

Использование технологий искусственного интеллекта в ДОМ.РФ

Григорий Грязнов

*Руководитель, Данные и новые технологии АО «ДОМ.РФ»
Заместитель директора АО «Банк ДОМ.РФ»*

Лаборатория искусственного интеллекта ДОМ.РФ

В 2023 году АО «ДОМ.РФ» сформировано отдельное структурное подразделение «Искусственный интеллект» специализирующееся на внедрении ИИ-технологий в бизнес-процессы компании и отрасли жилищного строительства.

Ключевые направления деятельности подразделения 2024 г.

- 1 **Optical Character Recognition (OCR)** – перевод текстовых данных в машиночитаемый вид

- 2 **Named Entity Recognition (NER)** – поиск именованных сущностей в неструктурированных текстовых данных

- 3 **Computer Vision (CV)** – распознавание и классификация графических образов на изображениях

- 4 **AI-помощники** – системы помощи принятия решений

Структура команды

11 ШЕ *Текущая численность команды*

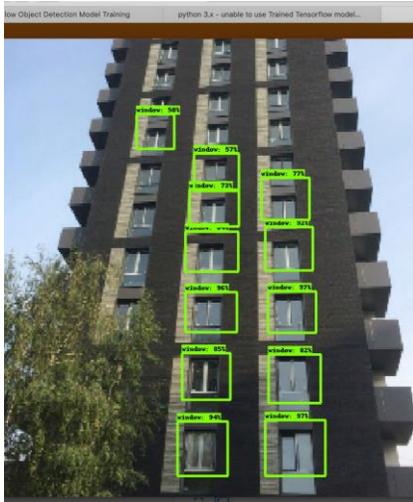


7 ШЕ *ML-инженеры*

4 ШЕ *Бизнес-анализ*

Примеры реализованных проектов

1 Анализ фотографий хода строительства



- Решение задачи **Image Detection**: определение ключевых элементов конструкции объекта
- **Image Comparison**: анализ темпов строительства с помощью сравнительного анализа фотографий
- **Стандартизация** ракурсов и качества изображений, загружаемых в систему
- **Обогащение** классических предиктивных моделей алгоритмами ИИ

2 Выявление именованных сущностей из нешаблонизированных документов



ручной разбор
документов



>10 часов на
анализ



вывод результата
за **20 минут**

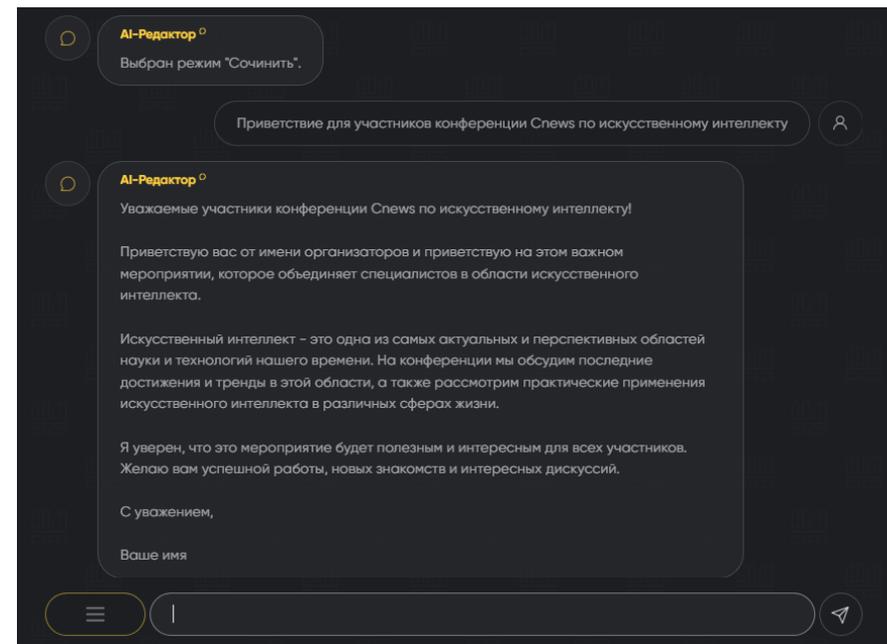
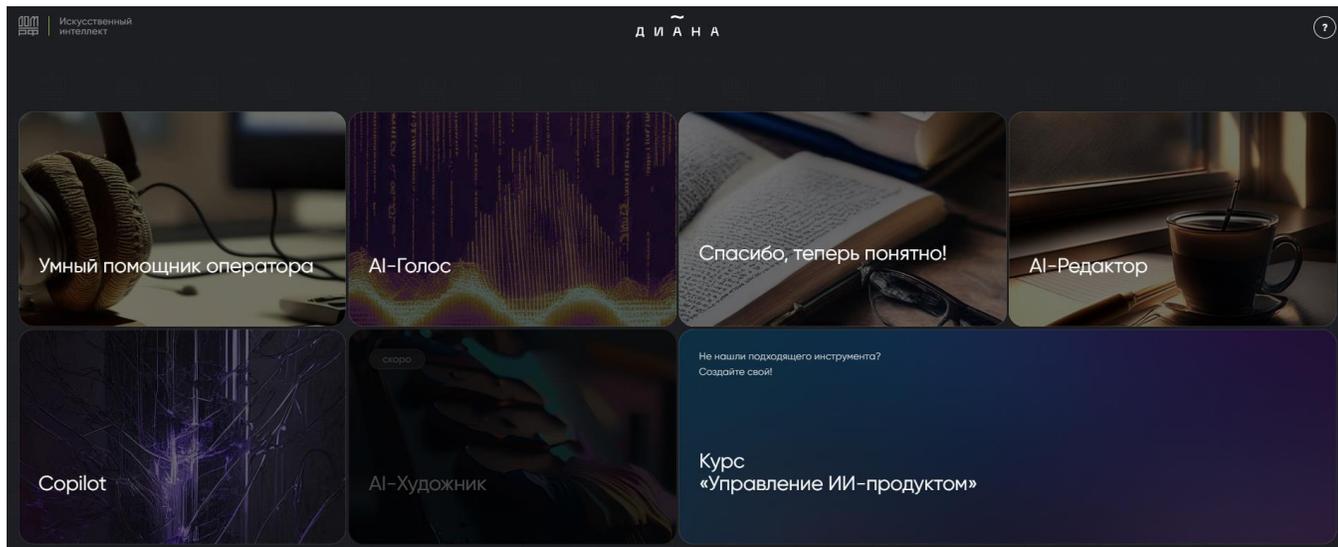


оптимизация
работы сотрудника

Прогноз Росстата по численности населения *(на начало года)*

Год	Низкий вариант прогноза		Средний вариант прогноза		Высокий вариант прогноза	
	тыс. человек	% от общей численности	тыс. человек	% от общей численности	тыс. человек	% от общей численности
2018	80 511.30	56.80	81 075.20	56.30	81 629.40	56.00
2019	79 581.00	56.40	80 296.50	55.80	80 999.80	55.30
2020	78 782.50	56.10	79 659.00	55.40	80 521.30	54.80
2021	77 966.10	55.80	79 009.80	54.90	80 037.50	54.30
2022	77 281.10	55.50	78 500.30	54.60	79 701.80	53.90
2023	76 681.50	55.40	78 083.60	54.40	79 466.60	53.60
2024	76 235.00	55.40	77 826.40	54.30	79 397.70	53.40
2025	75 934.10	55.50	77 722.20	54.20	79 489.40	53.30
2026	75 571.10	55.60	77 560.60	54.20	79 528.20	53.20
2027	75 264.40	55.80	77 462.70	54.30	79 636.90	53.10
2028	75 069.20	56.10	77 482.50	54.40	79 869.10	53.10
2029	74 956.00	56.40	77 588.60	54.50	80 191.20	53.20
2030	74 546.10	56.50	77 530.50	54.60	80 422.40	53.20
2031	73 944.40	56.50	77 311.10	54.60	80 546.40	53.10

AI-помощники



Базовый промт

Привет, у меня есть для тебя задание! У нас есть система чат-ботов, которая активирует ответ бота на основе определенных фраз. У каждого ответа бота есть несколько фраз, на которые он срабатывает.

Например, если в ответе есть информация о балансе, то ответ должен активироваться на целый набор фраз, например:

1. Покажи баланс
2. Как посмотреть баланс
3. Сколько денег на счету
4. Не могу посмотреть количество денег

Я буду писать тебе системное название ответа бота, которое чаще всего и содержит ключевые фразы. Твоя задача вычленить ключевые слова, а также составить словосочетания с возможными синонимами к ключевым словам, как я показал тебе выше в примерах про баланс и деньги.



Пользователь:

Для каких целей подходит эскроу?

GPT:

Ключевые слова и фразы:

Сферы применения эскроу
Цели использования эскроу-сервиса
Где применяется эскроу?
Области применения эскроу
В каких сделках используется эскроу
Почему выбирать эскроу-систему
Преимущества эскроу в различных областях
Эскроу в сфере недвижимости
Защита сделок через эскроу
Какие сделки подходят для эскроу?

Пользователь:

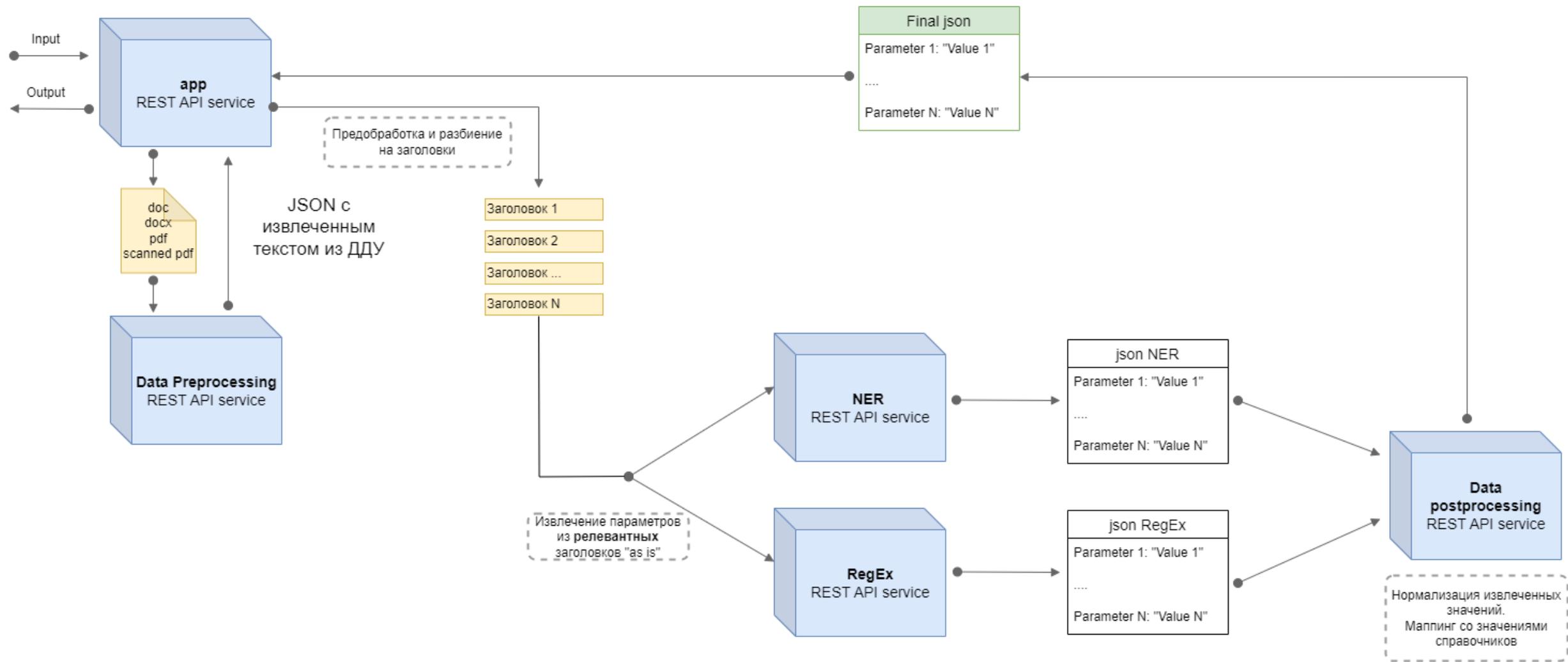
Статус справки?

GPT:

Ключевые слова и словосочетания:

Готовность справки
Проверить статус справки
Узнать, когда будет готова справка
Справка готова?
Получить информацию по справке
Что происходит с моей справкой в данный момент?
Какое текущее состояние моей справки?
Прошу сообщить о текущем прогрессе по справке
Пожалуйста, дайте мне актуальную информацию о статусе моей справки
Обновите меня по статусу моей справки

Извлечение сущностей из текстовых данных



Извлечение сущностей из текстовых данных

1



Bidirectional
Encoder
Representation
Transformer

Решаем задачи:

- **Masked Language Modeling (MLM)** – предсказание слова по контексту
- **Next sentence prediction (NSP)** – классификация, следует ли предложение за предыдущим

Пример:

Настоящий договор заключен между **MASK_1** и ООО «**MASK_2**»

2

RegExp – регулярные выражения, алгоритмической поиск в текстовых данных

```
/^([a-z0-9_\. -]+)@([a-z0-9_\. -]+)\.
```

3

Tesseract OCR – оптическое распознавание символов



Решаем задачи:

- Приведение к единому формату
- Возможность построения сквозных pipeline обработки данных

Извлечение сущностей из текстовых данных

Метрики качества (ground truth)

>80%
accuracy

Сущность	RegExp	ML	Result Accuracy
Стоимость	97%	99%	99%
Материал стен	0%	54%	54%
Материал перекрытия	0%	88%	88%
Фундамент	0%	89%	89%
Срок завершения строительства	100%	0%	100%
Этажность здания	0%	85%	85%
Этаж квартиры	97%	96%	97%
Количество комнат	82%	84%	84%
Общая площадь	77%	86%	86%
Жилая	100%	88%	100%
Кухня	100%	86%	100%
ИНН ЮЛ	100%	0%	100%
Адрес	61%	83%	83%
Номер помещения	100%	92%	100%
Кадастровый номер	100%	0%	100%

Обучение и методологическая поддержка

Внутренний курс для сотрудников

Программа обучения

Теория классического Машинного обучения

- Вводный практикум по языку программирования Python
- Основные задачи машинного обучения: классификация, регрессия, кластеризация.
- Обучение с учителем и без учителя.
- Классификация. Основные метрики.
- Решающие деревья
- Метод ближайших соседей (kNN)
- Линейный классификатор и логистическая регрессия
- Регрессия. Основные метрики.
- Метод наименьших квадратов. Линейная регрессия
- Нелинейная регрессия
- Отбор признаков. Метод понижения размерностей.
- Ансамблевые алгоритмы.
- Технология Bagging. Алгоритм случайного леса (Random Forest)
- Технология Boosting. Градиентный boosting.
- Обучение без учителя. Кластеризация и основные метрики.
- Нейронные сети и глубокое обучение.
- Алгоритмы NLP: tf-idf и токенизация.

Современные тренды искусственного интеллекта (ИИ)

- Генеративные алгоритмы (GPT). Интеллект или математическая модель? Где использовать и какие ограничения?
- GPT или BERT архитектура? В чем различия и кто победит?
- Как работают нейросети-художники MidJourney и DALL-E?
- Повышение эффективности бизнес-процессов за счет использования технологий ИИ.

Результаты



2
потока



1 038
слушателей



343
успешно завершили



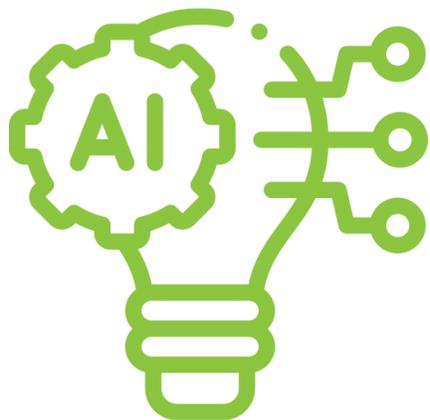
4.6
оценка слушателей

Экспертная группа Минстроя России по внедрению ИИ в строительной отрасли



АО «ДОМ.РФ» возглавляет экспертную группу по внедрению ИИ в отрасли

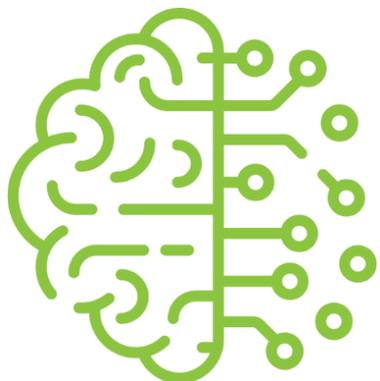
Ключевые направления деятельности:



- Разработка методологии оценки проникновения ИИ-технологий в отрасли
- Нормативно-правовое регулирование развития технологии
- Создание реестра отечественных ИИ-решений
- Обмен данными между участниками рынка
- Обучение и методологическая поддержка



stroi.ai@domrf.ru



Спасибо за внимание!



Григорий Грязнов

*Руководитель, Данные и новые технологии АО «ДОМ.РФ»
Заместитель директора АО «Банк ДОМ.РФ»*