

Национальный центр когнитивных разработок Платформы для прикладных решений на основе ИИ

Санкт-Петербург, 2021



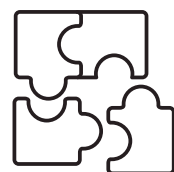
НАЦИОНАЛЬНЫЙ ЦЕНТР КОГНИТИВНЫХ РАЗРАБОТОК
УНИВЕРСИТЕТА ИТМО



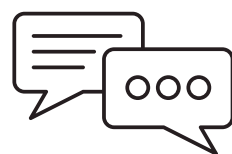
НЦКР



Центр компетенций Национальной технологической инициативы по направлению «Машинное обучение и когнитивные технологии».



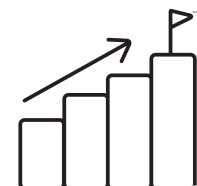
«Пан-университетское» структурное подразделение Университета ИТМО.



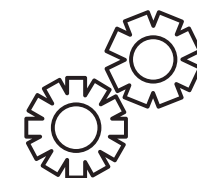
Консорциум промышленных партнеров и соисполнителей (Газпромнефть, Сбер, МТС, Мэйл.Ру, Банк «Санкт-Петербург», ЦРТ и пр.).



Более 200 штатных сотрудников (математики-алгоритмисты, дата-сайнтисты, дата-инженеры, разработчики ПО и девопсы, а также предметные специалисты).



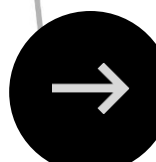
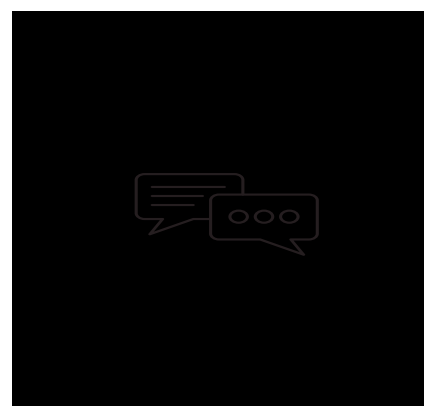
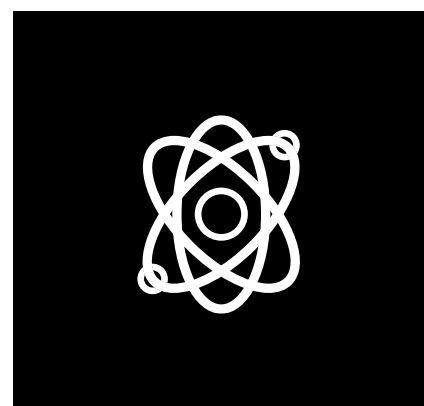
Совокупный доход от проектной деятельности 1+ млрд. руб. (2020).



Единичная разработка и кастомизация наукоемких **цифровых** решений под заказ, создание PoC и MVP, научный консалтинг и экспертиза.



ТЕХНОЛОГИЧЕСКИЕ ПРЕИМУЩЕСТВА: ПОЧЕМУ ПОЛУЧАЕТСЯ «БЫСТРО» И «ХОРОШО»



DataMall: Платформа управления большими данными



модель «биржи» данных
поверх различных источников
(может объединять хранилища данных
под Hadoop, Yarn, Spark, Storm, SAP HANA,
Cassandra и др.)



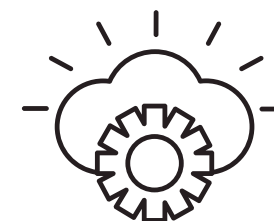
эффективное хранение, извлечение и
обработка данных в форме временных
рядов и графовых структур (существенно
быстрее, чем универсальные системы
хранения)



возможность структурирования
разрозненных данных в процессы
(Process mining)



по необходимости - сбор данных
из открытого Интернета
(социальные сети, сайты
объявлений, форумы и пр.) – с
возможностью интеллектуальной
настройки Краулера на
произвольный источник



быстрое конструирование и
проверка различных моделей ML
на основе данных в различных
формах
(ноутбуки Python, low-code-
программирование,
генеративный дизайн моделей с
помощью AutoML)

**Сокращение времени разработки цифровых объектов ИИ в 2-4 раза,
возможность массового создания типовых объектов ИИ для прикладных задач**



Идея DataMall - совмещение данных и моделей их обработки

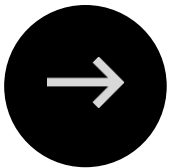
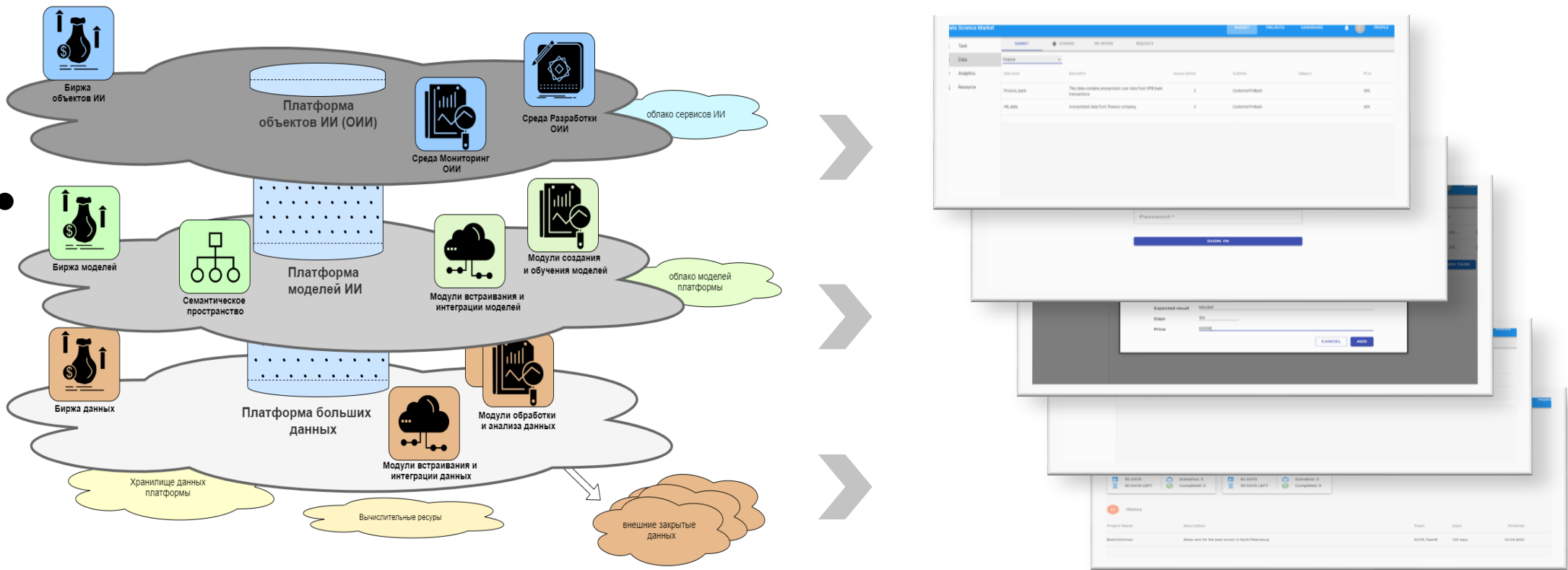
Повышение качества и снижение сроков разработки объектов ИИ за счет автоматизации лучших практик управления разнородными источниками больших данных и построения моделей ИИ

Уровни реализации платформы:

- уровень данных
- уровень моделей
- уровень готовых объектов

Направления использования:

- платформа «коллабораций»
- «биржа» данных, артефактов
- вычислительная платформа DDA – моделей



Идея DataMall - совмещение данных и моделей их обработки

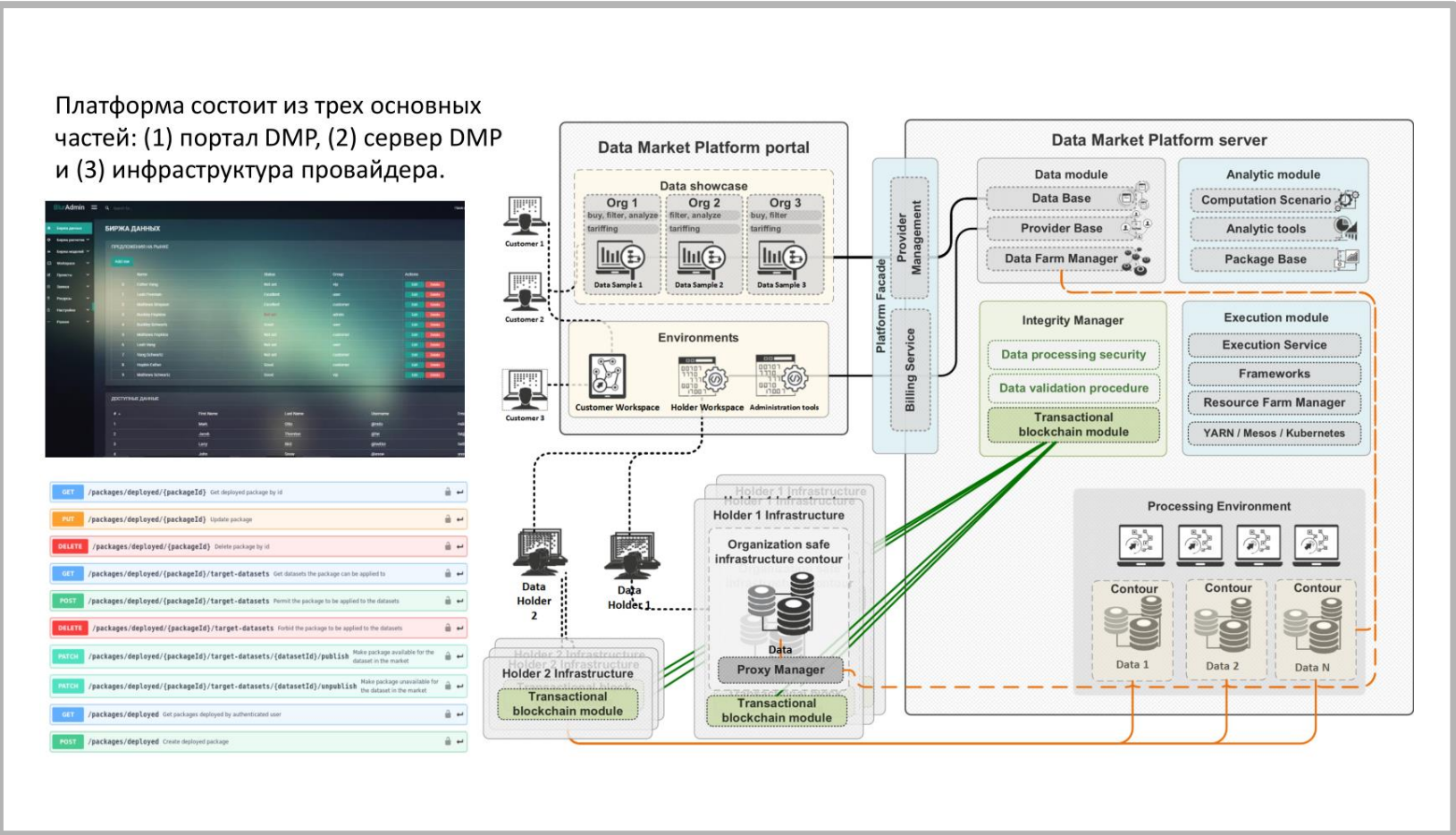
Повышение качества и снижение сроков разработки объектов ИИ за счет автоматизации лучших практик управления разнородными источниками больших данных и построения моделей ИИ

Конкуренты:

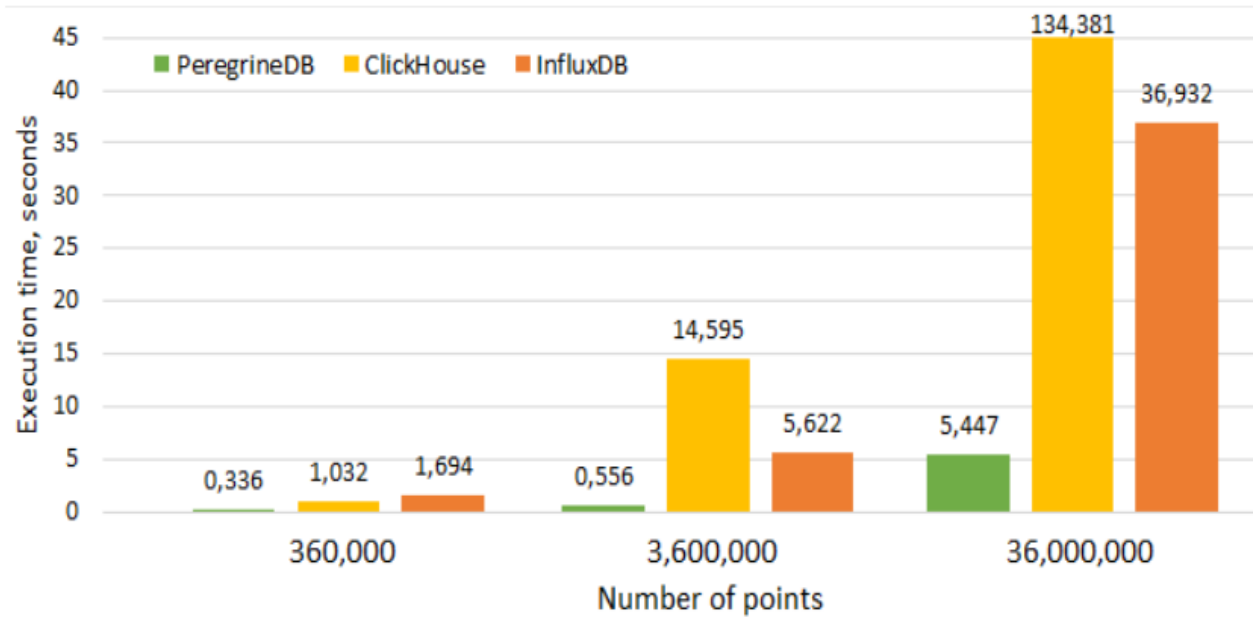
	DATA MALL	Qlik	Streamr	Databroker DAO	Dawex
Пакетная обработка	+	+	-	-	+
Потоковая обработка	+	-	+	+	-
Витрина	+	+ / -	-	-	+
Адаптеры	+	-	+ / -	+ / -	-
Сценарии	+	-	-	-	-
РРУД, «песочницы»	+	-	-	-	-



Технология Peregrine - кастомизация платформы DataMall

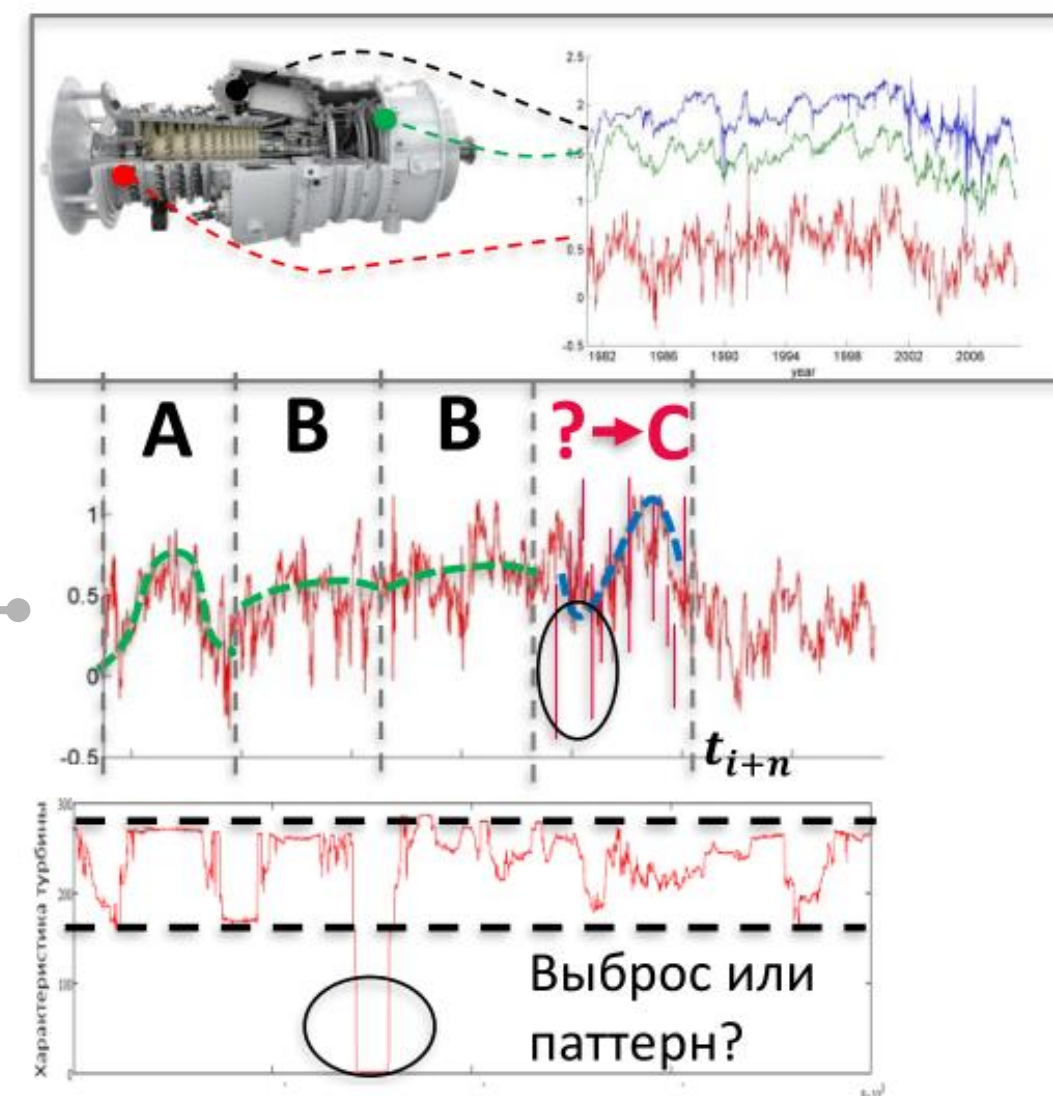


Конкуренция:



Технология Peregrine - кастомизация платформы DataMall

Очень быстрые операции с данными в форме временных рядов



Краулер - как обогащаются данные через поиск в Интернете

Задача: эффективный сбор полуструктурированных данных в Интернете

Решение: создание интеллектуального ядра краулера в рамках распределенной архитектуры:

Распределенная архитектура с поддержкой нескольких одновременных клиентов и возможностью «горячего» подключения новых узлов-обработчиков

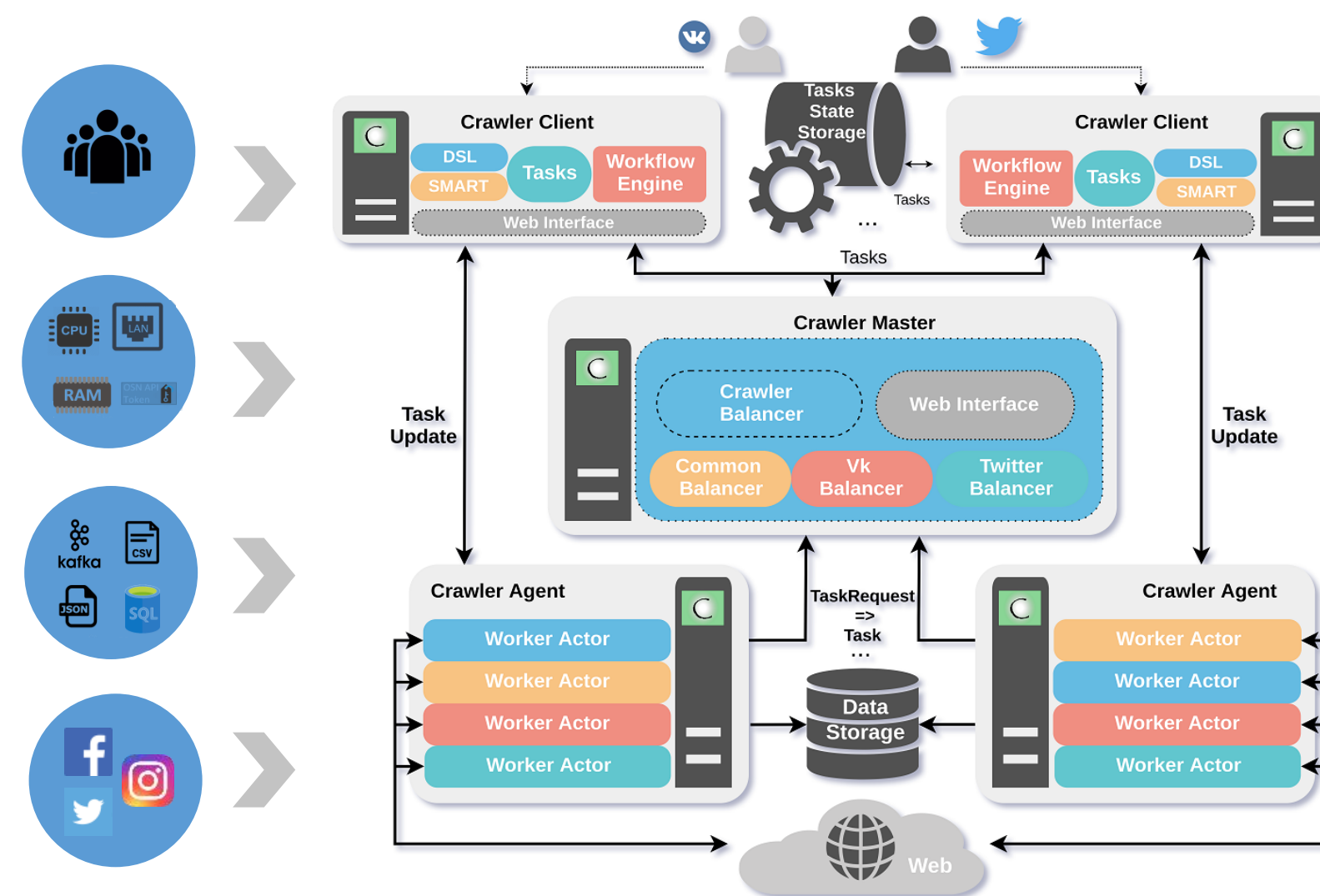
DSL и чат-бот для описания сценариев сбора данных

Workflow Engine - генерация задач и управление ходом исполнения

Балансировщики нагрузки для различных типов задач, контроль за расходом ресурсов и «вежливость» по отношению к сайту с собираемыми данными

Поддержка хранилищ и форматов: HDFS, Kafka, MongoDB, PostgreSQL

Краулер вызывается непосредственно из DataMall



Process Mining – как разрозненные данные собираются в процессы

Структурирование разрозненных данных в процессы

Реализация и расширение базовых механизмов интеллектуального анализа процессов (process mining)

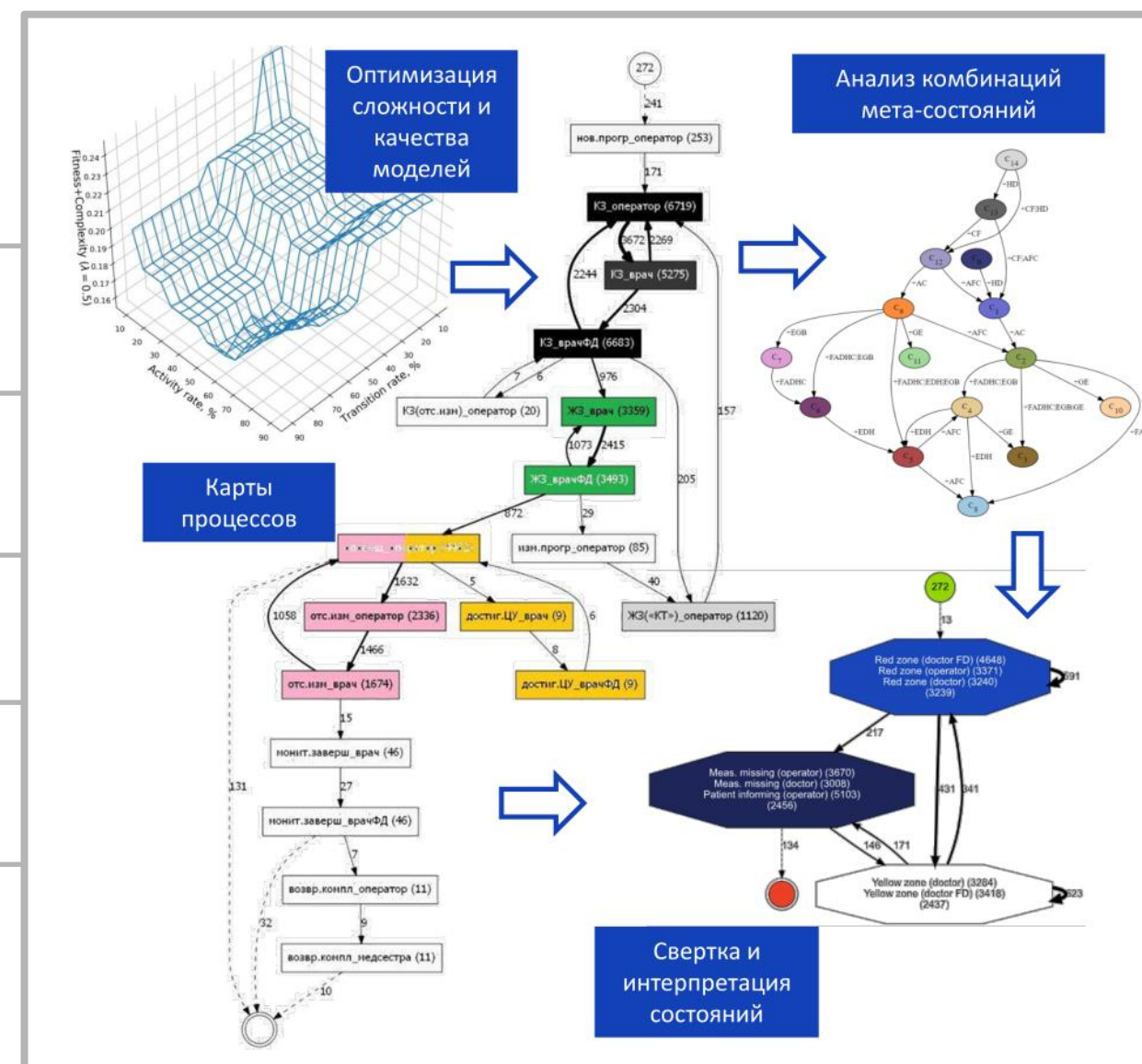
Автоматически регулируемая сложность и интерпретируемость карт процессов

Свертка и интерпретация циклических состояний с учетом сложности модели

Интеграция с предсказательными моделями состояний для интерпретации переходов

Широкий спектр применения: модели поведения, клинические пути, бизнес-процессы предприятий и пр.

Итог: автоматическая настройка инструментария интеллектуального анализа процессов для восстановления БП



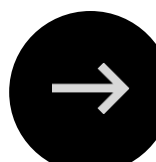
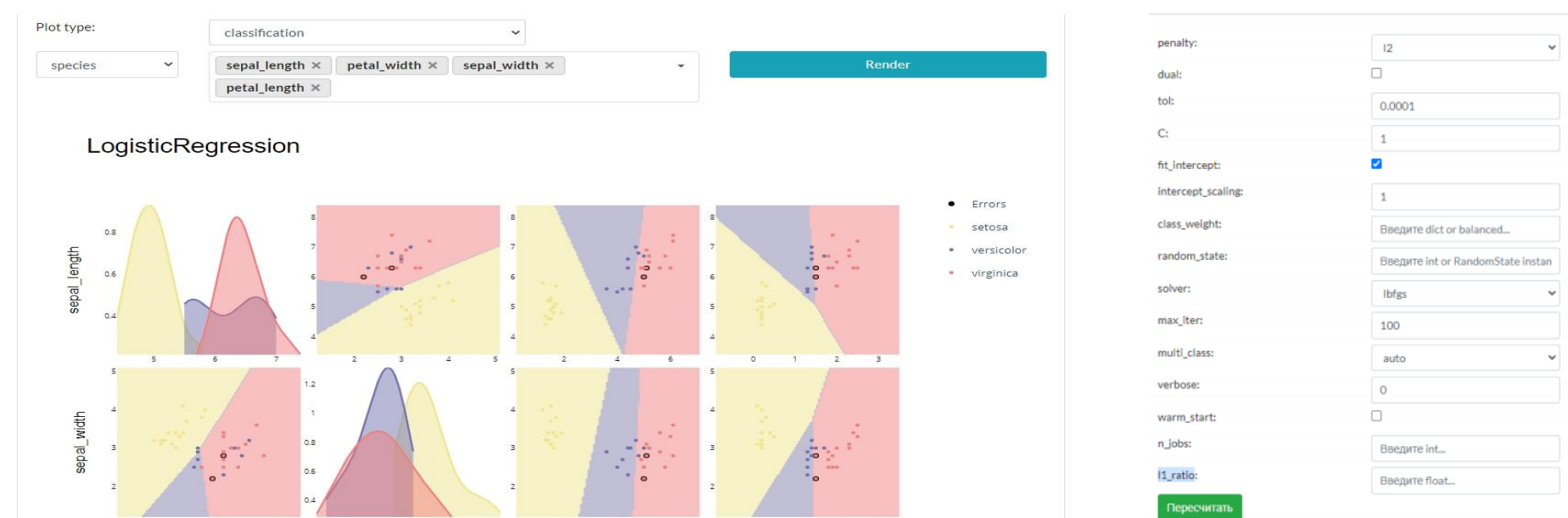
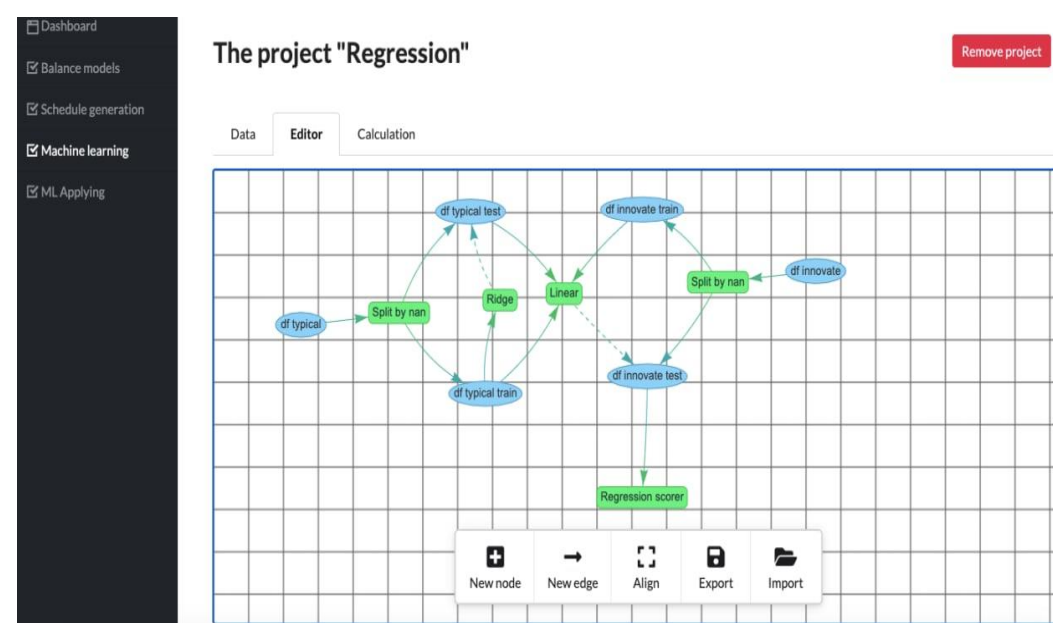
SMILE: Low-code платформа – создание моделей ML на данных

Low-code платформа - система для создания, настройки, проверки и использования моделей на данных с помощью ML:

нативное представление модели в форме графа с возможностью трансляции в код

использование существующих открытых фреймворков ML (более 100 процедур)

встроенные методы оценки и улучшения качества моделей



FEDOT: AutoML – Автоматическое создание моделей

FEDOT – фреймворк для автоматического создания моделей

Особенности фреймворка:

Variable Structure Pipelines

Основа алгоритма - GP

Поддержка Tabular, Timeseries данных

Поддержка гибкой модификации и расширения модулей фреймворка

Основная идея: (а) поддержка широкого набора задач и типов данных и (б) возможность замены отдельных модулей

В разработке:

Многокритериальная оптимизация

Расширенная предобработка данных

Поддержка NLP и CV задач

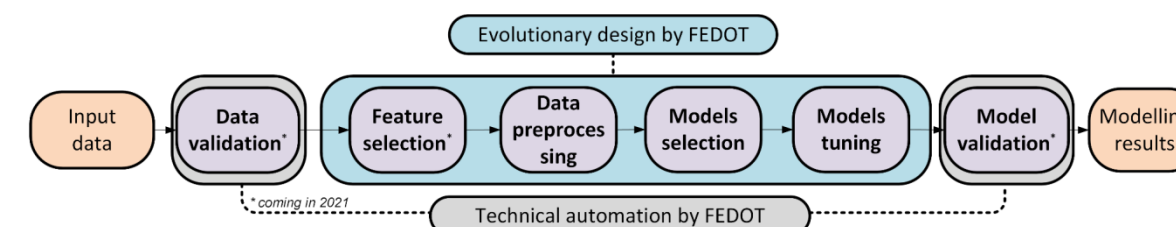
Preset-ы для определенных типов задач

WebGUI

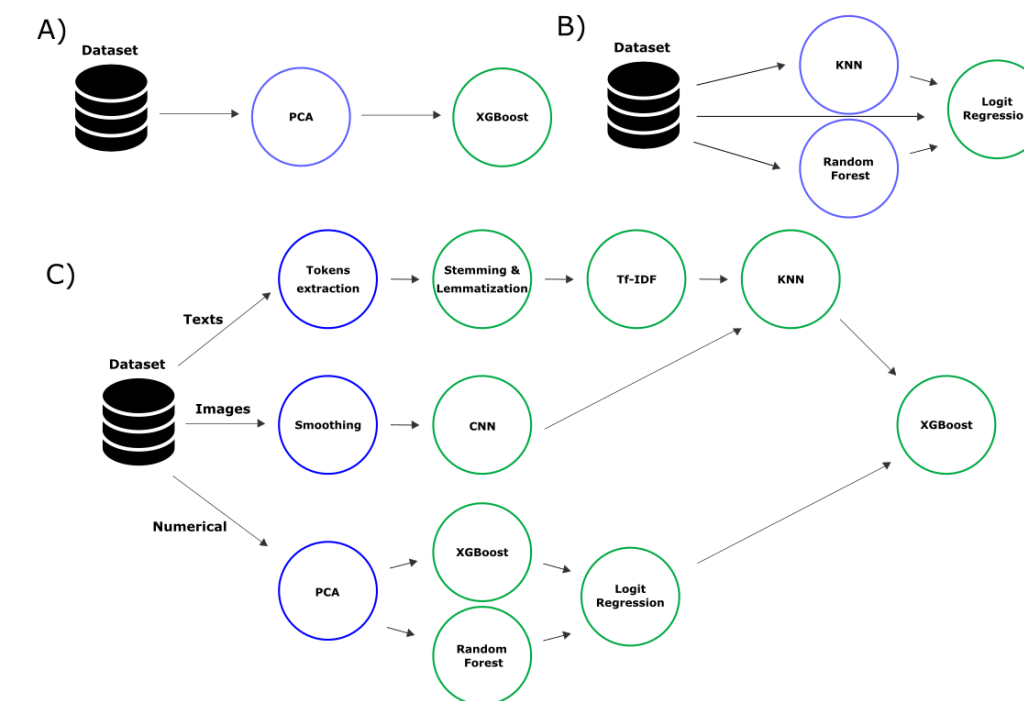
Масштабирование вычислений

<https://github.com/nccr-itmo/FEDOT>

Процедуры FEDOT вызываются из DataMall



FEDOT в контексте classic ML-pipeline



Типы пайплайнов, поддерживаемые FEDOT



Генеративный дизайн

13

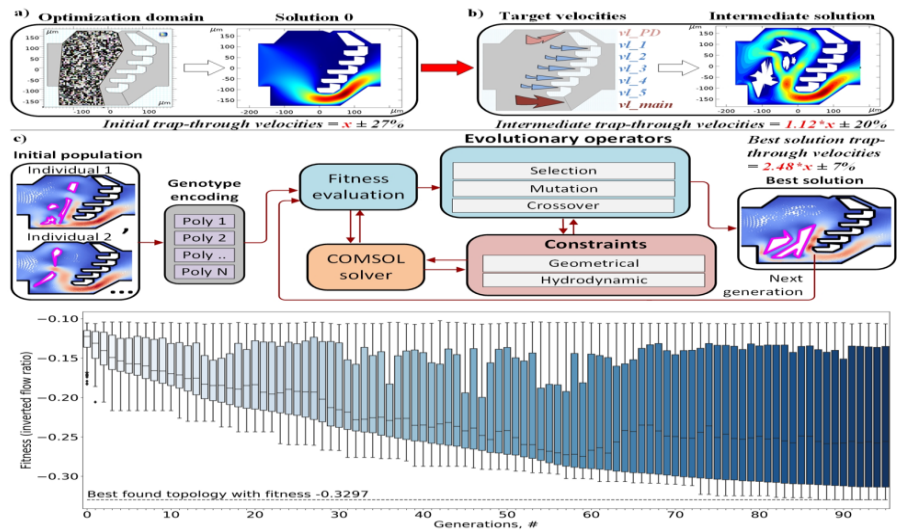
Генеративный дизайн – это подход, в рамках которого создается система усиленного интеллекта для специалиста (конструктора/дизайнера) и который позволяет создавать целую популяцию корректных решений для поставленных задач

От микро-масштаба:

Задача: проектирование микро-флюидных устройств для очистки крови.

Методы: алгоритмы генетического программирования для создания вариативной топологии препятствий, создающих условия для формирования потока крови с заданными характеристиками.

Результаты: даже относительно простые алгоритмы позволили сравниться и превзойти результаты экспертного проектирования.

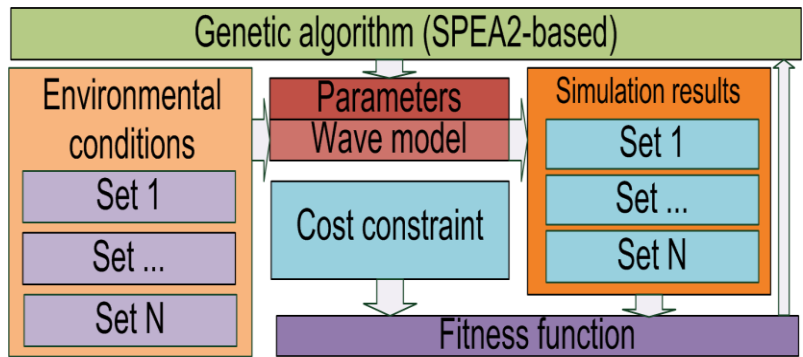


До макро-масштаба:

Задача: проектирование волнозащитных сооружений в порту.

Методы: многокритериальный генетический алгоритм для создания вариативной топологии волнозащитных сооружений, эффективность которых оценивается на основе волновой модели среды SWAN.

Результаты: реализованный многокритериальный алгоритм позволил автоматически разработать несколько вариантов конструкций, которые прошли проверку в испытательном бассейне.



Концептуальная схема применения многокритериального эволюционного алгоритма



Пример реализации: «выращенное» волнозащитное сооружение



КОМАНДЫ И ЛАБОРАТОРИИ



Директор – генеральный конструктор НЦКР
д.т.н. Бухановский Александр Валерьевич



Лаборатория
«Интеллектуальные технологии умного города»,
руководитель к.т.н. Митягин С.А.



Лаборатория
«Интеллектуальные технологии принятия решений»,
руководитель к.т.н. Иванов С.В.



Лаборатория «Алгоритмика сложных систем»,
руководитель к.т.н. Боченина К.О.



Лаборатория «Цифровое здравоохранение»,
руководитель к.т.н. Ковальчук С.В.

Лаборатория "Промышленный ИИ»,
руководитель к.т.н. Насонов Д.А.



Лаборатория
«Интеллектуальные технологии городского
планирования»,
руководитель Кудинов С.А.



Лаборатория «Моделирование
природных систем»,
руководитель к.т.н. Калюжная А.В.



Лаборатория «Когнитивная
невербалика»,
руководитель д.т.н. Басов О.О.



Лаборатория «Интерактивная визуализация»,
руководитель к.т.н. Карсаков А.С.

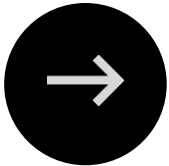


Национальный центр когнитивных разработок Университета ИТМО — один из лидирующих R&D центров в России, специализирующийся на машинном обучении и когнитивных технологиях и их внедрении в реальные секторы экономики.

НАПРАВЛЕНИЯ РАБОТЫ:

- когнитивная невербалика,
- алгоритмика сложных систем,
- технологии управления высокотехнологичными производствами,
- AutoML,
- AI для AR и VR,
- process mining + text mining ,
- цифровая урбанистика,
- Big Data

Более **200** научных сотрудников
Более **100** научных исследований и разработок
Более **150** успешно реализованных проектов
Более **50** патентов, ноу-хау, ОИС



Спасибо за ВНИМАНИЕ

+7 (812) 909 3156

nccr@itmo.ru

<https://actcognitive.org>

