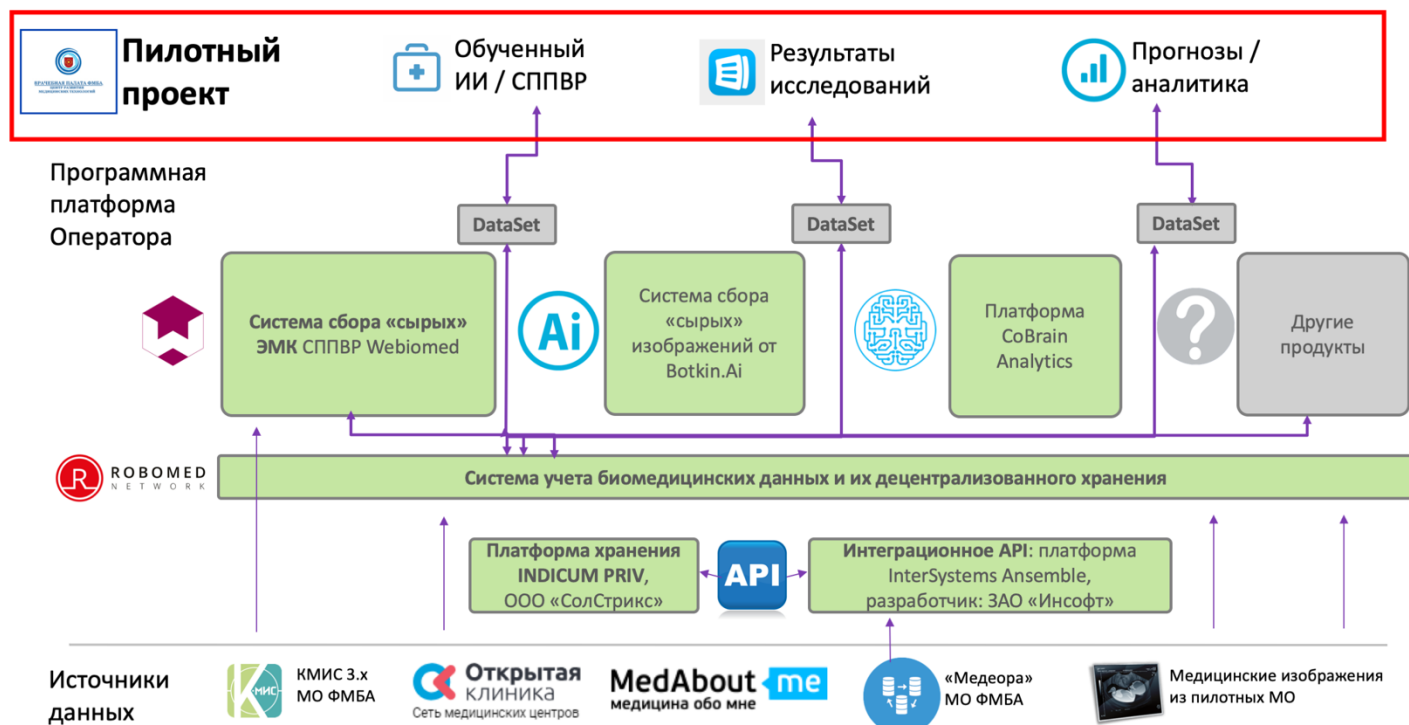


# Ассоциация «Национальная база медицинских знаний» совместно с «Врачебной палатой ФМБА» завершили пилотный проект создания оператора биомедицинских данных

Ассоциация «Национальная база медицинских знаний» совместно с АНО «Врачебная палата Федерального Медико-Биологического Агентства» (ФМБА России) в рамках созданного в начале 2019 г. консорциума реализована пилотный этап создания «Национального оператора биомедицинских данных». В состав консорциума, кроме НБМЗ и «Врачебной палаты», вошли ряд ИТ-компаний и разработчиков систем искусственного интеллекта.

Целью консорциума была практическая проверка возможности создания национального банка биомедицинских данных, которые можно было бы собирать из самых разнообразных источников данных и затем использовать для производства дата-сетов и машинного обучения, которые как известно являются одной из главных задач на пути создания в России собственного рынка искусственного интеллекта для медицины.

Архитектура пилотного проекта представлена на рис. ниже.



В качестве поставщика системы хранения биомедицинских данных и предоставления аппаратных мощностей выступила компания «СолСтрикс», <http://solstrix.ru/>.

В качестве платформ для сбора деперсонифицированных «сырых» медицинских данных с последующим производством дата-сетов из них применялись:

- Система **Webiomed.EHR**, предназначенное для сбора информации из электронных медицинских карт, <http://webiomed.ai>
- Система **Botkin.Ai**, предназначенная для сбора КТ легких и других медицинских изображений лучевой диагностики, <https://botkin.ai>
- Платформа **CoBrain Analytics**, предназначенная для обработки медицинских изображений, <https://cobrain.io/>

В качестве интеграционной шины применяется решение компании «**Инфсофт**».

Поставщиками медицинских данных выступили свыше 12 медицинских организаций, подчиненных ФМБА России и использующих медицинские информационные системы «**КМИС**» и «**Медео**ра». Также поставщиком данных согласилась выступить сеть медицинских центров «**Открытая клиника**».

В качестве биллинговой системы, учитывающей вклад поставщиков медицинских данных в накоплении единого Датасета, использовалось решение **Robomed Network**, <https://robomed.io/ru/>.

В результате **реализации пилотного проекта** на базе АНО «Врачебная палата ФМБА» компанией «СолСтрикс» был создан центр обработки данных, в котором установлено программное обеспечение Webiomed. С его помощью участники проекта осуществляли сбор деперсонифицированных медицинских данных для последующей их обработки, создания DataSet-ов и проведения машинного обучения.

Компанией «Инсофт» для консолидации данных была разработана интеграционная шина. Из участвующих в проекте МО была выполнена передача деперсонифицированных медицинских данных.

Во время сбора медицинских данных протестирована работа системы Robomed Network. С ее помощью собран реестр метаданных с различных медицинских учреждений и реализован поиск по этим данным, проверен механизм оценки содержимого медицинских данных как с точки зрения полноты заполнения медицинской документации, так и с точки зрения своевременности и точности работы медицинских сотрудников, проверена возможность прицельного удаленного доступа в хранилища медицинских данных, находящихся в медицинских учреждениях, через конкретные выборки из реестра метаданных. Реализована экономика обмена медицинскими данными, либо любыми производными пакетами данных между участниками сети с последующей фиксацией в платежном шлюзе.

Выполнена трехсторонняя интеграция МИС, сервисов MedAboutMe и объектного хранилища с использованием общей интеграционной шины. На первом шаге через специализированный сервис интеграционной шины в MedAboutMe из МИС переданы необходимые экземпляры СЭМД. На втором шаге через другой специализированный сервис шины MedAboutMe передал экземпляры СЭМД в хранилище. Таким образом, MedAboutMe продемонстрировал возможность стать еще одним источником данных, приходящих в хранилище через шину.

В итоге в тестовой версии платформы были накоплены сведения о 298 тыс. «цифровых двойников» пациентов, включая сведения о случаях обследования и лечения, анамнезе пациентов, зарегистрированных окончательных диагнозов и медицинских протоколах в формате структурированных электронных медицинских документов (СЭМДов).

Выполнена разметка и верификация собранной информации. В качестве пробы создана демонстрационная версия DataSet-a, содержащего данные об инфарктах и инсультах. На основании этого дата-сета проведено пробное машинное обучение, целью которого было создание математической модели, предсказывающей наличие развивающегося инфаркта или инсульта у пациента, обратившегося в медицинскую организацию. Решение поставленной задачи было осуществлено методом классификации. Напомним, что классификация в машинном обучении — это задача отнесения объекта к одному из заранее определенных классов на основании его формализованных признаков. Каждый из объектов в этой задаче представлен в виде вектора в N-мерном пространстве, каждое измерение в котором представляет собой описание одного из признаков объекта. В нашем случае из ЭМК пациента отобран ряд признаков, которые размечены в демо-версии. В дальнейшем возможно научить систему извлекать эти признаки из ЭМК автоматически, в том числе с помощью NLP-методов с алгоритмам машинного обучения. В ходе пилотного проекта проверено 6 различных алгоритмов машинного обучения (ML) и глубокого обучения (DL) для определения лучшего алгоритма решения поставленной задачи. По результатам предварительного моделирования получена оценка F1, равная 90.1%, что свидетельствует о высокой точности выбора самих признаков и алгоритма машинного обучения «*k Nearest Neighbor*».

Представленные результаты доложены наблюдательному совету ассоциации с целью уточнения дальнейших шагов по реализации проекта, включая демонстрацию результата инвесторам и работу с государственными органами власти в части создания частно-государственного партнерства для реализации полноценной промышленной фазы проекта.