

1. Команда проекта (ФИО, должность, телефон, почта)

Руководитель проекта: заведующий лабораторией Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ к.ф.-м.н. Безносиков Александр Николаевич (beznosikov.an@phystech.edu, +7 915 437-75-42).

Исполнитель проекта старший научный сотрудник МФТИ, к.ф.-м.н. Владислав Вячеславович Матюхин (vladmatyukh@gmail.com +7 925 991-59-49).

Исполнитель проекта научный сотрудник лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, аспирант 1 года обучения ФПМИ МФТИ Чежегов Савелий Андреевич

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 2 курса Змушко Филипп Александрович

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 1 курса Медяков Даниил Олегович

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 1 курса Молодцов Глеб Львович

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 1 курса Веприков Андрей Сергеевич

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 1 курса Солодкин Владимир Сергеевич

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 1 курса Богданов Александр Иванович

Исполнитель проекта ведущий инженер лаборатории Фундаментальных исследований искусственного интеллекта ФПМИ МФТИ, студент магистратуры 1 курса Былинкин Дмитрий Андреевич

2. Описание проекта

Краткое описание для сайта (2-3 предложения по структуре: Что? Для кого? Каким образом? Что благополучатели получат?)

Применение федеративного обучения ИИ для анализа медицинских данных и не только с целью популяризации данного нового направления обучения ИИ и развития его научной составляющей. Для решение актуальной задачи создается молодежная лаборатория состоящая из молодых ученых и студентов МФТИ. Лаборатория только в процессе создания.

Подробное описание проекта (опишите, что конкретно вы собираетесь делать).

Концепция федеративного обучения появилась не так давно с легкой руки исследователей Google, которые черпали мотивацию в довольно простой идее: большой объем данных для обучения ИИ хранится на устройствах пользователей (планшетах, смартфонах, ПК). Причем это персональные данные, использование которых может дать индивидуальные, персонализированные результаты для каждого отдельного клиента. У такой постановки есть и обратная сторона, которая порождает массу проблем от приватности используемых данных до организации процесса обучения на большом числе разнородных устройств неизвестных пользователей. В Google федеративное обучение началось с ИИ для анализа текстовых запросов на 100.000 телефонах. Но сейчас концепция обучения на уникальных частных данных проникла и в медицину (на данных больниц или носимых медицинских устройств), и в промышленность (на данных станков или камер наблюдения). Федеративные решения создают, как гиганты (Apple, Siemens, NVIDIA), так и стартапы (SAIL, Ferrum, OWKIN, Lynx.MD). Более того, концепцию федеративного обучения начинают продвигать на правительственном уровне крупных стран ЕС (<https://en.wikipedia.org/wiki/Gaia-X>) и США (<https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>), а один из ведущих и самых уважаемых ученых в области ИИ Ян Лекун выделил федеративное обучение, как одну из самых важных концепций в развитии будущего ИИ (<https://www.forbes.com/sites/robtoews/2020/10/12/the-next-generation-of-artificial-intelligence/?sh=4d14f60159eb>), поставив выше трансформерных архитектур, на которых базируются большие языковые модели.

Команда проекта является одной из немногих, кто активно исследует данную тематику в РФ. В рамках базовой кафедры ФПМИ в Институте системного программирования РАН (ИСП РАН) в прошлом году была открыта лаборатория проблем федеративного обучения (заведующий – руководитель проекта Безносиков А.Н.). В рамках работы данной лаборатории команда проекта совместно с Яндексом, Сеченовским университетом, ИСП РАН осуществила важный шаг в продвижении федеративного обучения. А именно, был поставлен модельный эксперимент, в котором, используя приватные медицинские данные ЭКГ снимков из разных источников, была обучена общая модель для детекции аномалий у пациента. Исследования были представлены в новостных сообщениях на платформах Яндекс, Habr, Vk, Т-Банк, Сnews, IXBT и многих популярных IT Telegram-каналов:

https://yandex.cloud/ru/blog/posts/2024/10/ml-models-and-federated-learning?utm_referred=https%3A%2F%2Fwww.google.ru%2F

<https://yandex.ru/company/news/02-08-10-2024>

<https://www.ixbt.com/news/2024/10/08/vpervye-v-rossii-jandeks-na-praktike-primenil-federativnoe-mashinnoe-obuchenie.html>

<https://habr.com/ru/news/849144/>

https://vk.com/wall-11283947_185428

<https://www.tbank.ru/invest/social/profile/basebel/8819440e-df4d-45ec-bd45-6451e6887008/?author=profile>

https://www.cnews.ru/news/line/2024-10-08_yandeks_pozvolit_partneram

Кроме этого, участники проекта имеют большое число публикаций в ведущих Q1 журналах и в просидингсах топ А* конференций по тематике федеративного обучения:

Q1 журнал: Beznosikov A. et al. On biased compression for distributed learning //Journal of Machine Learning Research. – 2023. – Т. 24. – №. 276. – С. 1-50.

Топ1 конференция по ИИ в мире: Beznosikov A. et al. Decentralized local stochastic extra-gradient for variational inequalities //Advances in Neural Information Processing Systems. – 2022. – Т. 35. – С. 38116-38133.

Топ1 конференция по ИИ в мире: Beznosikov A. et al. Distributed methods with compressed communication for solving variational inequalities, with theoretical guarantees //Advances in Neural Information Processing Systems. – 2022. – Т. 35. – С. 14013-14029.

Топ1 конференция по ИИ в мире: Beznosikov A., Takác M., Gasnikov A. Similarity, compression and local steps: three pillars of efficient communications for distributed variational inequalities //Advances in Neural Information Processing Systems. – 2024. – Т. 36.

Команда проекта является либо выпускниками, либо действующими аспирантами и студентами МФТИ, более того, активно преподает на ФПМИ различные курсы. В конце 2024 года на Физтехе открылся центр искусственного интеллекта. Из участников проекта была сформирована лаборатория Фундаментальных исследований искусственного интеллекта, где и планируется продолжить и исследования в области федеративного обучения.

За 2025 год предполагается перенести результаты, полученные для ЭКГ снимков на другие медицинские данные. В частности, показать, что федеративное обучение может помочь в ситуации, когда информация на одних и тех же пациентов хранится в различных удаленных и частных местах: например, ЭКГ снимки, анамнез – у одного поставщика данных, анализы – в другого, МРТ снимки – в третьего. Такого рода постановка эксперимента практически полностью отражает ситуацию хранения медицинских данных пациентов во многих регионах РФ. Приведенный выше эксперимент позволит показать еще одно применение федеративного обучения для реальных задач и поможет еще больше популяризировать федеративное обучение, как важный подход в развитии ИИ.

Кроме этого предлагается исследовать множество более фундаментальных и научных вопросов в области федеративного обучения:

- задачу автоматической детекции поставщиков “грязных” данных (например, выявление часто ошибающихся аннотатор – врачей размещающих данные);
- задачу приватности и защищенности персональных данных, участвующих в процессе обучения (например, создать довольно дешевые способы защиты приватности и оценить, есть ли смысл атаковать данные и модель ИИ, или это требует значительных ресурсов, превышающих ценность данных);
- задачу создания сети асинхронного общения участников федеративного обучения между собой для ускорения процесса обучения модели ИИ;
- и другие важные аспекты федеративного обучения, на которыми бьются ведущие мировые группы.

3. Актуальность

Зачем нужен этот проект, какую проблему он решает?

см. пункт 2

4. Кто или что является благополучателем результата проекта?

В качестве благополучателя может выступать ФПМИ МФТИ и МФТИ в целом – см. пункт 5

Также в качестве благополучателя в более далекой перспективе могут выступать потенциальные пользователи федеративного обучения – см. пункты 2 и 8.

5. Планируемый результат проекта

Непосредственный результат

Формальным результатом работы в 2025 году должна стать минимум одна статья (в Q1-Q2 журналах и просидингах А*/А конференций – см. пример в пункте 2) в области федеративного обучения, а также несколько выступлений на ведущих мировых конференциях по ИИ. Кроме этого предполагается провести модельный эксперимент (с соответствующим новостным освещением), показывающий применимость федеративного для реальных прикладных задач.

Как результат проекта влияет на развитие МФТИ и/или комьюнити вокруг него.

Федеративное обучение – перспективная тема исследований, важный шаг в развитии искусственного интеллекта, как в России, так и в мире. МФТИ, как один из ведущих исследовательских центров, всегда был на передовых ролях в перспективных исследованиях. Поэтому кажется естественным выпускникам Физтеха не просто развивать федеративное обучение, но и делать это в рамках МФТИ с прицелом на поддержание ведущих позиций Физтеха в научном сообществе в области ИИ.

Стоит отметить, что как в команде проекта, так и в целом в лаборатории фундаментальных исследований ИИ все сотрудники моложе 30 лет, а большая

часть молодежи 25. Более того, команда проекта активно преподает на ФПМИ, в том числе читает курсы на весь поток ФПМИ, а также на базовых кафедрах ФПМИ (ИСП РАН, ИППИ РАН). В рамках данных курсов освещаются темы в том числе связанные с федеративным обучением, за которые студенты часто цепляются и хотят начать работу в этих направлениях. К нам приходит много мотивированных студентов. Мы стараемся поддерживать хотя бы незначительно любые начинания и попытки первых исследований студентов и молодых ученых. Все больше сильных студентов 3-4 курсов уходят в индустрию, даже не попробовав себя в науке, хотя ровно для раннего старта в науку и задумывался Физтех. С помощью даже небольших финансовых вливаний мы хоть как-то можем удерживать ребят вокруг интересных научных задач и в некотором смысле популяризировать науку среди студентов МФТИ.

6. План реализации проекта, его этапы и их сроки.

На данный момент работы ведутся по всем направлениям. Примерный сроки представления результатов

- май 2025: окончание работ над 3-4 научными статьями в области федеративного обучения, отправка работ на рецензирование в ведущие журналы и мировые конференции;
- август 2025: окончание проведения модельного эксперимента применимости федеративного обучения для реальной прикладной задачи
- июль - декабрь 2025: выступление на ведущих мировых конференциях по ИИ с результатами проекта
- октябрь - декабрь 2025: новостное освещение модельного эксперимента применимости федеративного обучения для реальной прикладной задачи
- сентябрь 2025 - февраль 2026: выход научных публикаций с результатами проекта

7. Бюджет проекта а. *Общий бюджет проекта* б. *Сумма, запрашиваемая от ФЦК МФТИ;* с. *Есть ли софинансирование? В каком объеме и его источник? Укажите потенциальных партнеров, в том числе тех, которые поддерживают не только деньгами.* d. *Почему проект не может быть полностью профинансирован из бюджета МФТИ или иных источников? е. Прикрепите отдельным от презентации xls-файлом таблицу с построчной расшифровкой всего бюджета, в которой укажите, на какие статьи расхода вы запрашиваете денег у ФЦК, а на какие планируете получить спонсорское финансирование;*

Общий бюджет проекта – 10 000 000 рублей.

Сумма, запрашиваемая от ФЦК – 5 000 000 рублей.

Как упоминалось в пункте 2, команда проекта сотрудничает с Яндексом, Сеченовским университетом и ИСП РАН. Сеченовский университет и ИСП РАН предоставляют данные для проведения экспериментов по федеративному обучению. Со стороны Яндекса поступает маркетинговая поддержка: все новостное освещение, приведенное выше – результат работы пиар-офиса Яндекса. Также со стороны ИСП РАН и Яндекса в проекте задействовано несколько внешних

сотрудников (не указанных в пункте 1): 2 junior, 1 middle, 1 senior программисты со стороны ИСП РАН, 1 middle программист со стороны Яндекса.

Несмотря на появление центра ИИ в МФТИ (откуда и планируется основное финансирование лаборатории в будущем), получение государственных средств на фундаментальные научные исследования планируется не раньше середины лета 2025 года (зарплаты сотрудникам поступят не раньше сентября). Работа с потенциальными бизнес-заказчиками только разворачивается. На ближайшее полгода (возможно, немного больше) команде проекта необходимо финансирование для продолжения функционирования.

| Статья расходов | Сумма |
|---|--|
| <i>Заработные платы (с учетом налогов и резервов)</i> | <i>5 000 000 (пять миллионов) рублей</i> |

8. Долгосрочное развитие проекта

Является ли проект разовой акцией или планируется его повторение на регулярной основе? Планируете ли вы в будущем искать другие источники финансирования?

Как уже было описано выше, развитие федеративного обучения является важной фундаментальной ступенью в развитии искусственного интеллекта нового поколения. Важно приумножать успехи в этой области не только в мировом сообществе, но и в России. Поэтому работа в этой области будет продолжена как по научной составляющей, так и по промышленной. Предполагается искать спонсирование, как со стороны государства (РНФ и другие фонды), так и со стороны промышленных заказчиков (Сбер, Т-Банк, Яндекс, Газпромнефть). В идеале есть надежда, что бизнес поддержит не только исследования в данной области, но и не побоится массово использовать решения для реальных задач и приложений.

9. Подразделение МФТИ, через которое будет проходить финансирование проекта. *Укажите, какие договорённости есть с руководителем подразделения. В отдельных случаях мы можем запросить одобрение курирующего проректора. Укажите, кто будет заниматься документооборотом для оплаты счетов через МФТИ. а. ФИО и контакты руководителя проекта/подразделения, который будет подписывать ФЛС. б. ФИО и контакты ответственного исполнителя по проекту.*

Проект будет реализовываться в лаборатории Фундаментальных исследований искусственного интеллекта, руководитель проекта Безносиков А.Н. (beznosikov.an@phystech.edu, +7 915 437-75-42).

Ответственным исполнителем будет являться Матюхин Владислав, старший научный сотрудник лаборатории, vladmatyukh@gmail.com +7 925 991-59-49.

10. Как ваш проект будет способствовать популяризации деятельности ФЦК МФТИ среди студентов, сотрудников и выпускников, какие конкретные действия вы планируете для этого предпринять.

Планируется указывать ФЦК, как источник финансирования в публикуемых в рамках проекта статьях – см. примеры статей выше в пункте 2.

Планируется указывать ФЦК, как источник финансирования в новостях, связанных с темой проекта – см. примеры новостей выше в пункте 2.

Планируется указать ФЦК, как спонсора в презентациях и постерных выступлениях на научных конференциях и семинарах.

Как уже упоминалось выше, к нам в лабораторию приходит большой поток молодежи, нацеленной заниматься наукой, и мы стараемся материально их поддерживать. Раз в полгода мы устраиваем мероприятия по привлечению студентов, часто участвуем в таких мероприятиях со стороны ФПМИ и МФТИ. Планируется упоминать на таких мероприятиях ФЦК, как одного из спонсоров финансирования студентов во время их первых шагов в науке.