Руководитель направления по развитию
бизнеса в области ИИ Cloud

Моб. тел.: +7 (916) 315-62-20
Офис: +7 (495) 260-10-81
E-mail: iataraskin@sbercloud.ru
Адрес: Деловой дом на Моховой –
Москва, ул. Воздвиженка, д. 4/7, стр. 1