

# ITMO

## Опенсорс в ИТМО: как дела в 2024?

Никитин Николай

лидер движения ITMO.OpenSource,  
руководитель лаборатории автоматического машинного  
обучения, ИТМО

# Что есть в ИТМО?



Свои **открытые проекты** в области ИИ - и не только.

>35 библиотек и фреймворков.

Крупнейшее в России академическое опенсорс-сообщество.

>1200 участников.

Уникальный цикл опенсорс-митапов: барных и не только.

Этот - уже 8-й по счету.

Авторские исследования ландшафта опенсорса

Про последнее упомянуло более 40 крупных СМИ.

И многое другое

# С чего мы начинали?

В 2020 мы имели:

Одиночные открытые библиотеки;

Минимум внешних участников;

Отсутствие популяризации решений;

Что сделали:

запустили ряд проектов по созданию Open Source инструментов в области ML (как крупномасштабных, так и студенческих).



Суть

# Что имеет по проектам:



Из чего состоит	Кто использует	Востребованность	Зачем?
35+ фреймворков и библиотек в области ИИ, готовые к использованию, снабженные тестами и документацией.	Разработанные инструменты используются в десятках стран - в том числе сотрудники NVIDIA, Uber, Google Cloud, JetBrains, Databricks.	1500+ звезд, 300к скачиваний (ТОП 1 в РФ среди университетов).	<ul style="list-style-type: none"><li>• Польза обществу;</li><li>• Обучение профессионалов;</li><li>• Вдохновение тех, кто участвует;</li><li>• Технологический суверенитет.</li></ul>

**Репозиторий:** <https://github.com/aimclub>

# Сообщество ITMO OpenSource



Сообщество разработчиков и пользователей открытого научного кода в ITMO:

- [https://t.me/itmo\\_opensource](https://t.me/itmo_opensource)  
(чат ITMO Opensource)
- [https://t.me/scientific\\_opensource](https://t.me/scientific_opensource)  
(канал “Научный опенсорс”)

Помощь в решении проблем с созданием открытых решений, их популяризации, анонсы мероприятий и полезных материалов.

Более 1000 участников.



<- Join ITMO.OpenSource

# С чего начать «вкатываемся»?

У научных команд часто возникают типовые проблемы:

- Как написать понятный readme?
- Как и зачем реализовать автотесты?
- Где и как можно рассказать о своём проекте?

Поэтому мы создали репозиторий с лучшими практиками, руководствами и шаблонами в области Open Source.

<https://github.com/aimclub/open-source-ops>

🔗 Основные разделы:

## Инструкции

- С чего начать разработку open-source библиотеки;
- Зеркалирование GitHub -> GitLab;
- Мультиязычные README;
- Создание документации;
- Настройка ботов для репозитория.

## Шаблоны

- Типовой шаблон README для open-source проектов.

## Лучшие практики и примеры

- Организация управления open-source проектом;
- Полезные ссылки для авторов open-source библиотек;
- Советы по работе в Pull Request-ax.

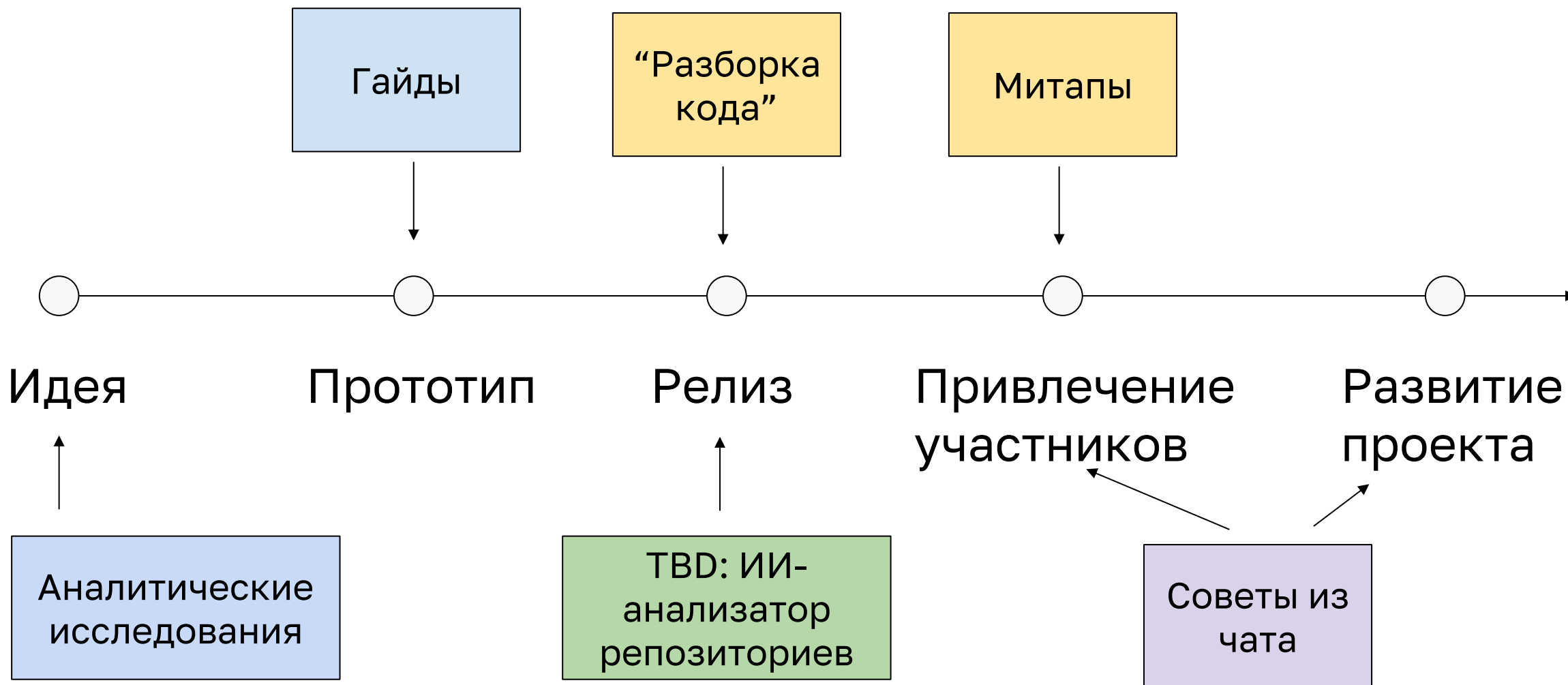
## Открытый код и наука

- Где опубликовать научную статью про OS-разработку?.

## Открытый код в ИТМО и не только

- Репозитории научных подразделений и лабораторий;
- Pet-проекты, связанные с наукой;
- Научно-популярные посты о open-source в ИТМО.

# “Технология” поддержки опенсорса в ИТМО



# Откуда взять денег на опенсорс?

Гранты КОД-ИИ: Фонд содействия Инновациям, до 10 млн. рублей, около 15 поддержанных проектов в год.

Примеры выигравших проектов от участников сообщества:

- Библиотека для **автоматизации генерации сверточных нейронных сетей** для задач предсказательного моделирования природных сред - Юлия Борисова
- Разработка фреймворка **прикладных инструментов анализа данных на основе байесовских сетей** - Ирина Деева
- Открытая библиотека для **автоматического решения задачи кластеризации мультимодальных данных на GPU** - Сергей Муравьев



# Ещё про митапы

- **Предлагайте форматы, в которых хотите поучаствовать**  
Недавно провели митап по Байесовским методам, который был инициирован обсуждением в чате;
- **Предлагайте темы in advance**  
Даже если нет объявленного митапа, доклады можно предлагать в issue;
- **Онлайн тоже можно**  
Половина наших митапов - онлайнновые и в них могут поучаствовать все.

# Неклассические способы популяризации

- Фреймворк FEDOT (а равно и другие участники его экосистемы - FEDOT.Industrial, GOLEM) были награждены специальным призом премии в области ИИ "Гравитация".
- Фреймворк VAMT был награжден премией за лучший научный проект на GitVerse – «Звезда GitVerse»



ИТМО

# Аналитические исследования



Начала с “Хабра”



81.73

Рейтинг

Подписаться

ИТМО

IT's MOre than a University



itmo\_nsslab 4 апр в 17:04

## Open Source в российском ИИ: исследование ландшафта

Простой 7 мин 2.6K

Блог компании ИТМО, Open source\*, GitHub\*, Машинное обучение\*, Искусственный интеллект

Аналитика

Привет, Хабр!

Это исследователи из NSS Lab ИТМО и создатели сообщества [ITMO.OpenSource](#). Мы считаем, что важно делать не просто науку, а открытую науку, результатами которой можно легко воспользоваться. Поэтому [пишем](#) об открытом коде, проводим митапы «Scientific Open Source» и разрабатываем научные решения с открытым кодом. Эта статья написана на материале исследования, которое мы презентовали на последнем митапе. В ней поговорим об открытом научном коде, посмотрим, как с ним обстоят дела в разных организациях, и поделимся подборкой репозиторий, которые могут решать различные научные задачи.

Зачем вообще нужен научный открытый код?

Продолжили - лендингом



Содержание

Связаться с авторами

## Использование ML/Data Open Source в России

Проанализировали решения, инструменты и их применимость. Выделили главных игроков, обозначили проблемы и перспективы.

Основные инсайты по итогам исследования

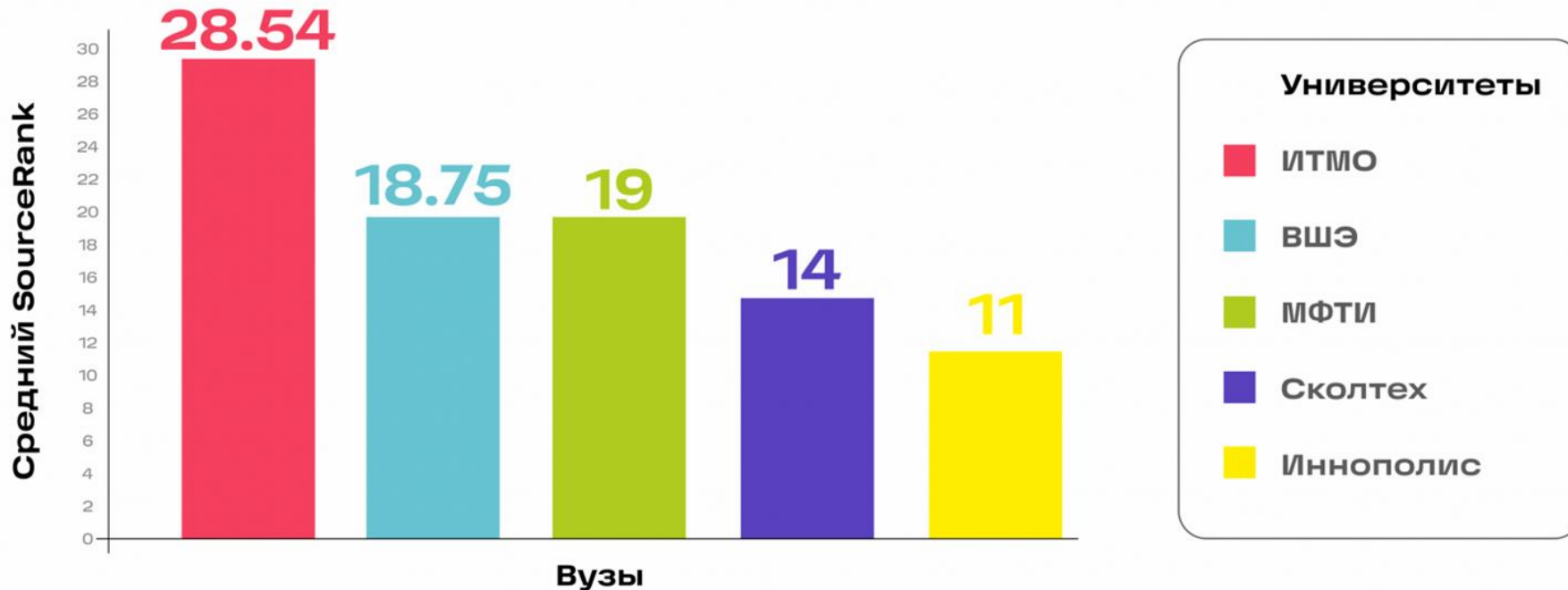
- [1] Опенсорс - не российский, а общий**  
Почти все компании-разработчики открытых решений в Data/ML находят пользователей не только на внутреннем, но и на международном рынке. Востребовано то, что лучше работает без привязки к конкретной стране.
- [2] В России создается много сильных открытых инструментов**  
Они востребованы в технологическом стеке крупных компаний и частных пользователей. Их авторы – как «бигтех», так и небольшие команды, в том числе академические.
- [3] Использование публичных API GitHub позволяет собрать много интересных данных о том, кто, где и когда использует открытые проекты**
- [4] Опенсорс – это не только про код**  
Публикации моделей, данных и бенчмарков тоже важны – многие активно используемые датасеты создаются российскими компаниями.
- [5] В компаниях есть энтузиасты опенсорса – и их много**  
Хотя они продвигают свои идеи и участвуют в развитии своих и чужих проектов, им все же сильно не хватает системности и поддержки (в том числе и схем монетизации).
- [6] Продвигать свои проекты сложно**  
Есть много конференций и митапов, но все же форматов и площадок недостаточно. Кроме того, у потенциальных участников открытых проектов не хватает и квалификации. Здесь могут помочь университеты и их студенты – опыт их участия в разработке опенсорса обычно позитивный.

# А как у других?

Организация	Инфоканалы	Самые популярные репозитории
ИТМО AIM.CLUB	ИТМО OpenSource + канал “Научный опенсорс”	<a href="#">FEDOT</a> , <a href="#">BAMT</a> , <a href="#">FEDOT.Industrial</a>
ВШЭ	“Открытый код ФКН ВШЭ” <a href="https://t.me/hse_cs_opensource">https://t.me/hse_cs_opensource</a> , 620 подписчиков	<a href="#">hsemotion</a> , <a href="#">roerich</a> , <a href="#">probaforms</a>
МФТИ, СПС	СПС, <a href="https://t.me/SciProgCentre">https://t.me/SciProgCentre</a> , 500 подписчиков	<a href="#">DeepPavlov</a> , <a href="#">kmath</a>
Сколтех	Новые официальные форки бывшего репозитория Сколтеха	<a href="#">ttry</a> , <a href="#">h2tools</a>

# А как у других - сравнение

Нормализованное значение критерия SourceRank




$$SR_{UNI} = SR_1 + \frac{SR_2}{2} + \dots + \frac{SR_n}{n}$$

где  $SR_{UNI}$  – итоговое значение критерия SourceRank для университета,  $SR_n$  – значение метрики SourceRank для n-ного проекта.

Source Rank - кумулятивный критерий, учитывающий как базовые признаки - наличие readme и лицензии, частоту релизов, количество зависимых репозиторий и число контрибьюторов. Для уравнивания нишевых и массовых проектов SourceRank использует логарифмическую шкалу при расчёте некоторых её составляющих.

# Поговорили с экспертами



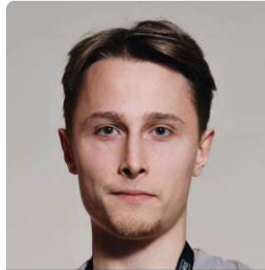
**Максим Савченко**  
Сбер, Управляющий директор, руководитель Центра инструментов машинного обучения Лаборатории AI

[О себе →](#)




**Сергей Бережной**  
Яндекс, Директор по взаимодействию с разработчиками

[О себе →](#)



**Александр Чиков**  
Т-Банк, Тимлид

[О себе →](#)



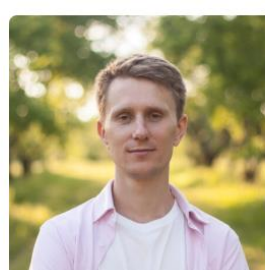
**Юрий Кацер**  
Рокет Контрол, Руководитель направления DS в обогащении

[О себе →](#)



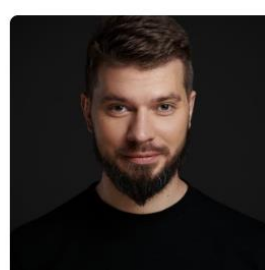
**Александр Сидоров**  
Wildberries, Head of Data Science

[О себе →](#)




**Александр Волынский**  
VK, Технический менеджер Cloud ML Platform, архитектор VK Cloud

[О себе →](#)



**Алексей Смирнов**  
CodeScoring, Founder & CEO

[О себе →](#)



**Александр Нозик**  
МФТИ, Директор центра научного программирования

[О себе →](#)

Вовлекли в исследование представителей бигтеха и академии

# Покопали открытые данные

↳ Методика оценки ML/Data-решений

Общая информация

Критерии степени развития проекта

Критерии степени используемости

## Критерии степени используемости проекта

Оценка того, насколько проект используется (часто ли его добавляют в избранное, часто ли пишут пользователи, часто ли скачивают и цитируют, широка ли география использования)

forks\_count

количество форков

Stars

Количество звезд (log)

open\_issues\_count

количество открытых issues

Downloads last month

количество скачиваний за последний месяц (если есть на рупі)

Временная

динамика звезд

# Разузнали, кто чем занят

↳ Опенсорс-сообщества

Корпоративные

Академические

Студенческие

- + центр “ИИ в Химии” ИТМО-  
<https://github.com/ai-chem>
- + Новый ФизТех ИТМО -  
<https://github.com/swarmtronics>

- **ИТМО AI Talent Hub, Учебно-научные лаборатории МФ ТиНТ**

Другой аспект опенсорса в ИТМО – использование его при обучении студентов бакалаврианта и магистратуры. В учебно-научных лаборатория МФ ТиНТ ([LISA](#), [CTLab](#)) реализуются многочисленные опенсорсные мини-проекты в сфере ИИ, проводятся семинары. В магистратуре AI Talent Hub опенсорсные проекты создаются как выпускные работы студентов, а также как результат специализированных треков хакатонов.

[Группа ВК](#), [AI Talent Hub](#),

- **ФОСС Кружок**

Это сообщество объединяет не только студентов, но и школьников - а также организует масштабный конкурс проектов.

[Канал](#)

- **GoOpenSource**

Коллективная база знаний и сообщество открытого кода, идея которого – сделать открытый код доступным бизнесу. Реализовали идею российской витрины программного обеспечения с открытым кодом, первыми контрибьюторами проекта стали студенты Иннополиса.

[Сайт](#), [канал GOS](#), [канал GoOpenSource](#)



# Поговорили о перспективах

## Что дальше?

[1] ИИ все больше проникает в рабочую деятельность любого разработчика. Базовый уровень такого взаимодействия – помощники общего назначения вроде ChatGPT или более специализированные – например, CoPilot от GitHub.

Однако, сейчас появляется много решений нацеленных именно на помощь в создании опенсорс-проектов, например:

- RepoAgent

Инструмент для авто-генерации документации;

- RepoGPT

Индивидуальный разработчик) – инструмент, предназначенный для суммаризации файлов кода (.py) и документации (.md) в репозитории;

- llm-code-review

(индивидуальный разработчик) инструмент для ревью pull request-ов с помощью БЯМ

- bug-agent-llm

LLM-агент, способный самостоятельно исправлять ошибки из по их описанию;

В следующем году:  
попробуем сделать  
ИИ-ассистента  
опенсорсера.

[2] Какие задачи могут быть решены такими инструментами:

- Чат на основе документации
- Code Review
- Автоматизированное исправление ошибок

- Актуализация примеров
- Анализ результатов тестирования

- Вопросы и ответы по коду и документации

- В перспективе - резкое повышение автоматизации разработки проекта в целом

# Куда ещё сходить?

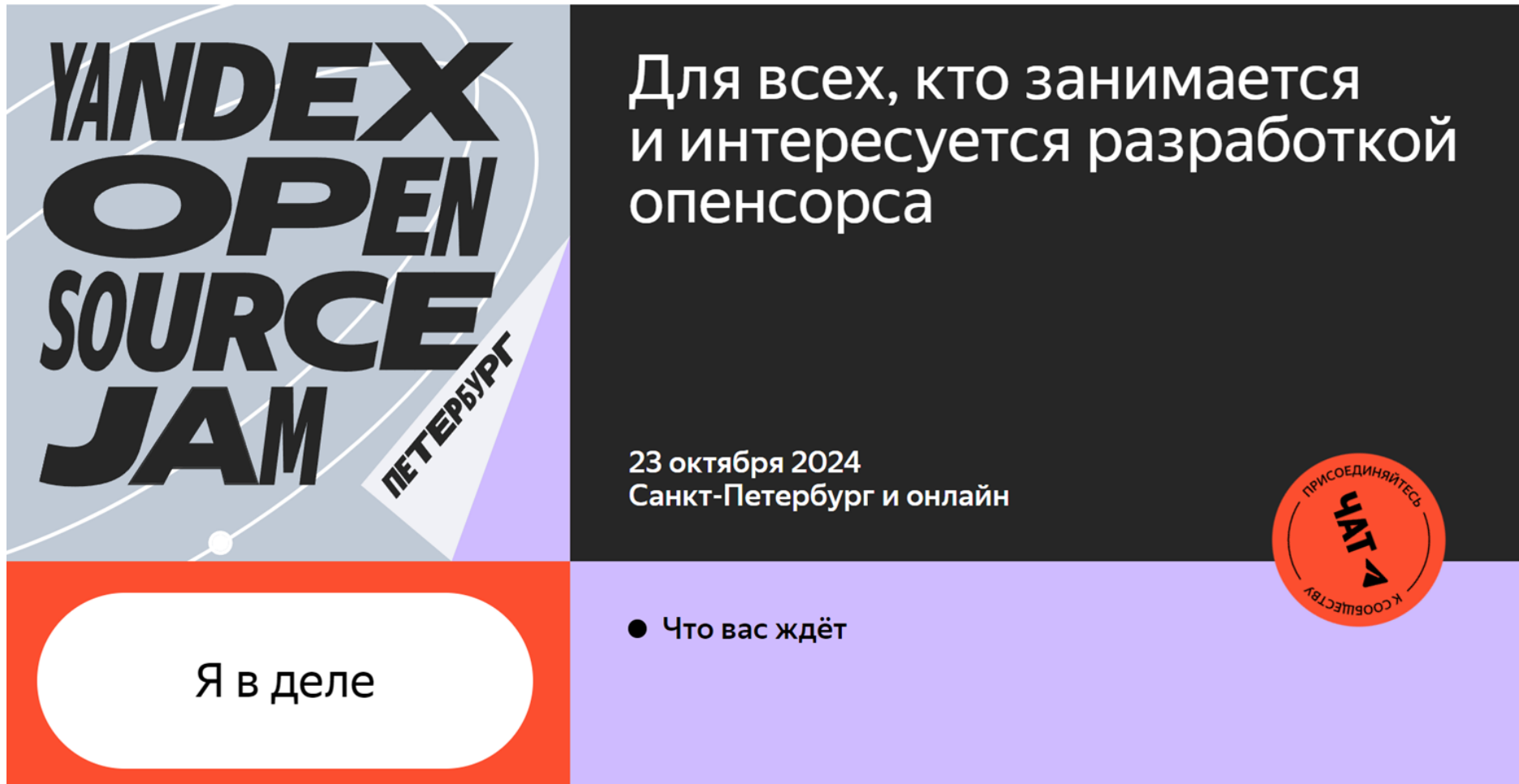
Open Source 

Что вас ждёт

Программа

Место

 [Добавить в календарь](#)




**YANDEX  
OPEN  
SOURCE  
JAM**

**ПЕТЕРБУРГ**

Для всех, кто занимается  
и интересуется разработкой  
опенсорса

23 октября 2024  
Санкт-Петербург и онлайн

 ПРИСОЕДИНЯЙТЕСЬ  
К СОБЫТИЮ

● **Что вас ждёт**

**Я в деле**

<https://opensource.yandex/jam/>

Спасибо за внимание

**iTMO** *re than a*  
**UNIVERSITY**

[nnikitin@itmo.ru](mailto:nnikitin@itmo.ru)