



iPavlov Artificial Business Intelligence
(ABI) Platform - Smart Search

iPavlov.ai

О проекте

Функции:

- Кластеризация текстовых коллекций
- Маршрутизация текстовых коллекций
- Умный поиск информации

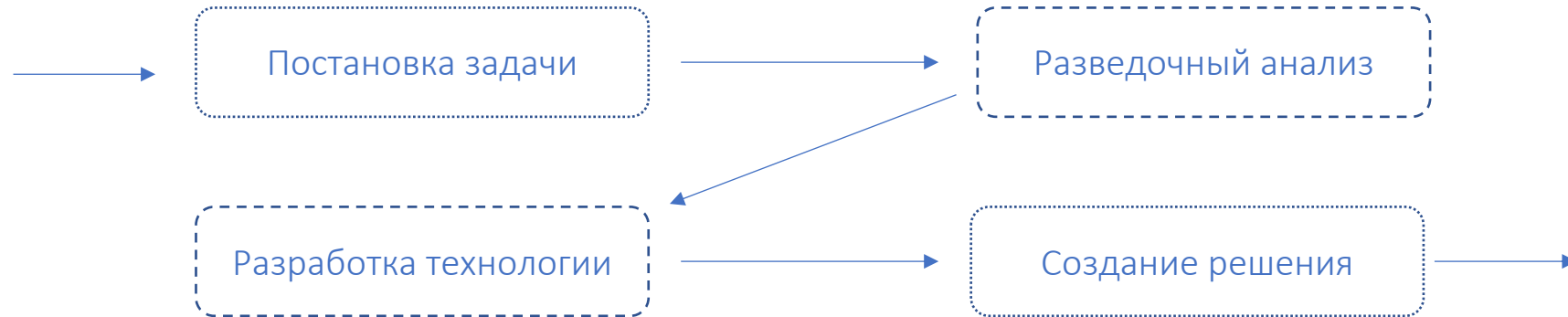
Проект построен на основе библиотеки [TopicNet](#) (библиотека автоматизированного тематического моделирования)

Преимущества TopicNet:

- Содержит функционал необходимый разработчику
- Позволяет использовать сложные сценарии обучения и подготовки моделей
- Возможность встраивать собственные критерии качества для обучения моделей
- Возможность встраивать собственные регуляризаторы как этапы обучения модели
- Построение базовых решений повышенной точности
- Автоматизация стандартной стартовой работы исследователя в области анализа и обработки данных
- Минимизация затрат в предметную область



Актуальность проекта



- Снижение «рутинной работы»
- Снижение процента ошибок, обусловленных человеческим фактором
- Отсутствие технологических аналогов на рынке SDK продуктов
- Современная тенденция быстрого получения информации
- Бесперебойность работы
- Возможность создания под индивидуальный запрос



История

МФТИ выиграл грант на создание ИИ-проекта «Нейроинтеллект iPavlov».



Постепенное разделение проекта на 2 равноценные команды:
Команда коммерциализации РИД (Лоран Аюпян)
Лаборатория: исследовательская составляющая (Михаил Бурцев)



Подписано значимое соглашение с руководством холдинга «Швабе» (Госкорпорации «Ростех»), в рамках которого был создан НИЦ АО "Швабе" в МФТИ.



Реализация проекта iPavlov в рамках НТИ, при индустриальной поддержке Сбербанка. Главная цель проекта: создание open-source библиотеки DeepPavlov, доступной для скачивания по всему миру.



В результате успешных переговоров с Дмитрием Дэйн, сооснователем и техническим директором компании Virgil Security, Inc. (Manassas, VA, USA) была открыта международная лаборатория интеллектуальных криптографических систем в рамках ЦК НТИ по направлению «Искусственный интеллект» на базе МФТИ.



Цель: коммерциализация новых продуктов и технологий, улучшение качества и эффективности научно-исследовательских и опытно-конструкторских работ между лабораториями МФТИ и промышленными предприятиями Холдинга.

2017

2017
2020

2018

2019

2019

2019

2018
2020

2020

2020

2020

2020

Реализован широкий пласт работ по привлечению партнеров, клиентов и инвесторов (в том числе: Альфа-Банк, ПАО «МегаФон», АО ПО УОМЗ, ПАО «Ростелеком», SAP, Медиалогия, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России, ПАО «РГС Банк» и др.)

Подписание трехстороннего соглашения о совместном развитии высокотехнологичной области искусственного интеллекта (ИИ) | РФПИ, МФТИ и Национальным центром больших данных и искусственного интеллекта имени Короля Абдель Азиза Королевства Саудовская Аравия.



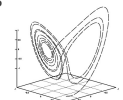
Выпущена первая версия рыночного продукта «Платформа создания мультифункциональных чат-ботов iPavlov», а также первая версия ПО для многофункционального медицинского терминала.



На заседании наблюдательного совета АСИ, президент и председатель правления Сбербанка России Герман Греф доложил об успехах iPavlov.

В ходе заседания проектного комитета НТИ было рекомендовано проработать состав учредителей ООО «Айпавлов».

МФТИ принял стратегически-важное решение о создании спин-офф компании ООО «Айпавлов», для продолжения коммерциализации ИИ-разработок и вывода их на глобальные рынки.



iPavlov: 5 направлений развития компании

NLP

ЦА для автоматизации HR процессов

Маршрутизатор запросов

Сервис по анализу транзакционных данных

Модели intent recognition, NER, sentiment analysis, etc.

Smart City

Интеллектуальные дорожно-транспортные системы

Решения «Умное ЖКХ»

Решения «Обработка обращений населения»

Решения «Общественная безопасность»

Smart Clinic

Создание антропоморфных роботов-ассистентов

Система анализа медицинских карт и изображений

Цифровой комплекс нейрореабилитации больных

Автоматизированная система управления потоками клиентов

ABI

Сервис по анализу временных рядов

Разработка ПО для телеметрических комплексов различных технических и «умных» объектов

Сервис по анализу мультимодальных неструктурированных данных

CV

Сервис по распознаванию объектов и их сегментации

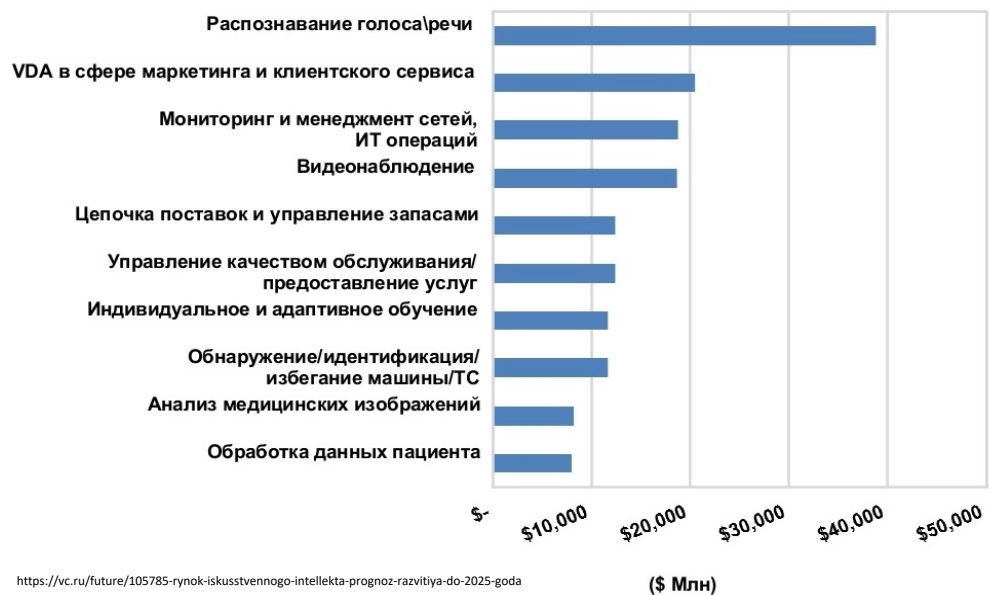
Сервис по распознаванию и обработке документов

Система компьютерного зрения в области розничной и оптовой торговли

Анализ отрасли цифровых ассистентов

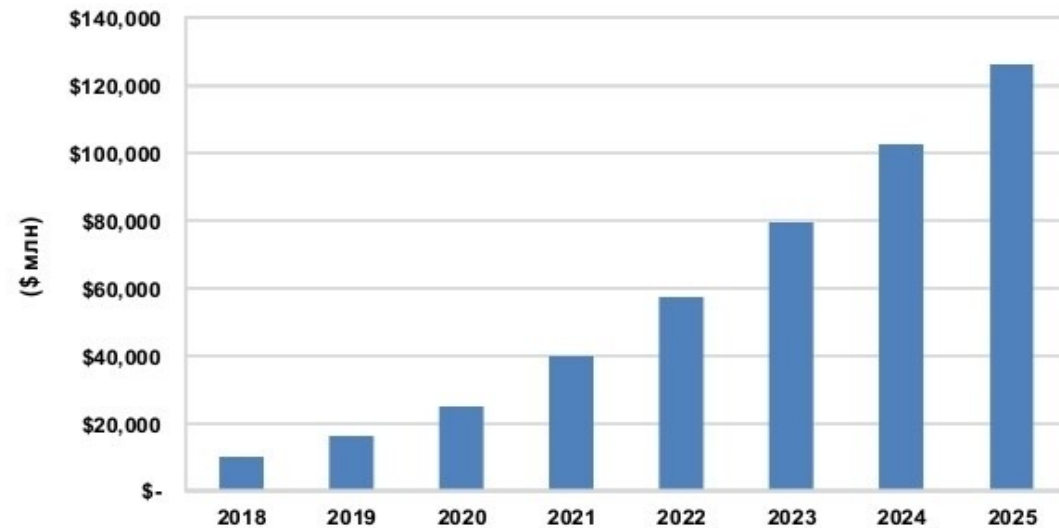
По исследованиям Gartner - в следующие три года все больше организаций станут использовать виртуальных помощников для сотрудников. Среди отраслей, наиболее заинтересованных во внедрении - страхование и сферу финансовых услуг. Также виртуальные ассистенты используются в ИТ-секторе, отделах клиентского обслуживания и информационных запросов

Таб. 1.2. Совокупный мировой доход от AI ПО, 10 лучших кейсов использования AI: 2018-25 гг



<https://vc.ru/future/105785-rynok-iskusstvennogo-intellekta-prognoz-razvitiya-do-2025-goda>

Таб. 1.1. Годовой доход от ПО с AI, мировой рынок: 2018-2025



<https://vc.ru/future/105785-rynok-iskusstvennogo-intellekta-prognoz-razvitiya-do-2025-goda>

Прогнозируется, что ежегодный глобальный доход от программного обеспечения (ПО) вырастет с 10,1 млрд долларов в 2018 году до 126 млрд. долларов в 2025.

Цифровые ассистенты по прогнозам принесут около 4,8 млрд. долларов годовой прибыли в 2025 по сравнению с всего 876,3 млн. долларов в 2018.

Цель и задачи проекта

Цель:

Создание библиотеки TopicNet - автоматического построения тематических сетей в прикладных задачах текстовой аналитики и анализа финансовых транзакционных данных. Это необходимо для снижения порога входа при использовании иерархических интерпретируемых тематических моделей в прикладных задачах и получения автоматизировано-настраиваемых базовых моделей высокого качества.

Задачи:

- Построение иерархических тематических моделей по большим динамически обновляемым текстовым или транзакционным данным
- Рубрикация данных контактного центра: выделение интенгов в текстовых обращениях клиента.
- Векторизация юридических лиц – клиентов банка на основе информации об их транзакционной активности с последующим поиском потенциальных конкурентов, партнеров, поставщиков и клиентов.



Развитие проекта

| Этап | Перечень работ | Результаты |
|--|--|---|
| Этап 1: текстовые и транзакционные данные | <ul style="list-style-type: none"> Создание библиотеки TopicNet - автоматизированный тематический анализ | <ul style="list-style-type: none"> Первая версия библиотеки Topic Net, документация по ее использованию |
| Этап 2: мультимодальные данные: изображения, временные ряды, тексты, транзакции. Внутренние библиотеки | <ul style="list-style-type: none"> Создание части библиотеки TopicNet с поддержкой английского языка, визуализацией результатов; Первая версия анализа изображений печатных документов | <ul style="list-style-type: none"> Вторая версия библиотеки Topic Net, документация по ее использованию: обработка документов |
| Этап 3: развитие внутренней библиотеки и реализация первых продуктов на ее основе | <ul style="list-style-type: none"> Создание библиотеки предобученных моделей Создание внутренних технологических составляющих ряда продуктов: информационный поиск, рубрикация диалогов, сегментации и профилирования клиентов | <ul style="list-style-type: none"> Третья версия библиотеки с предобученными моделями, документация по ее использованию. Внутренний технологический стек для сервисов. |
| Этап 4: улучшение продуктов, реализация новых на их основе. Добавление нового функционала в библиотеку | <ul style="list-style-type: none"> Создание сервисов: выделения тем диалога, профилирования клиентов / спецификации | <ul style="list-style-type: none"> Готовая для развертывания и обучения на стороне заказчика система. Документация по ее обучению и использованию |
| Этап 5: новые продуктовые направления | Создание сервисов: <ul style="list-style-type: none"> вопросно-ответная система dewarping документа профилирование пользователя | <ul style="list-style-type: none"> Готовая для развертывания и обучения на стороне заказчика система. Документация по ее обучению и использованию |
| Этап 6: новые продуктовые направления | <ul style="list-style-type: none"> Создание библиотеки локально аппроксимирующих моделей временных рядов | <ul style="list-style-type: none"> Четвертая версия библиотеки Topic Net с частью обработки временных рядов Документация по ее использованию |
| Этап 7: новые продуктовые направления (интеграция мультимодальных данных в одну систему) | <ul style="list-style-type: none"> Упаковка сервисов в продуктовые линейки | <ul style="list-style-type: none"> Готовая библиотека для масштабирования проектной деятельности, документация по ее использованию Готовые для развертывания и обучения на стороне заказчика сервисы и документация по их обучению Свидетельство о регистрации РИД |

Реализация проекта



Этап 3-4. Мониторинг и контроль разрабатываемой платформы.
Ожидание сдачи всего функционала согласно ТЗ, доработка индивидуальных потребностей



РОСАТОМ

Этап 3. Мониторинг и контроль разрабатываемой платформы. Ожидание сдачи всего функционала согласно ТЗ



Этап 1-2. Индустриальный партнер, заказчик технологии, договор заключен, проект сдан, оплата за проект получена



Этап 1-2. Индустриальный партнер, заказчик технологии, договор заключен, первый этап сдан, оплата за первый этап получена



Этап 1-2. Индустриальный партнер, заказчик технологии, договор заключен, первый этап сдан, оплата за первый этап получена



Предполагаемые результаты

- Библиотека TopicNet по автоматизированному созданию тематических сетей для масштабирования проектной деятельности



- Адаптированная и апробированная система анализа покупательской активности клиентов банка.



- Адаптированная и апробированная система анализа и рубрикации диалогов контактного центра.



- Адаптированная и апробированная система анализа внутренних процессов персонала



- Адаптированная и апробированная система поиска по большому количеству данных. Распознавание запросов на естественном языке и свободной форме



- Лицензируемый модуль TopicNet для капитализации и коммерциализации научных исследований по созданию тематических сетей



Сравнение с аналогами по конкретным результатам

| | Gensim LDA | SkLearn LDA | VW LDA | TopicNet |
|---|------------|-------------|---------|-----------|
| Автоматизированное построение модели | Нет | Нет | Нет | Да |
| Возможность построения иерархий тем | Нет | Нет | Нет | Да |
| Скорость обработки коллекции | Высокая | Низкая | Высокая | Высокая |
| Возможность выбора стратегии обучения тематической модели | Нет | Нет | Нет | Да |
| Число потоков построения модели | 1 | 1 | 1 | 2 и более |
| /Зависит от коллекции/ Декоррелированность тем | < 0.7 | < 0.7 | < 0.7 | < 0.9 |
| /Зависит от коллекции/ Доля интерпретируемых тем | До 70% | До 70% | До 70% | До 90% |
| /Зависит от коллекции/ Разреженность тем | < 0.6 | < 0.6 | < 0.6 | < 0.9 |
| Число модальностей | 1 | 1 | 1 | 5 и более |

Научная база проекта

- 1 Open Source библиотека Topic Modeling: BigARTM
- 5 Области внедрения методов AI: банкинг, телеком, документооборот, гос. корпорации, архив
- 3 Ключевых партнера в области Artificial Intelligence
- 5 Успешно завершенных проектов в области Artificial Intelligence
- 8 Докладов на международных конференциях
- 18 Публикаций в изданиях, рецензируемых Scopus, WoS

Ресурсы:

- 1 Библиотека обработки фотографий и сканов документов iDog
- 2 Системы разведочного анализа текстовых и транзакционных данных
- 2 Лицензируемых модуля предобработки и анализа текстовой информации



Лаборатория машинного интеллекта МФТИ ведет постоянную работу в направлении создания технологических компонента для последующей капитализации результатов научной деятельности и их коммерциализации

Коммерциализация результатов

Результаты первого этапа:

- Масштабирование проектной деятельности на новых клиентов
- Увеличение потока финансирования лаборатории
- Зарегистрированный РИД в формате лицензируемого кода, распространяющегося по лицензионным договорам среди заинтересованных технологических команд и компаний

Результаты последующих этапов:

- Создание отдельных сервисов и продуктов промышленными партнерами или компаниями, представляющими spin-off Лаборатории машинного интеллекта

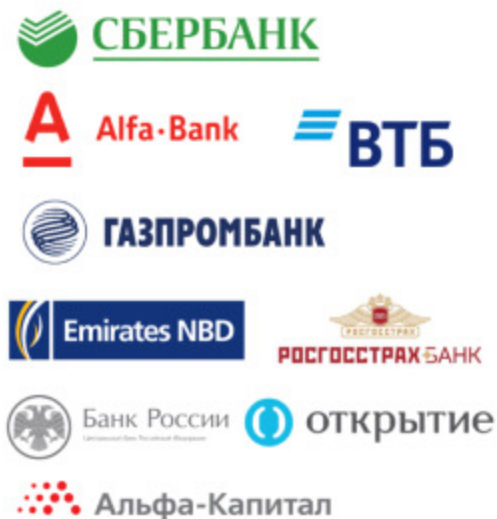
Целевая аудитория:

- Банки
- Телеком
- Медицина
- Гос. Учреждения
- Архивы
- Музеи
- Образовательные учреждения

Telecom



Банковский сектор



IT компании/ retail



Прочее



Контактная информация

Адрес: Россия, 141701, Московская обл.,
г. Долгопрудный, Институтский переулок 9 с.3,
МФТИ, корп. «Физтех.Цифра»

Тел: +7 (498) 713-91-67

Почты:

info@ipavlov.ai

pr@ipavlov.ai

partner@ipavlov.ai

hr@ipavlov.ai

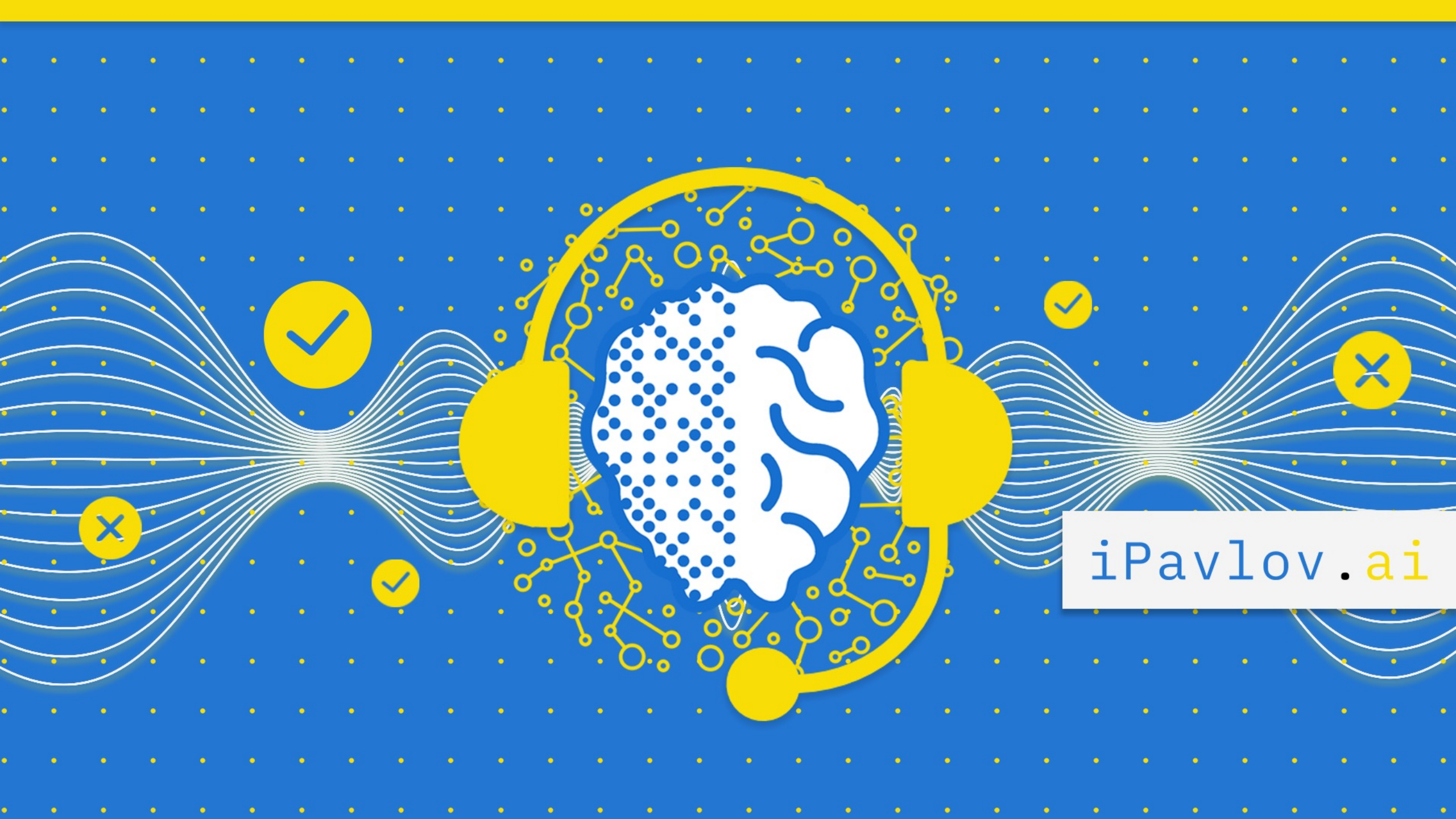
shvabe.ai@mipt.ru

iPavlov.ai



НИЦ АО «Швабе» в МФТ





iPavlov.ai