

**Разработка открытого ПО проведения
воспроизводимых экспериментов при обучении
наукам о данных**

**рук. проекта Хританков А.С.
кафедра АТП МФТИ**

Краткое описание проекта

Решается актуальная задача повышения конкурентоспособности МФТИ в области наук о данных. А именно, в части воспроизводимости экспериментов и командной работы. Проблема воспроизводимости и доверия к результатам – одна из важнейших в современном состоянии исследований в области и индустрии.

Проект реализуется совместной командой студентов ФПМИ и преподавателей кафедры ATP ФПМИ с опытом выполнения сходных проектов. Будет разработана программная система с открытым исходным кодом для проведения повторяемых экспериментов, обучения студентов наукам о данных и организации совместной работы исследователей.

Проект поддержан на кафедрах Алгоритмов и технологий программирования, Интеллектуальных систем, а также на уровне школы ФПМИ в целом.

Запрашиваемое финансирование – 395 тыс. руб., срок реализации – 3 мес.

Команда проекта



Антон Хританков, к.ф.-м.н.
R&D, ML, Agile, SoftEng
руководитель проекта
доцент кафедры АТП



Олег Ивченко, асп.
BigData, DevOps, Linux
тех. обеспечение
документооборот



Виктор Яковлев, к.ф.-м.н.
R&D, Bioinformatics
учет расходов по ФЛС
зав. кафедры. АТП

Команда проекта

— — —



Артем Ухов

МФТИ 5 курс

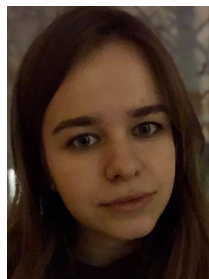
Python, Java, Cloud
ведущий разработчик



Никита Ухов

МФТИ 5 курс

DevOps, Java, Python
Ведущий разработчик



Ирина Кудряшова

МФТИ 5 курс

Frontend, Linux, Scala
веб-разработчик



Андрей Беликов

МФТИ 5 курс

Python, Linux
разработчик

Предпосылки выполнения проекта

Развитие наук о данных и вовлечение широкого круга начинающих и специалистов


Воспроизводимость и повторяемость эксперимента служат способом повышения доверия к результатам исследований

Применение достижений наук о данных в индустрии требует уверенности в результатах прикладных исследований

Большинство исследований выполняется в исследовательских коллективах

Актуальность

Базовая нерешенная проблема воспроизводимости и верификации результатов в области наук о данных

- ожидания и рекомендации конференций NIPS/NeurIPS по обеспечению повторяемости
 - инициативы paperswithcode.com, codalab.org и др. в области систематизации и повторения исследований
 - появление интернет сообществ по проблеме, например, REPA
 - возникают решения от индустрии MLOps от Google Cloud, AIOps от Microsoft Azure, REP/REANNA от Yandex/CERN
- 

Решаемая проблема

МФТИ не представлен в области воспроизводимых исследований в науках о данных

Нет подходящего решения для воспроизводимых экспериментов для учебы и исследований в команде



Ожидаемые результаты проекта в целом

— — —

- Реализована **программная система с открытым исходным кодом** для проведения повторяемых экспериментов в области наук о данных
- Сформировано **сообщество пользователей и разработчиков**, развивающих систему
- **Система использована на кафедрах МФТИ** и за пределами МФТИ. Представлены доклады по результатам применения.
- **Представлен бренд МФТИ и ФЦК МФТИ** в области повторяемых исследований в науках о данных
- **Вовлечены студенты и магистранты** в практику воспроизводимых исследований

План реализации

— — —

Май –
Сентябрь
2020

Этап 1. Реализация прототипа для привлечения первичных пользователей среди кафедр

- Реализация первой версии системы на основе имеющегося опыта
- Проведение презентаций и встречи с кафедрами МФТИ
- Разработка сайта проекта, упоминание поддержки ФЦК МФТИ

Сентябрь –
Ноябрь
2020

Этап 2. Создание сообщества вокруг открытого проекта

- Проведение хакатона для доработки системы с указанием ФЦК
- Определение процессов управления и развития open source проектом

Октябрь –
Февраль
2021

Этап 3. PR и продвижение проекта

- Выступление на отраслевых конференциях, участие в семинарах Moscow Python Meetup, REPA, OSS Conf., AltLinux-OSS и др.
- Публикация заметок, статей на образовательных и отраслевых ресурсах с указанием ФЦК МФТИ

Ожидаемые результаты. Этап 1.

— — —

- Разработана первая версия MLDEV для решения задач:
 - Проведение повторяемого эксперимента
 - Подготовка эксперимента для научной статьи, курсовой работы
 - Автоматизирован расчет эксперимента в вычислительном облаке общего доступа
- Создание проекта с открытым исходным кодом на ее основе
 - Исходный код, документация доступны на одном из популярных хостингов для проектов с открытым кодом для использования
 - Сайт с описанием и документацией проекта доступен в Интернет
- Уточнение постановок задач для второго этапа (хакатон)
 - Во взаимодействии с заинтересованными сторонами проекта уточнены требования по применению в лабораториях и на кафедрах МФТИ
 - Подготовлены задачи для решения на этапе 2

Запрашиваемое финансирование. Этап 1

— — —

МФТИ, кафедра АТП ФПИ

помещения, содействие подразделений института и участие сотрудников поддержавших кафедр во внедрении – в рамках текущего финансирования МФТИ

Участники проекта

собственное аппаратное и программное обеспечение, аккаунты в интернет ресурсах

Физтех-школа ФПИ

40 тыс. руб.

ФЦК МФТИ – запрашиваемое финансирование – Этап 1

355 тыс руб.

Реализация этапов 2 и 3

— — —

Возможность продолжения финансирования определяется по результатам этапа 1

В текущей заявке финансирование этапов 2 и 3 не запрашивается

Поддержка и интерес к проекту



Стрижов В.В.

д.ф.-м.н.,
профессор, кафедра
Интеллектуальных
систем ФПМИ

Поставленные перед MLDEV задачи

Совместный анализ результатов экспериментов в области машинного обучения (графики, диаграммы, таблицы) для направления исследования кортикограмм (ECoG)

Проведение экспериментов на множествах выборок, критериев качества и алгоритмов для обсуждения повторяемых результатов экспериментов

Поддержка подготовки курсовых работ / статей в области машинного обучения для обучения студентов на кафедре

Автоматическое проведение экспериментов и совместного обсуждения их результатов и текстов работ студентами между собой и с консультантами

Поддержка и интерес к проекту



Яковлев В.В.

к.ф.-м.н., зав. каф.
Алгоритмов и технологии
программирования ФПМИ

Поставленные перед MLDEV задачи

Использовать общедоступные и предоставленные добровольно (public computing) вычислительные ресурсы для проведения экспериментов для выполнения курсовых и дипломных работ

Обучать студентов выполнению воспроизводимых вычислительных экспериментов (на курсе Автоматизация программирования кафедры) с применением разработанной системы

Долгосрочное финансирование

- Заявка на проект в рамках программы 5–100 – разработка программных средств повышения качества и увеличения числа публикаций по результатам исследований, проводимых научными группами и студентами
- Дальнейшее развитие системы волонтерами из сформированного сообщества, участие в Google Summer of Code (2021)
- Получение спонсорского финансирования от компаний–партнеров МФТИ
- Привлечение сотрудников кафедр / базовых организаций / студентов, магистрантов и аспирантов к работе по проекту

Контакты

— — —

Антон Хританков, к.ф.-м.н.

руководитель проекта

доцент кафедры АТП

Email: anton.khritankov@acm.org

LinkedIn:

<http://linkedin.com/in/anton-khritankov>

Виктор Яковлев, к.ф.-м.н.

руководитель подразделения

проекта

зав. кафедры АТП

Email: iakovlev.vv@mipt.ru