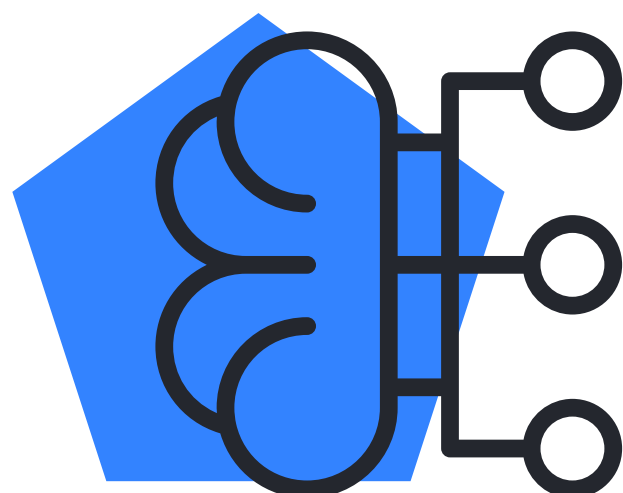




Цифровая трансформация





**Машинное обучение
теперь неотъемлемая
часть бизнеса
и без него «обойтись»
не выйдет**



**Основа обучения
модели – качественные
и объемные данные**



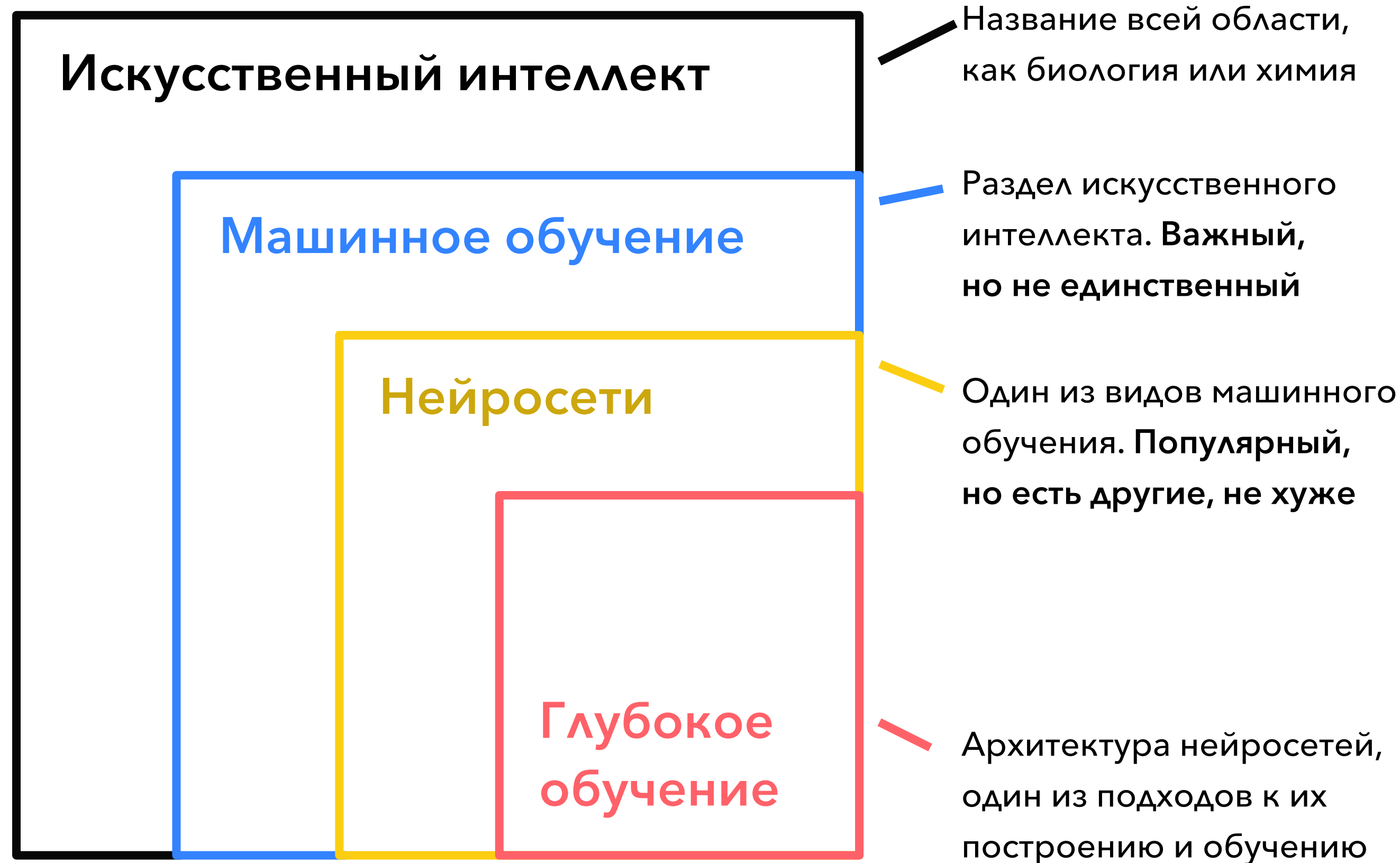
**Единственный способ
добиться успеха –
регулярно транслировать
результаты в мир**

Иначе не бывает

Сравнивать можно
только вещи одного
уровня, иначе получаются
изречения формата

«
Что лучше:
машина
или колесо?

Нельзя отождествлять
термины без причины!



Количество публикаций по машинному обучению растёт с каждым годом

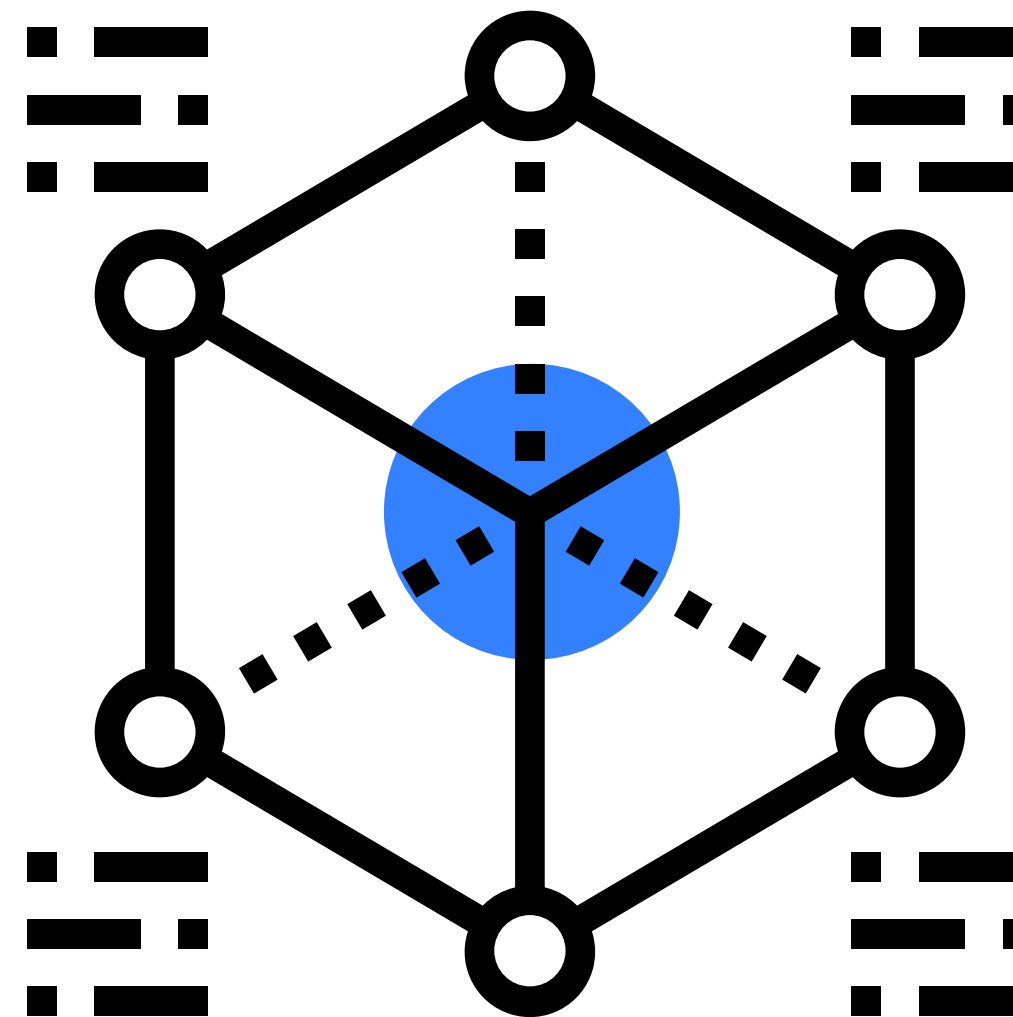


Задачи машинного обучения

Данные

Изображения
Видео
Тексты
Сигналы
Табличные данные
Среда, агент

Искусственный интеллект

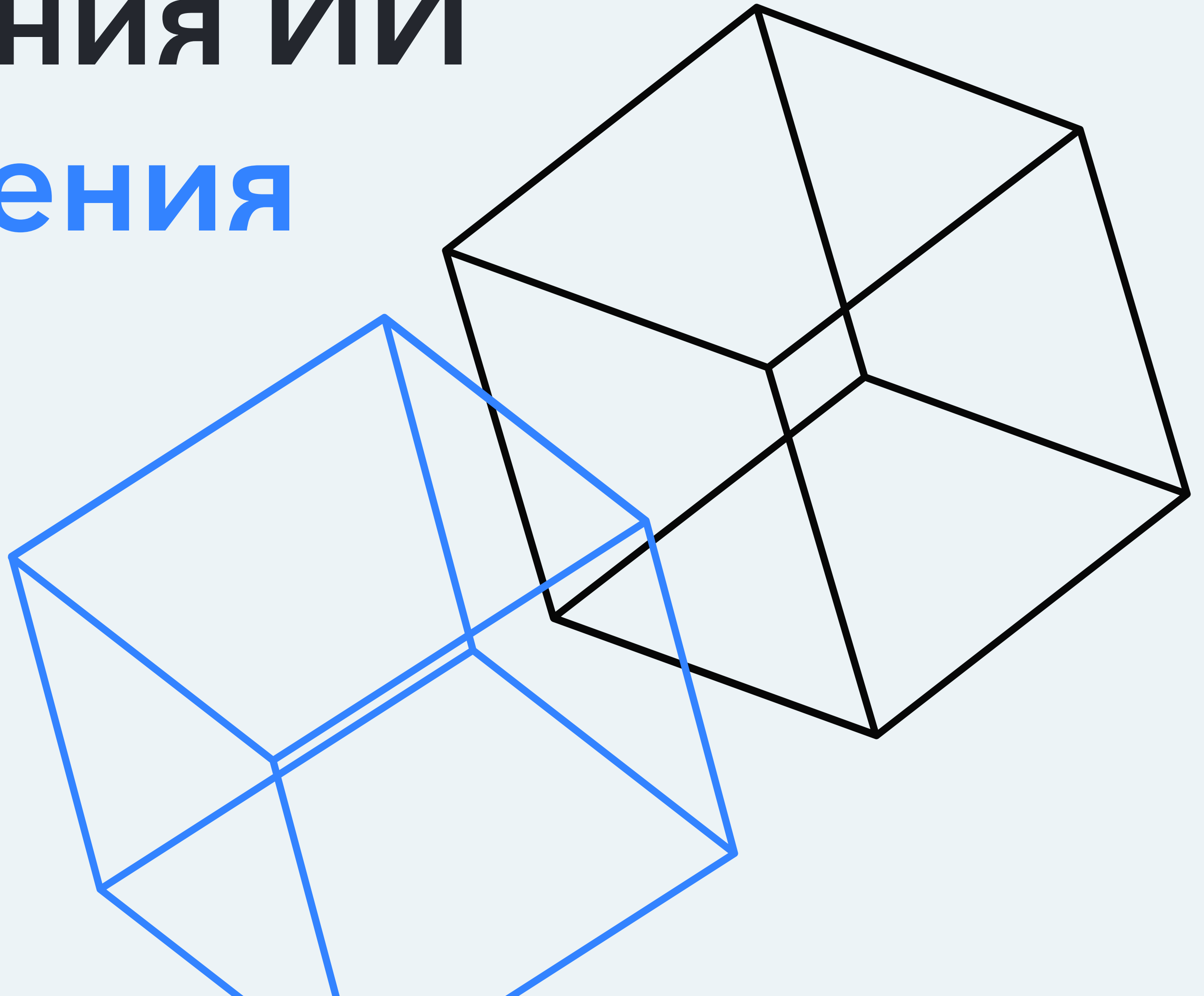


Решения

Классификация
Детекция
Сегментация
Генерация
Регрессия
Оптимальное
взаимодействие
со средой

Приложения ИИ

Изображения



Изображения Классификация



Аляскинский маламут



Сибирская хаски

Но если данных недостаточно...



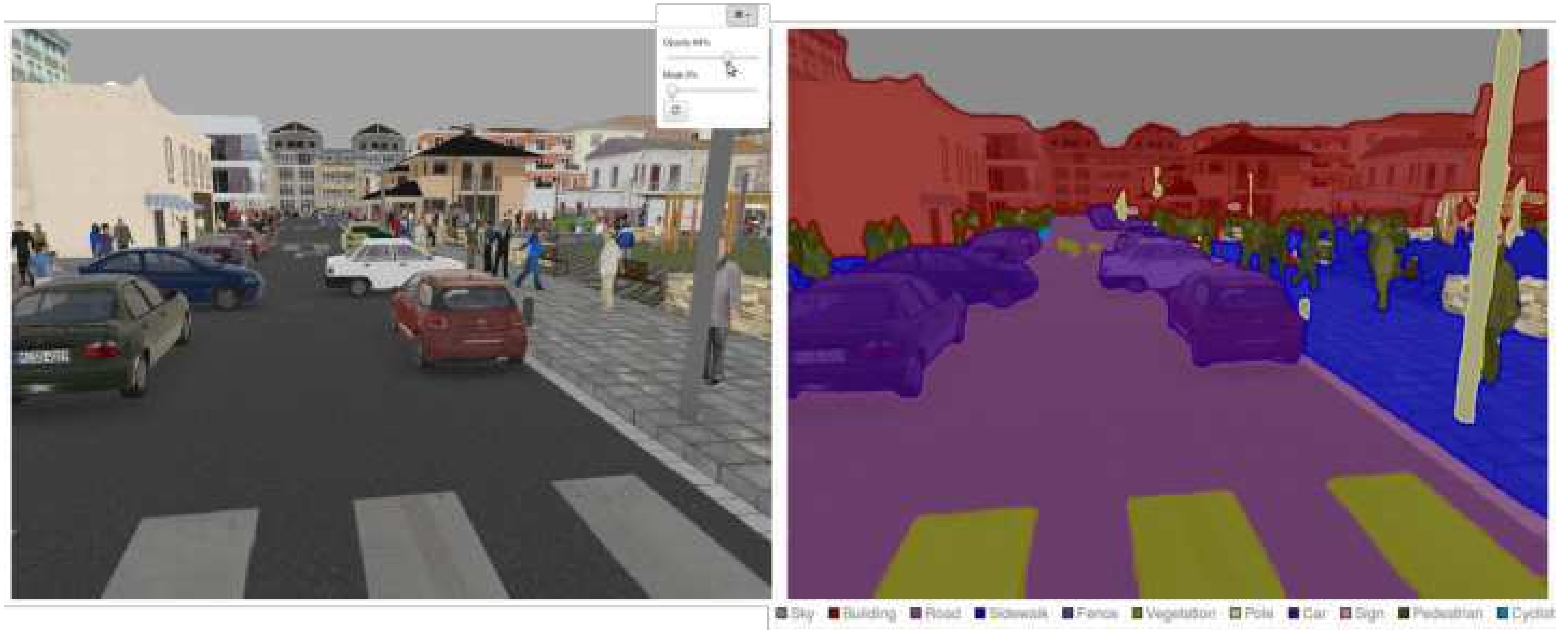
Изображения

Детекция



Изображения

Сегментация



Изображения Генерация



**Какая из двух
фотографий
настоящая?**

Изображения

Генерация видео



Изображения

Генерация подписей к изображениям



«мужчина в черной футболке играет на гитаре»



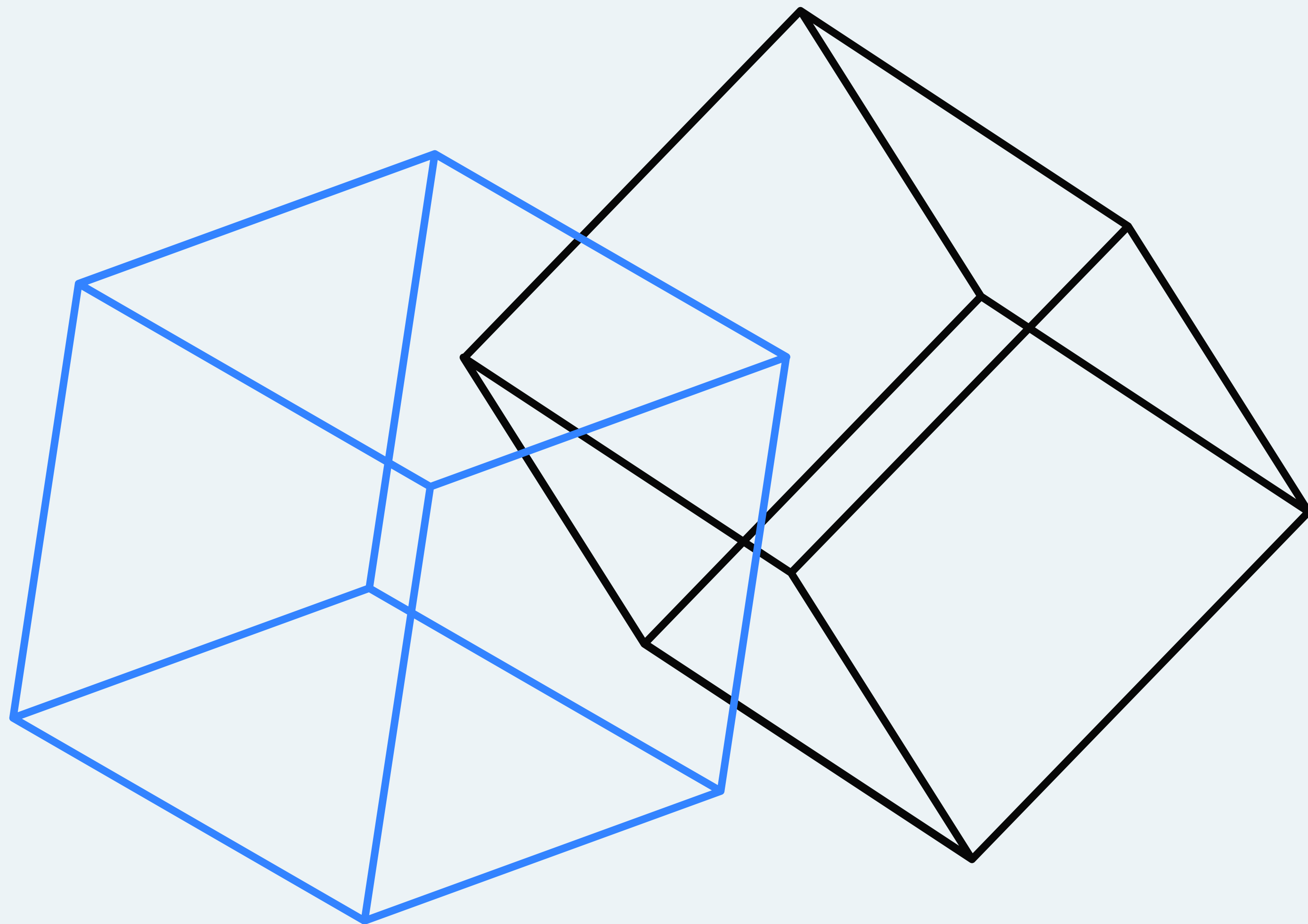
«строитель в оранжевом защитном жилете работает на дороге»



«две девочки играют с игрушкой лего»

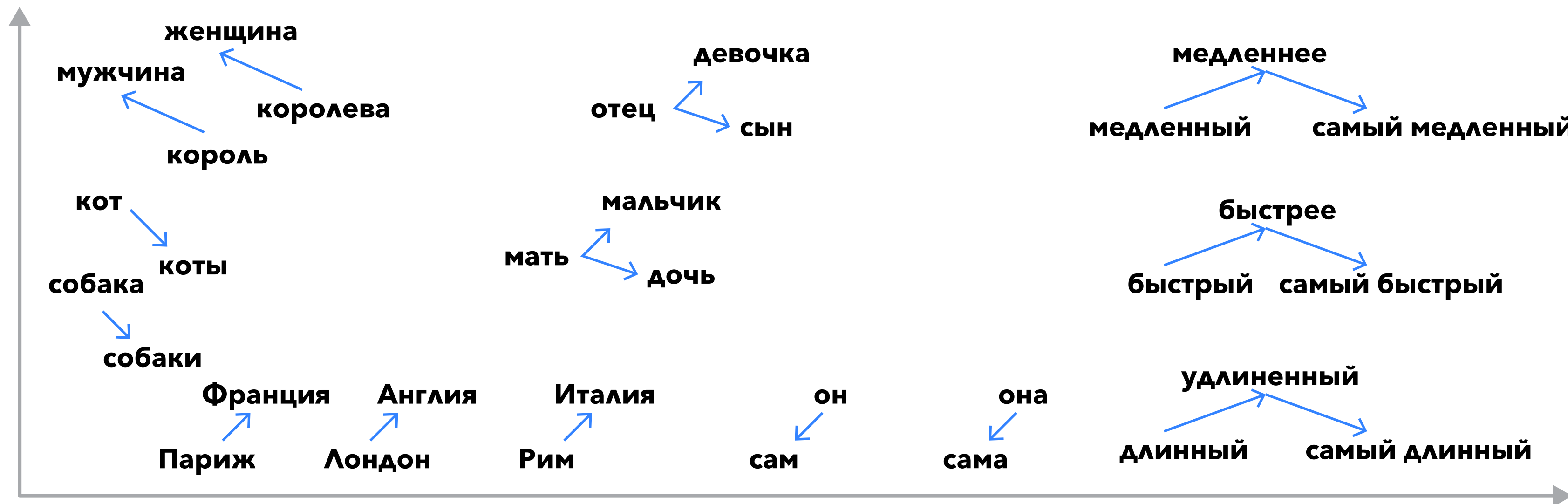
Приложения ИИ

Тексты



Тексты

Отображение слов в метрическое векторное пространство



король - мужчина + женщина = королева

доктор - мужчина + женщина = медсестра

женщина - мужчина = она - он

Тексты

Машинный перевод



It was a large lovely garden, with soft green grass. Here and there over the grass stood beautiful flowers like stars, and there were twelve peach-trees that in the spring-time broke out into delicate blossoms of pink and pearl, and in the autumn bore rich fruit. The birds sat on the trees and sang so sweetly that the children used to stop their games in order to listen to them. "How happy we are here!" they cried to each other.

№1

Это был большой прекрасный сад с мягкой зеленой травой. Тут и там по траве выдержал красивые цветы как звезды, и были двенадцатью персиковыми деревьями, которые весной вспыхнули в тонкие цветы гвоздики и жемчуг, и осенью принес богатые плоды. Птицы сидели на деревьях и пели так сладко, что дети привыкли для остановите их игры, чтобы слушать их. "Насколько счастливый мы здесь!" они кричали друг другу.

№2

Это был большой прекрасный сад с мягкой зеленой травой. Кое-где над травой стояли прекрасные цветы, похожие на звезды, и было двенадцать персиковых деревьев, которые весной расцветали нежными розовыми и жемчужными цветами, а осенью приносили богатые плоды. Птицы сидели на деревьях и пели так сладко, что дети останавливали свои игры, чтобы послушать их. «Как мы счастливы здесь!» - кричали они друг другу.

Какой перевод сделан нейросетью?

Тексты

Генерация текста

Заданная тема



Recycling is good for the world, no, you could not be more wrong

Текст, сгенерированный алгоритмом

"Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources."

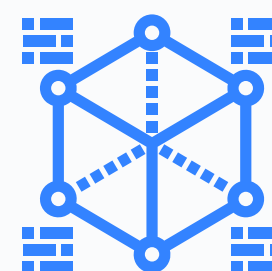
Speech-to-text

Распознавание речи

Input



Звуковой сигнал



Нейронная сеть



Текст

Output

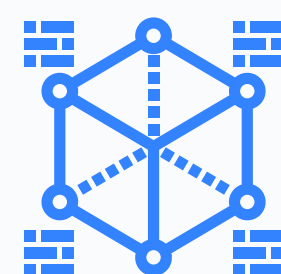
Speech-to-text

Генерация речи

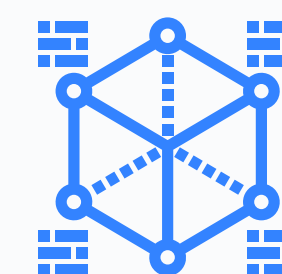
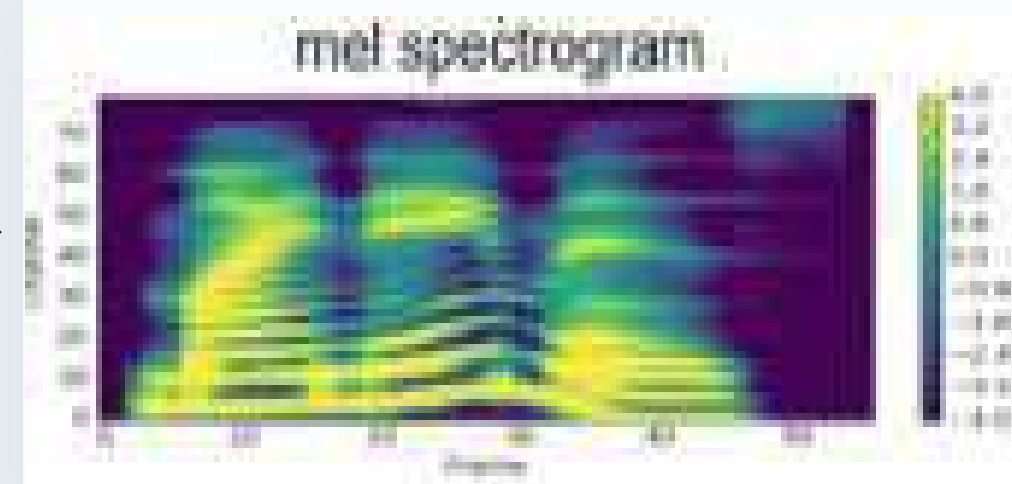
Input



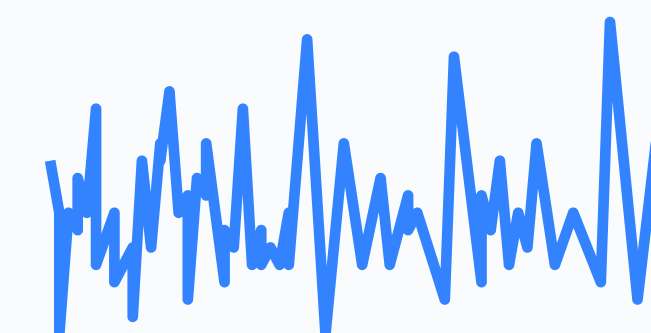
Текст



Нейронная
сеть 1



Нейронная
сеть 2

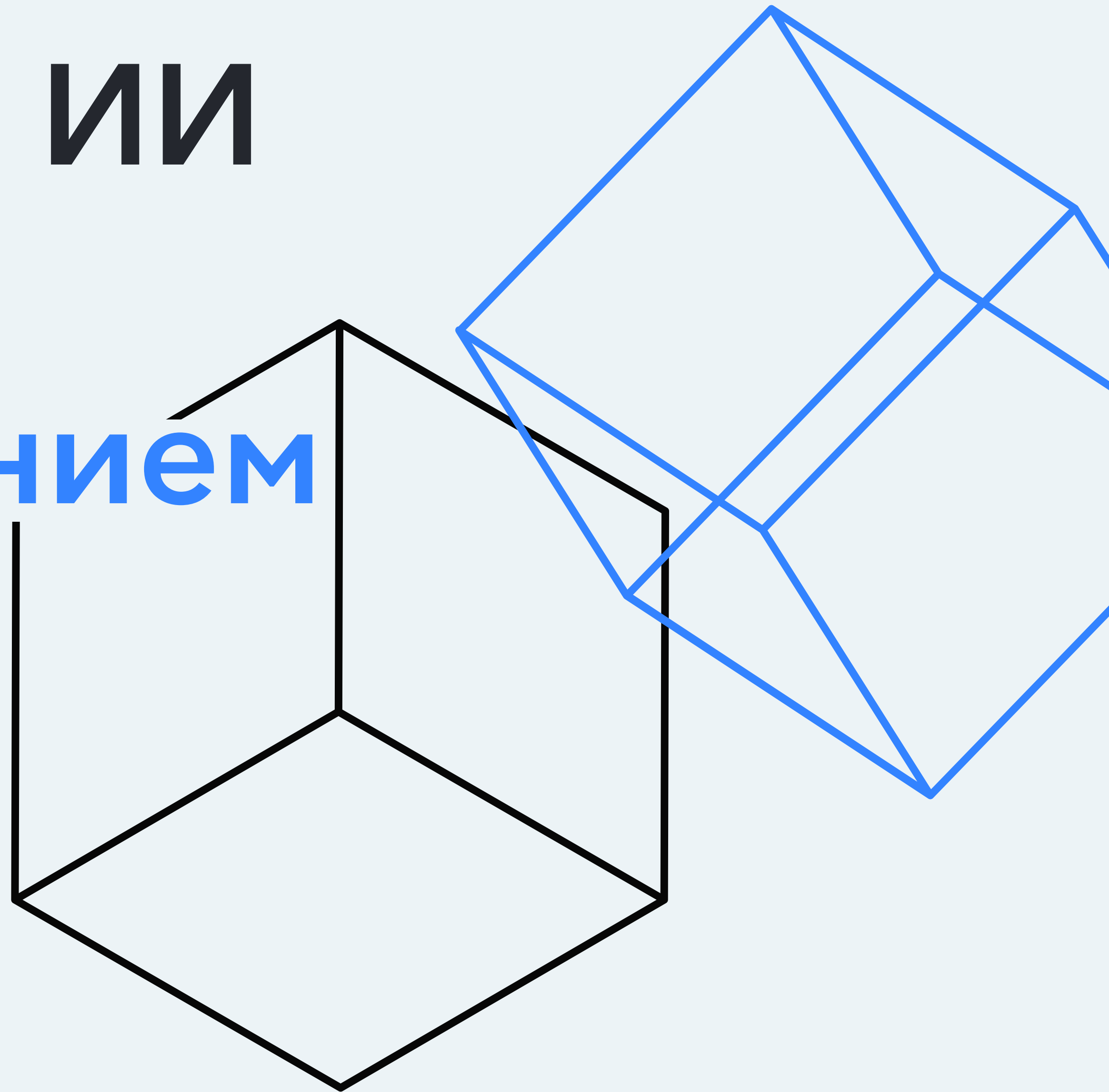


Звуковой сигнал

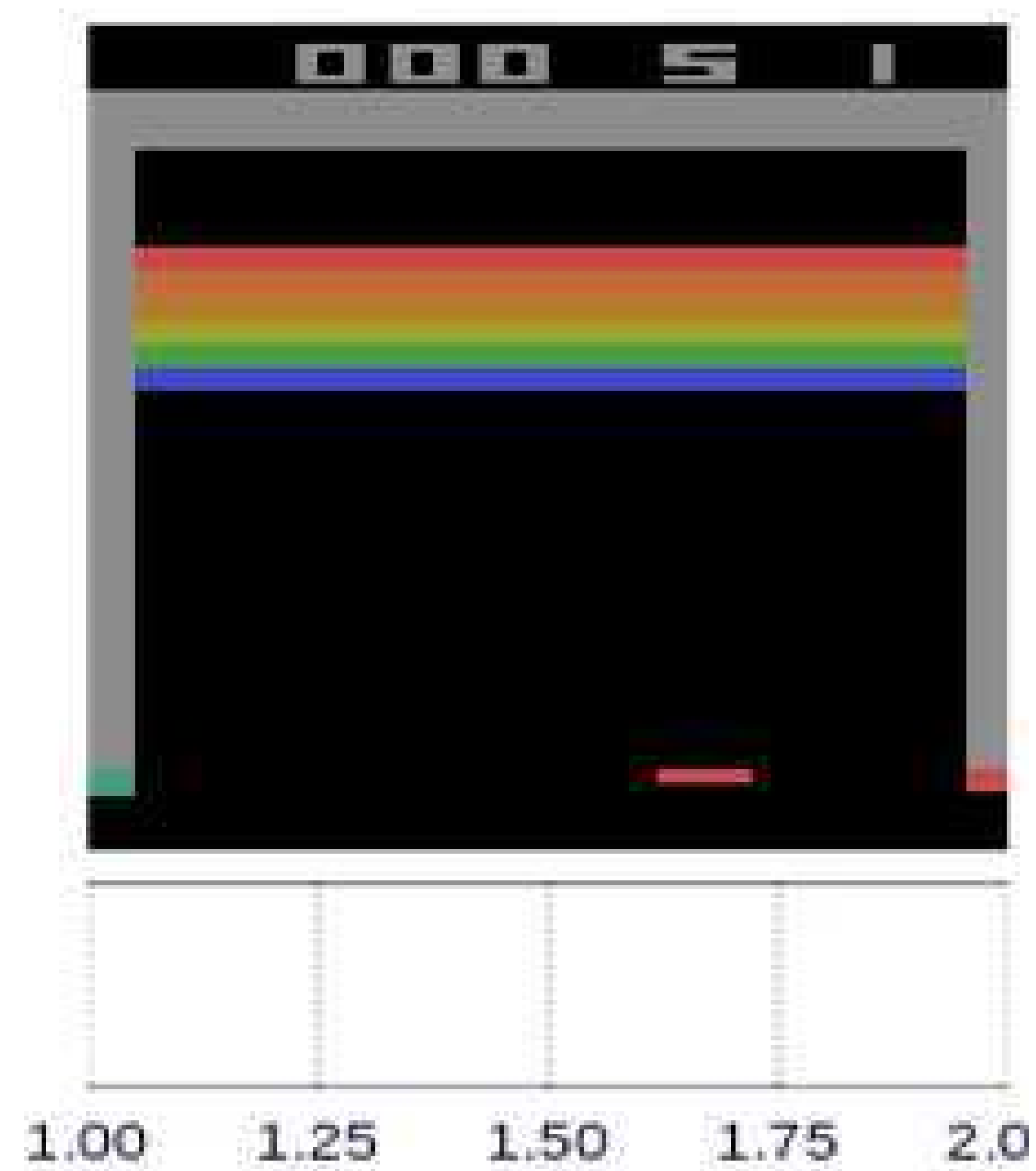
Output

Приложения ИИ

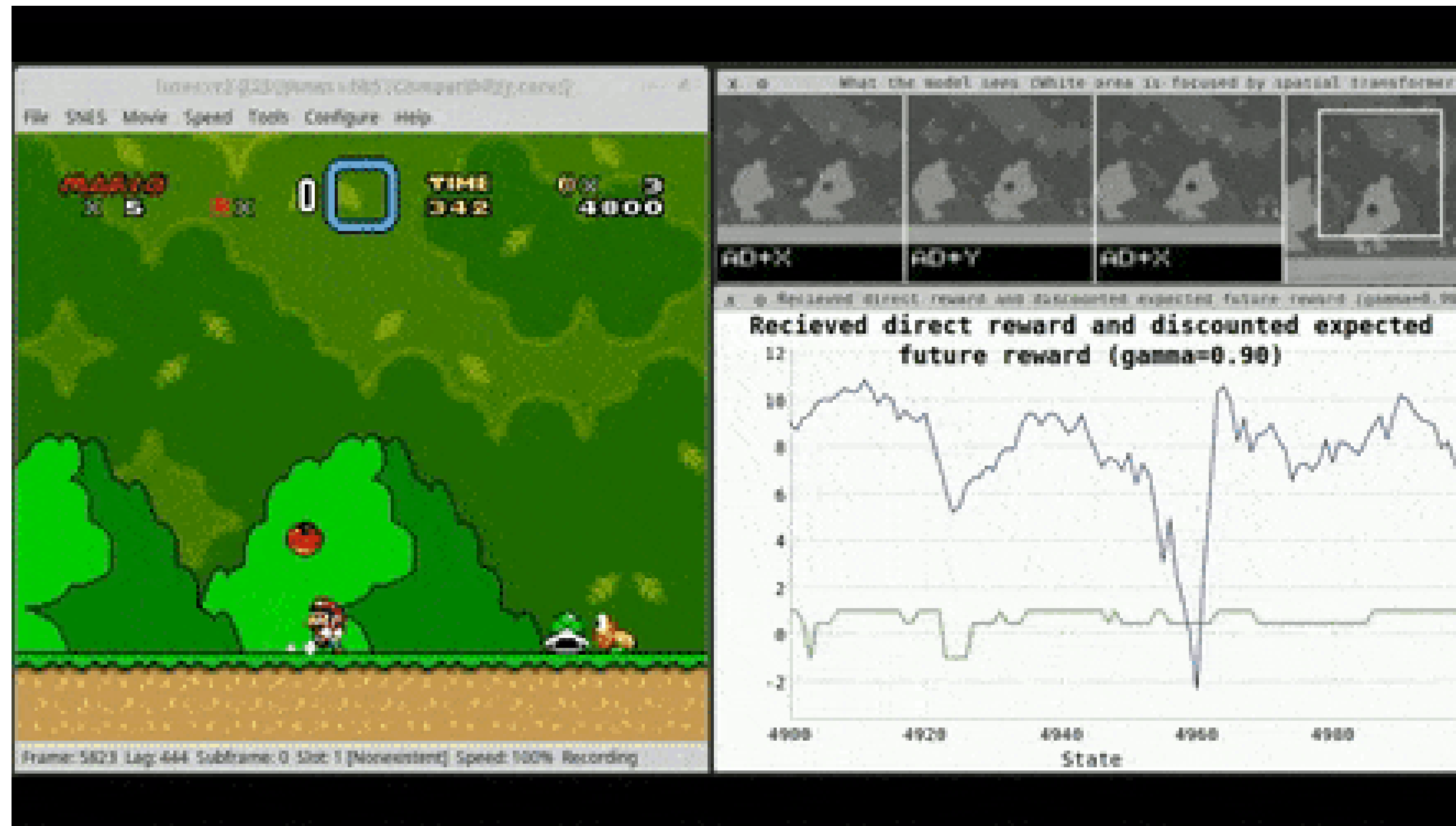
Обучение с подкреплением



Обучение с подкреплением



Обучение с подкреплением



Хронология событий



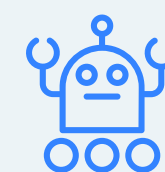
Май 2014

«Потребуется
не менее 10 лет,
прежде чем
компьютер выиграет
у человека в Го»



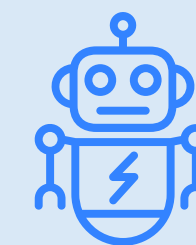
Октябрь 2015

AlphaGo
обыгрывает
чемпиона Европы
(2-го дана)



Март 2016

AlphaGo обыгрывает
18-кратного чемпиона
мира (высшего 9-го
дана)



Октябрь 2017

Новая версия AlphaGo,
которая вообще никак
не учитывает весь
человеческий опыт игры
в Го, со счетом 100:0
выигрывает у старой
версии

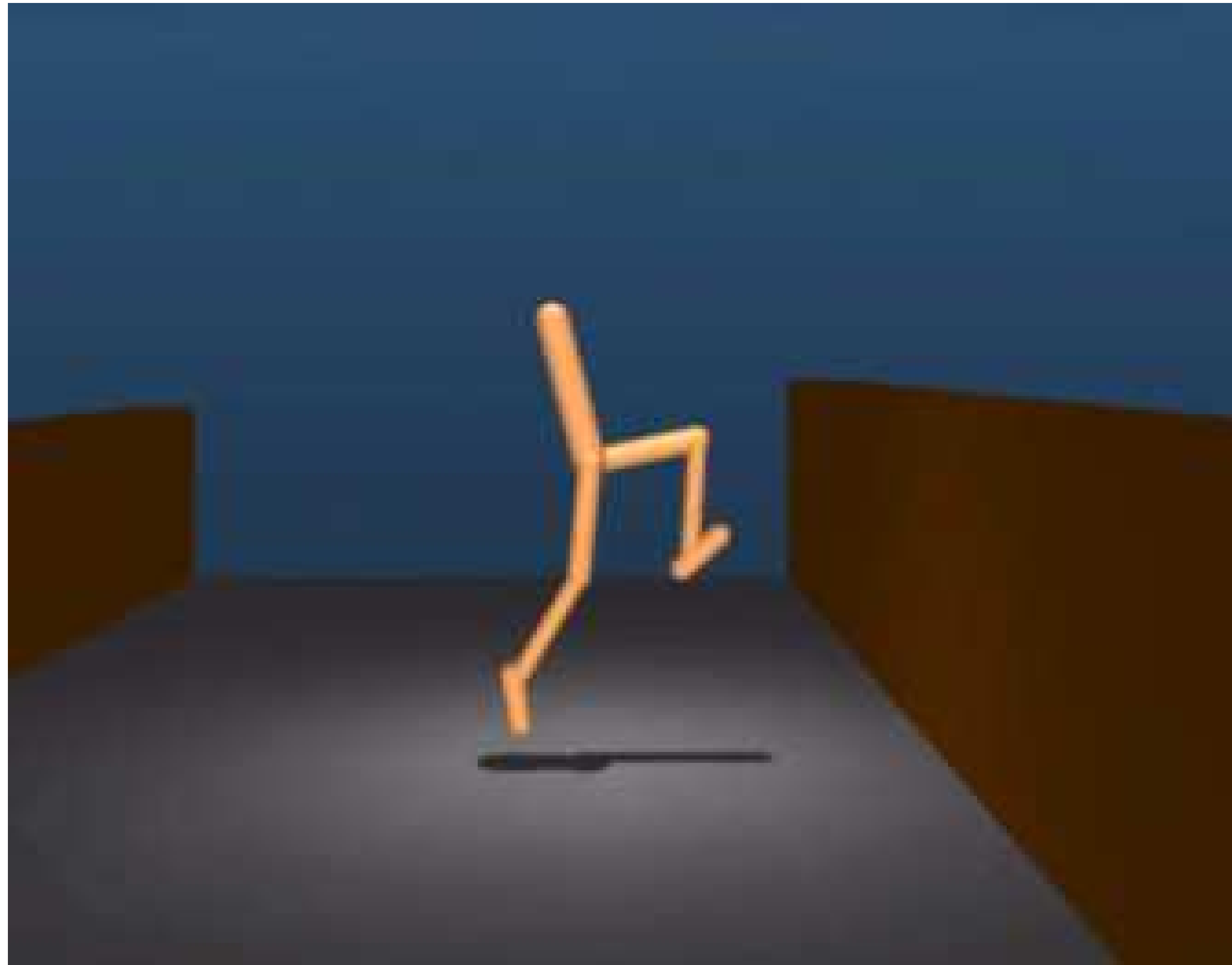


А люди уже даже не понимают ходов компьютера.
Всего лишь за 3 года профессия «игрока в Го» стремительно пролетела
с позиции «мы - элита, и нейросети нас нескоро заменят» до «вот же ж
беда, че ж нам теперь делать-то».
That was fast...

Обучение с подкреплением



Обучение с подкреплением

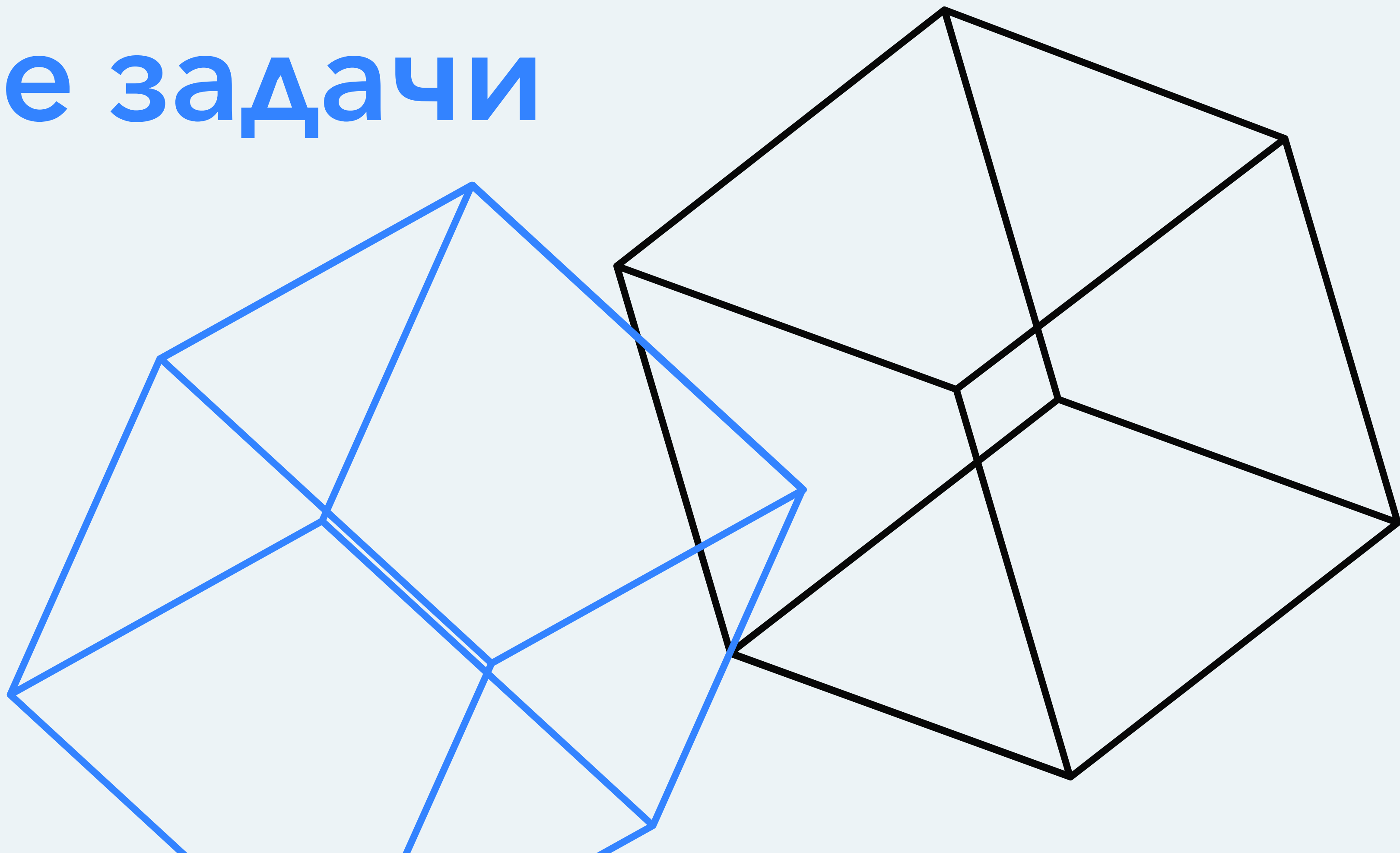


Обучение с подкреплением: примеры

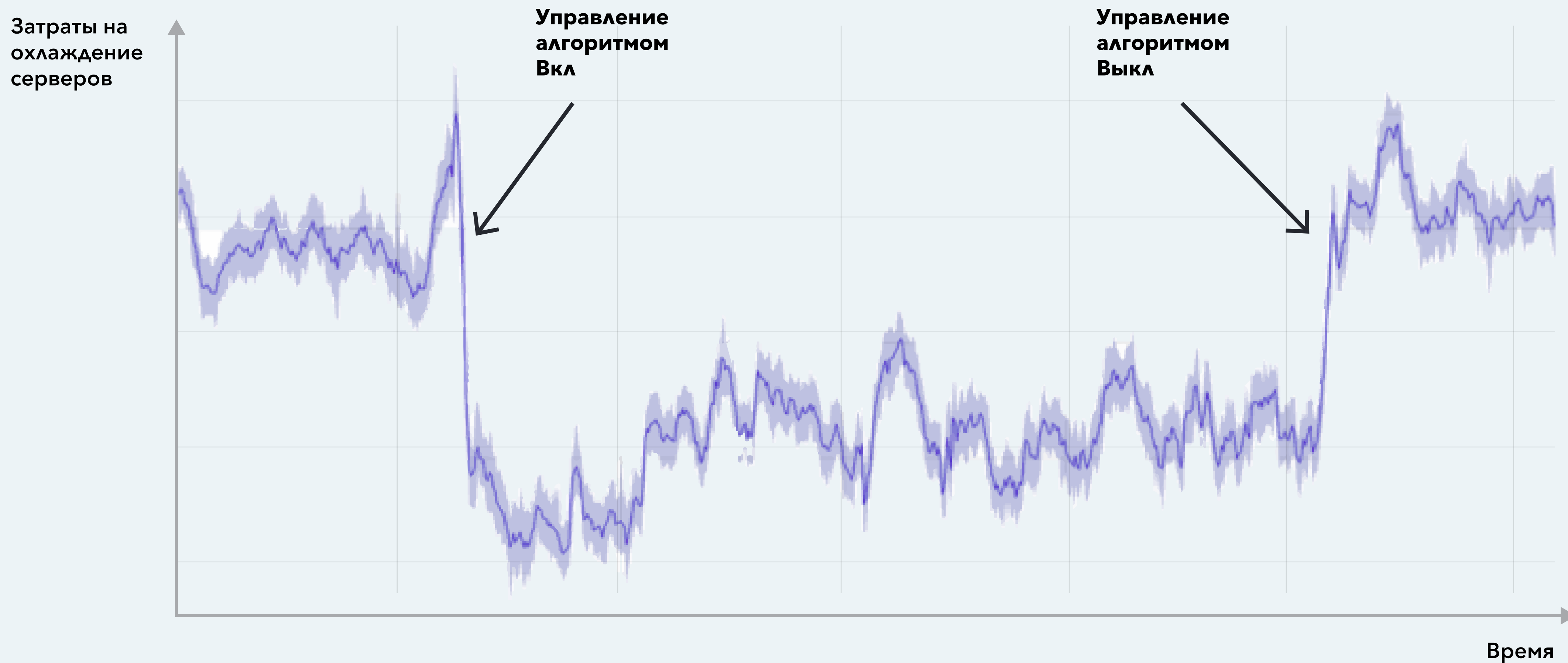


Приложения ИИ

Другие задачи



ИИ для оптимизации расходов



Google Uber

facebook.

DeepMind UBER AI Labs facebook research

TensorFlow

PYRO

PYTORCH

Лидеры

- Машинное обучение влияет на все сферы бизнеса
- Развивают технологии

- **Open source**

TensorFlow \geq **1 800** contributors

PyTorch \geq **900** contributors

Pyro \geq **50** contributors

Яндекс

SAMSUNG

amazon



SAMSUNG Research

@поиск



Yandex
CatBoost

Уверенные пользователи

- Машинное обучение в продуктах сервисах
- Используют технологии
- Стремятся к инноваторам
 - Catboost ≥ 90 contributors
 - Samsung AI Research



teradata.

Аутсайдеры

- Машинное обучение как услуга
- Закрывают технологии
- Продают инструменты

Машинное обучение и бизнес

Google Uber
facebook

Яндекс SAMSUNG
Apple @mail.ru amazon

SAP sas
teradata

DeepMind UBER AI Labs
TensorFlow PYRO
PYTORCH
facebook research

 a @поиск
Yandex CatBoost
SAMSUNG Research

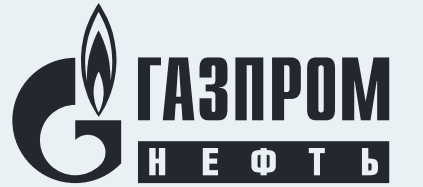
Единственно
верный
способ

(открытость

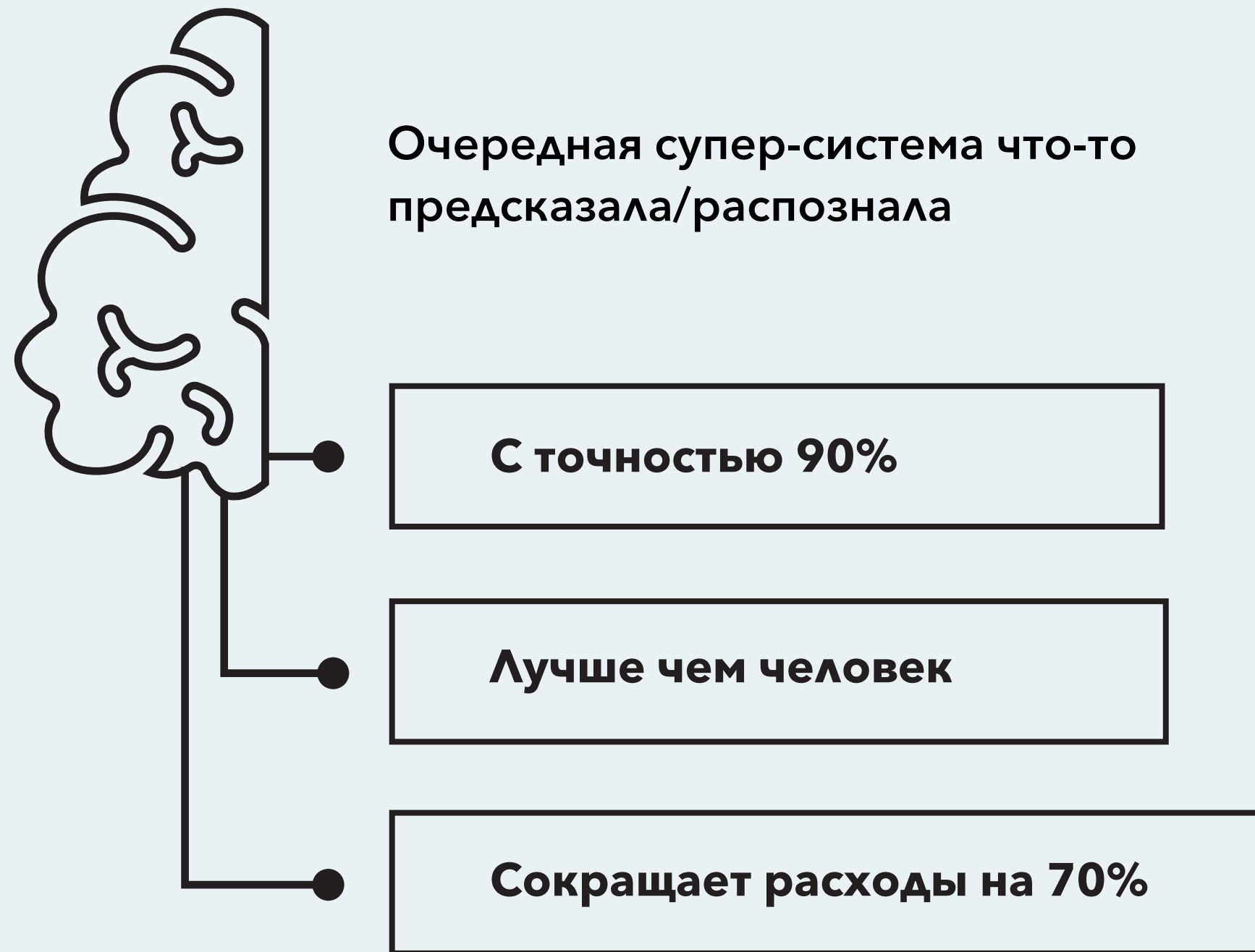
>> рассказывать
всем



На что обращать внимание в любых смелых заявлениях?



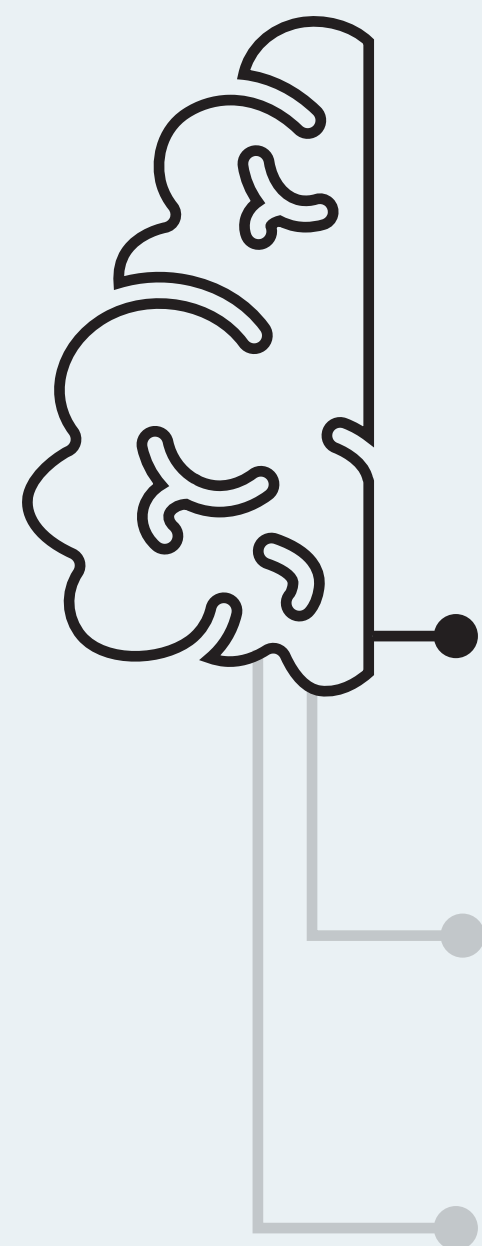
Заявление



На что обращать внимание в любых смелых заявлениях?



Заявление



Очередная супер-система что-то предсказала/распознала

С точностью 90%

Лучше чем человек

Сокращает расходы на 70%

На самом деле

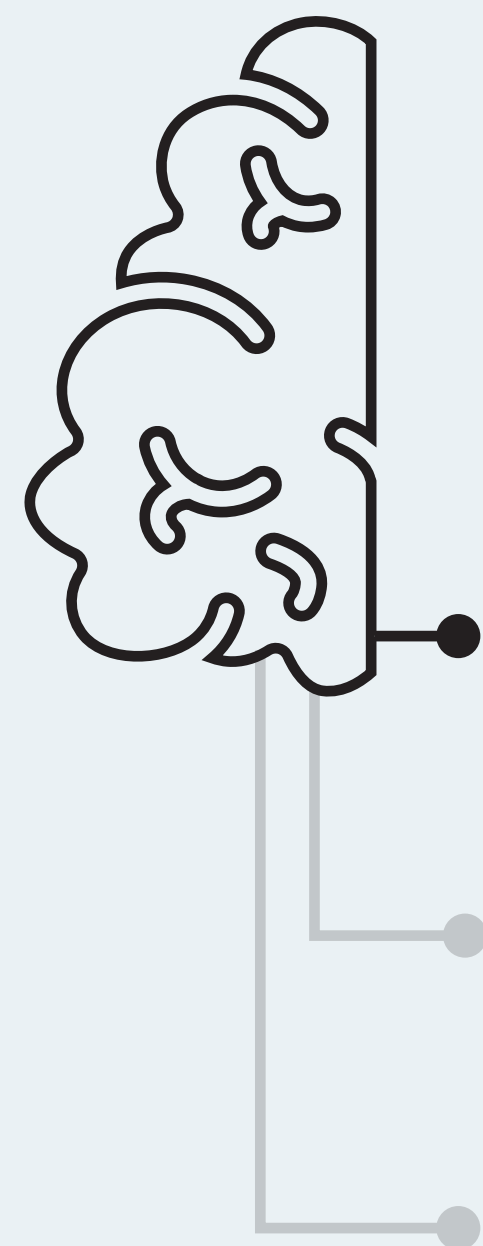
Подаем для проверки системы данные по 10 насосам, где мы точно знаем, что 1 находится в аварийном состоянии



На что обращать внимание в любых смелых заявлениях?



Заявление



Очередная супер-система что-то предсказала/распознала

С точностью 90%

Лучше чем человек

Сокращает расходы на 70%

На самом деле

Подаем для проверки системы данные по 10ти насосам, где мы точно знаем что 1 находится в аварийном состоянии



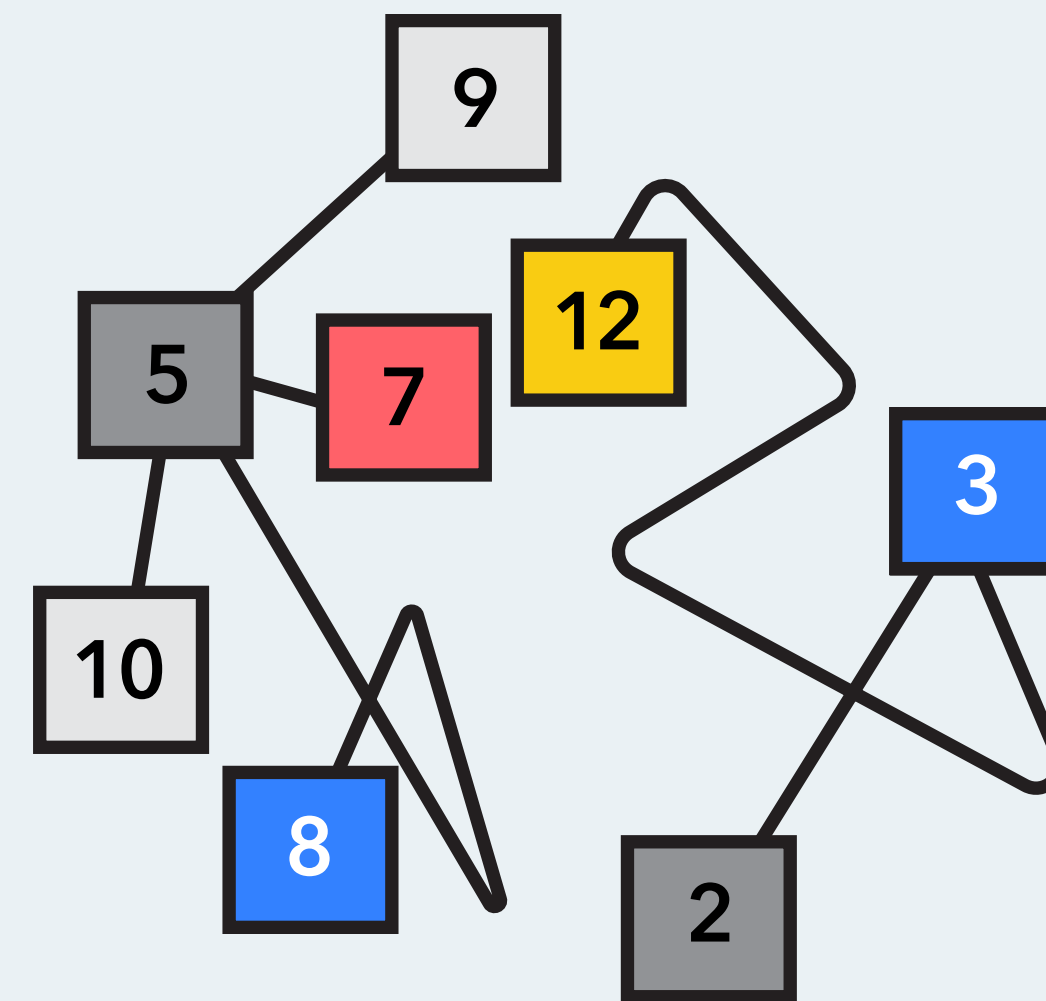
Система говорит, что все 10 насосов работают отлично

И ведь ошиблись-то мы только в 1/10, а значит точность 90%

Для оценки моделей DS



**Нужны правильные метрики,
которые демонстрируют качество
алгоритмов**



Задачи и метрики, о которых мы
поговорим в следующих лекциях

Class imbalance



Sensitive

Non-sensitive

Confusion Matrix

Accuracy [ACC]

$$\frac{TP+TN}{TP+FN+FP+TN}$$

Balanced Accuracy [BA]
 $0,5 \cdot (TPR+TNR)$

Weighted Accuracy [WA]
 $w \cdot TPR + (1-w) \cdot TNR$

F-Measure

$$\frac{2PR \cdot TPR}{PR+TPR}$$

Precision [PR] = $\frac{TP}{TP+FP}$

G-Mean

$$\sqrt{TPR \cdot TNR}$$

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP (17)	FN (03)
	Negative	FP (04)	TN (12)

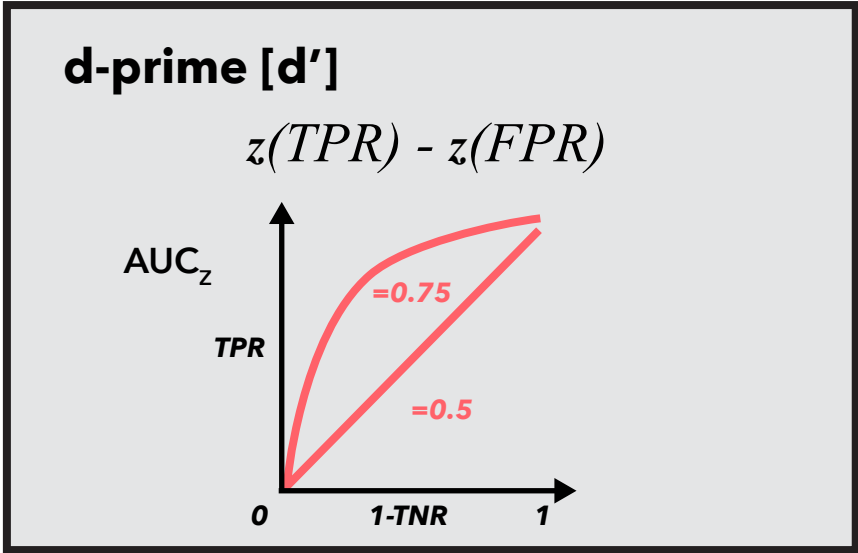
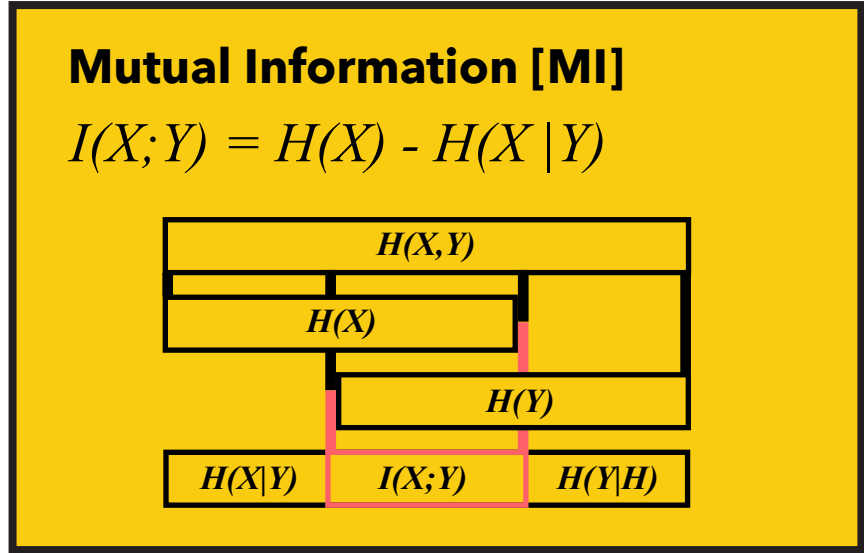
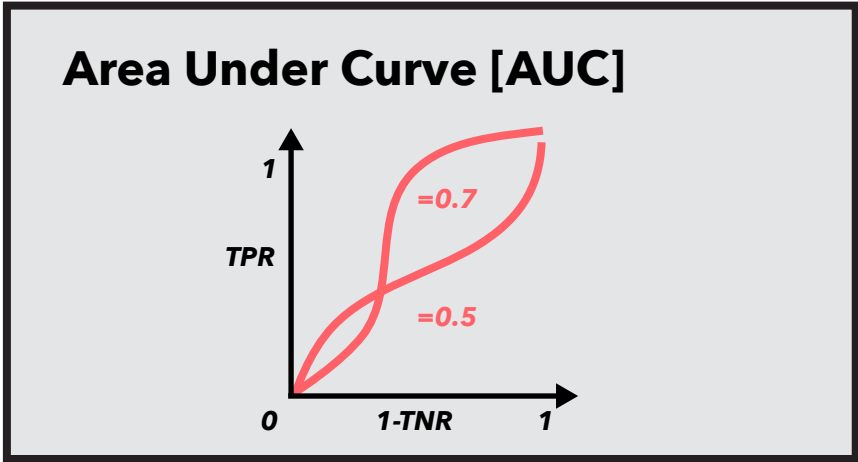
True Positive Rate [TPR] = $\frac{TP}{TP+FN}$

True Negative Rate [TNR] = $\frac{TN}{TN+FP}$

Matthews Correlation Coefficient [MCC]

$$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{P \cdot N \cdot (TP+FP)(TN+FN)}}$$

$P = TP+FN$ $N = TN+FP$



$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$

The sum of squared residuals (SSE)

$$SSE = \sum (y_i - \hat{y}_i)^2$$

$$R^2_a = \frac{(n-1) R^2 - k}{[n - (k+1)]}$$

$$\hat{\sigma}^2 = \frac{SSE}{n - (k+1)} = MSE$$

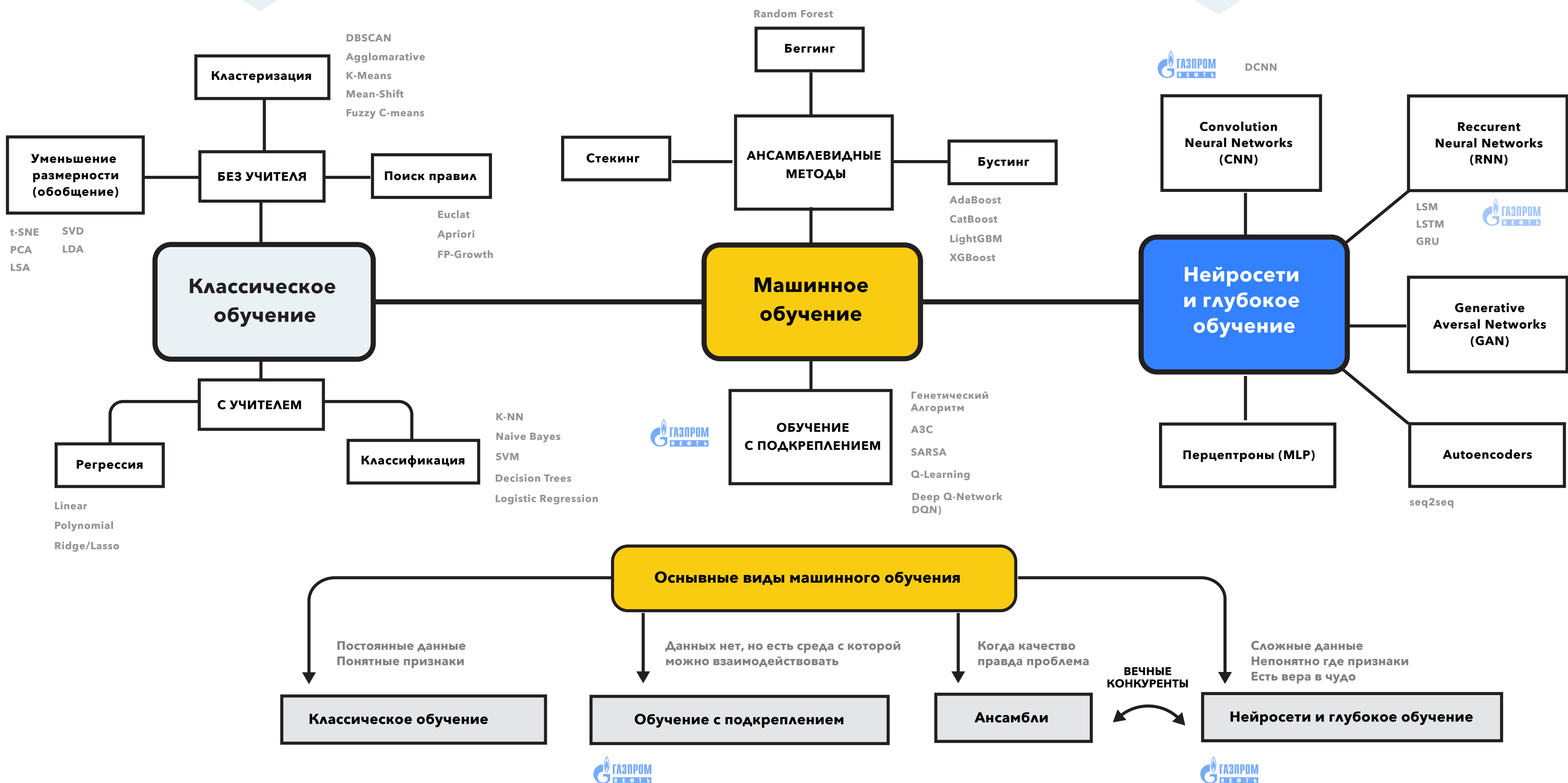
The sum of squared total residuals (SST)

$$SSE = \sum (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SSE}{SST}$$

$$f = \frac{R^2 / k}{(1 - R^2) / [n - (k+1)]}$$

$$f \geq F_{\alpha, k, n - (k+1)}$$



Заявление

Очередная супер-система что-то предсказала/распознала

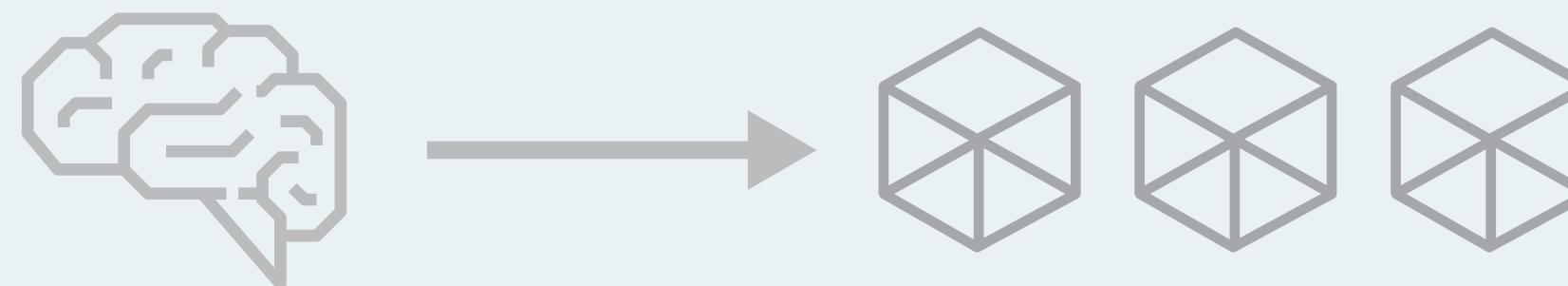
С точностью 90%

Лучше чем человек

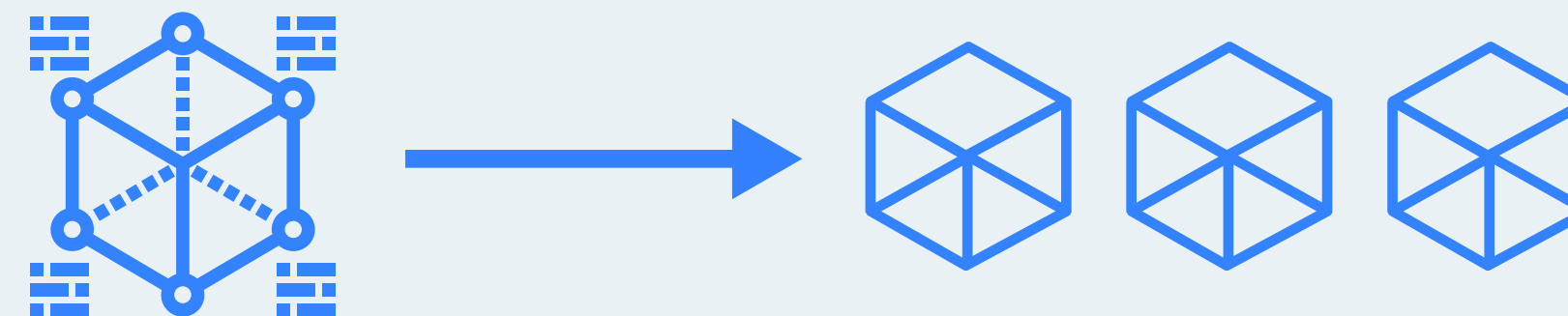
Сокращает расходы на 70%

На самом деле

Есть работа человека по разметке нефтесодержащих пластов



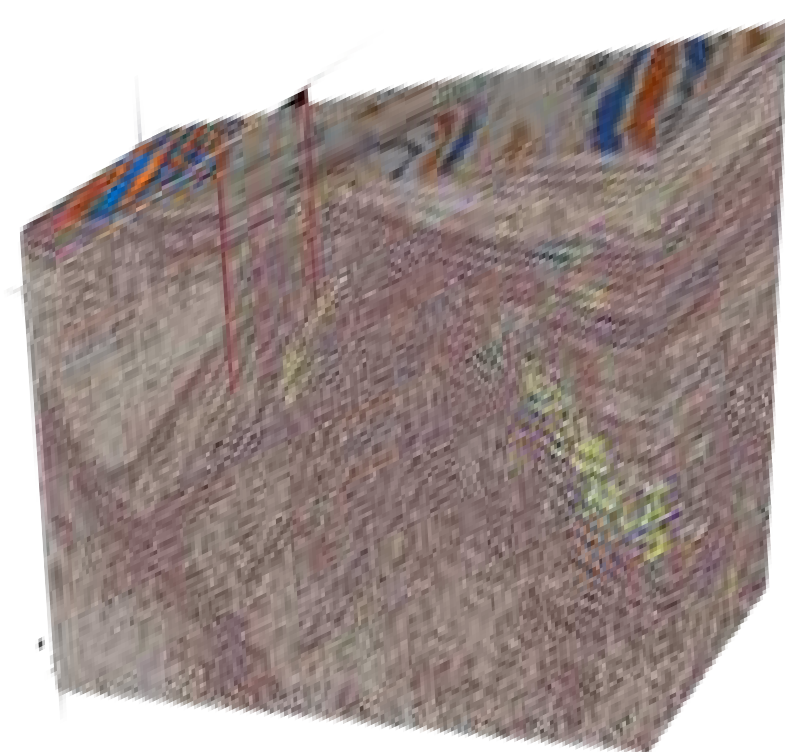
Есть алгоритм по разметке нефтесодержащих пластов



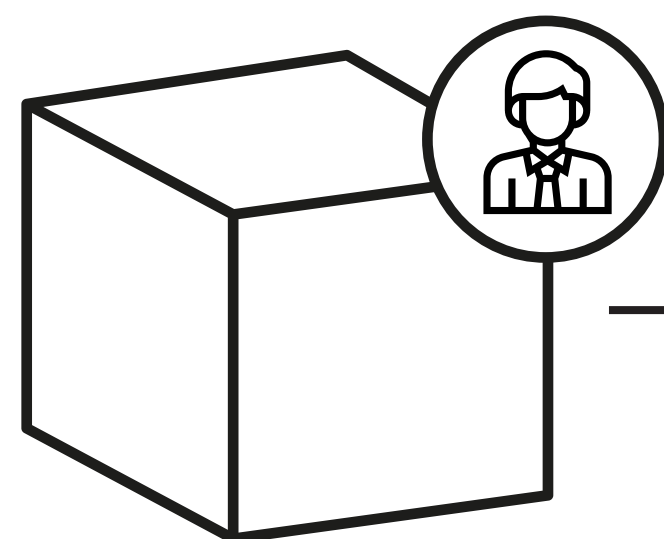
Для того, чтобы проверить, а действительно ли алгоритм «лучше человека», должен существовать исследованный эталон, про который точно (путем бурения) на каждом участке известно все

Ближайшая задача на 2019

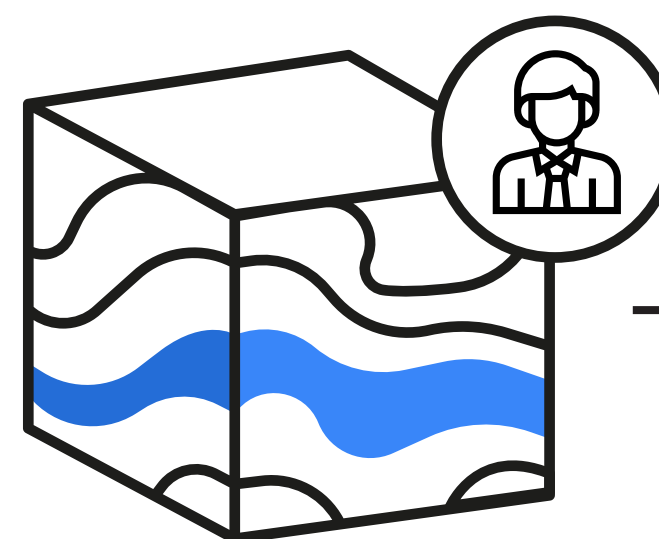
В текущем workflow специалистов «незаметно/неощутимо» интегрировать модели, которые дадут сокращение времени на процессах



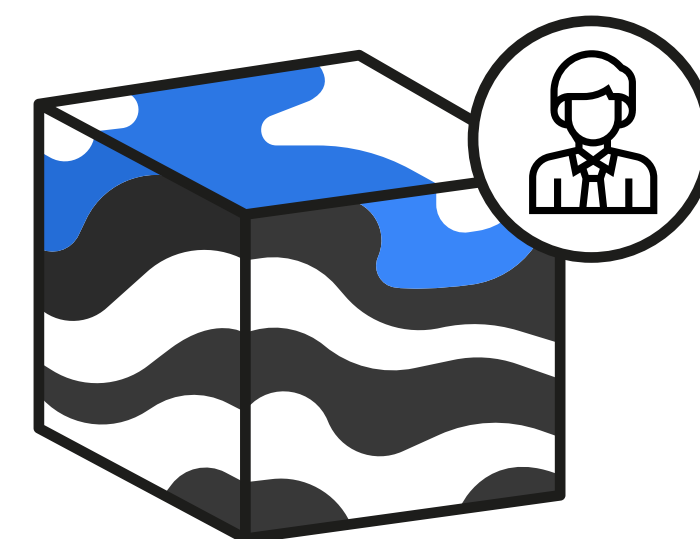
Ничего нет, разметки нет –
полная неопределенность



Добавили разметки.
Сделали модели очистки
шума – стало чуть понятнее



Добавили данных и разметки.
Сделали модели разметки
фаций – стало уже интересно

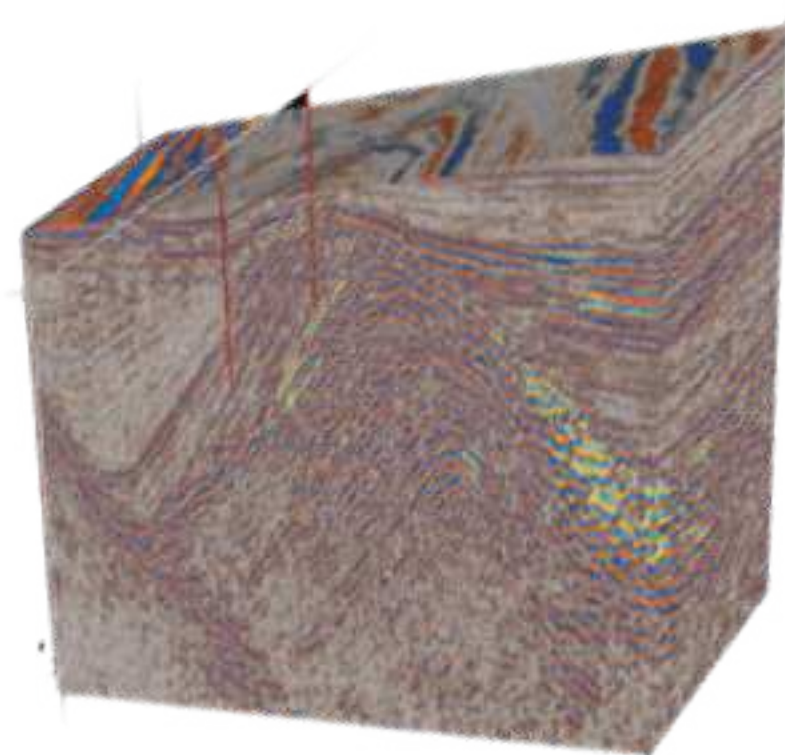


Добавили данные каротажей
и смогли восстановить еще
недостающие данные и выявить
пропущенные пропластки

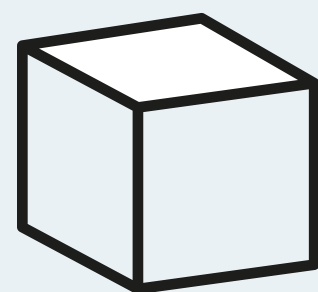
Ближайшая задача на 2019

В текущем workflow специалистов «незаметно/неощутимо» интегрировать модели, которые дадут сокращение времени на процессах

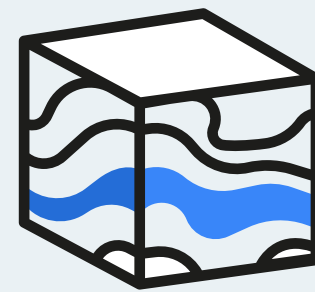
Цифровая платформа (например: Когнитивный геолог)



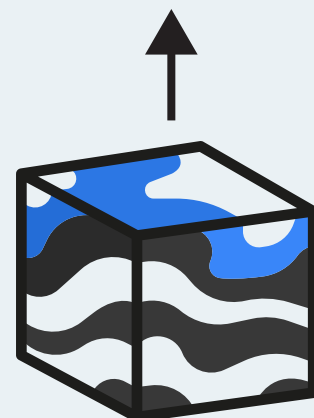
Ничего нет, разметки нет –
полная неопределенность



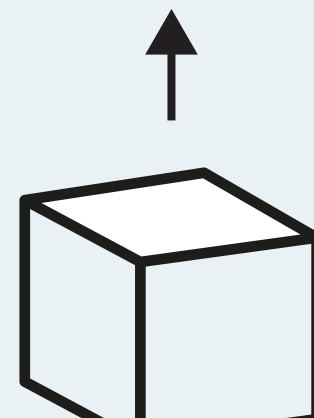
Модель очистки
сейсмики от шума



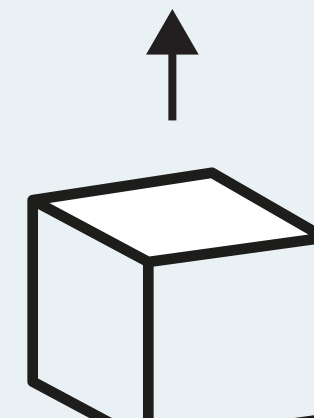
Модель разметки
фаций



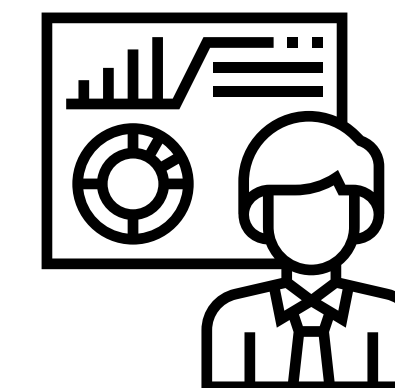
Модель
выделения
пропластков



Новые модели
дальше по workflow



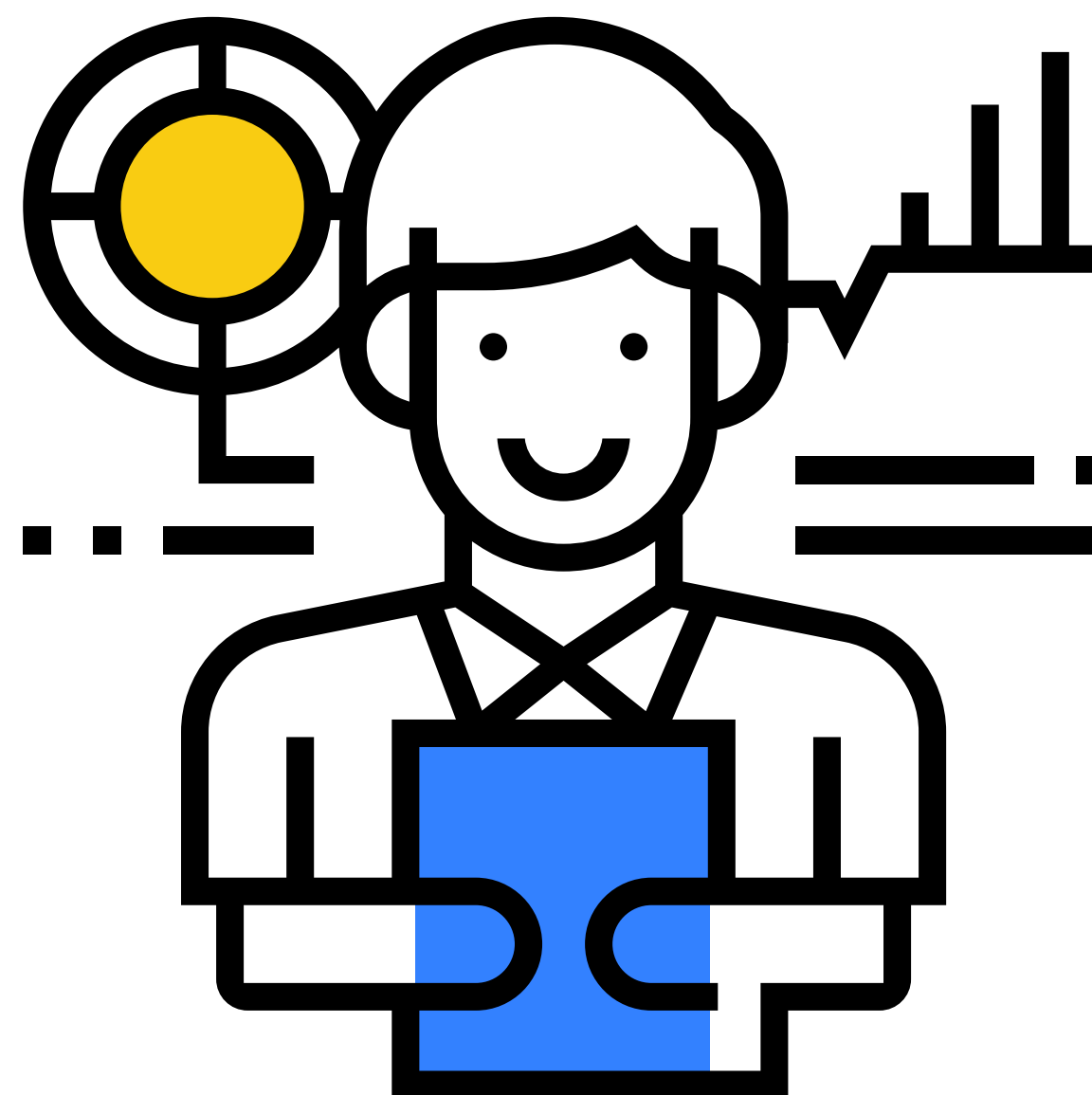
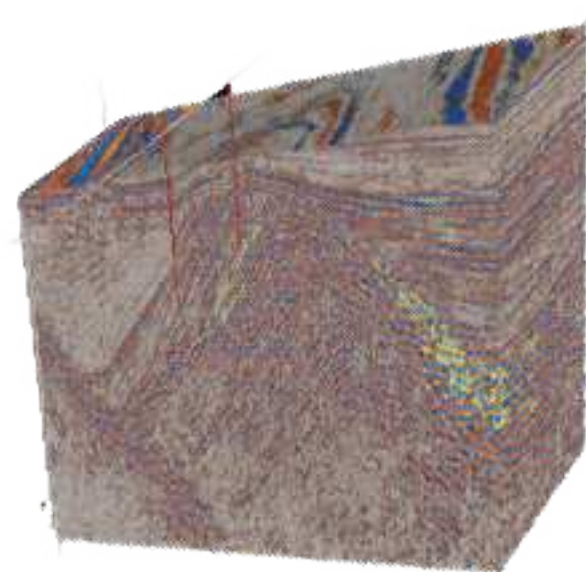
Новые модели
дальше по workflow



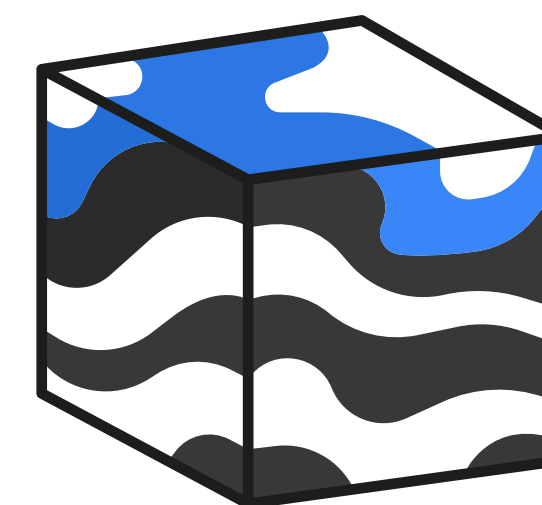
На выходе вся
информация
для проверки и
анализа эксперту

Зачем нужен датасет?

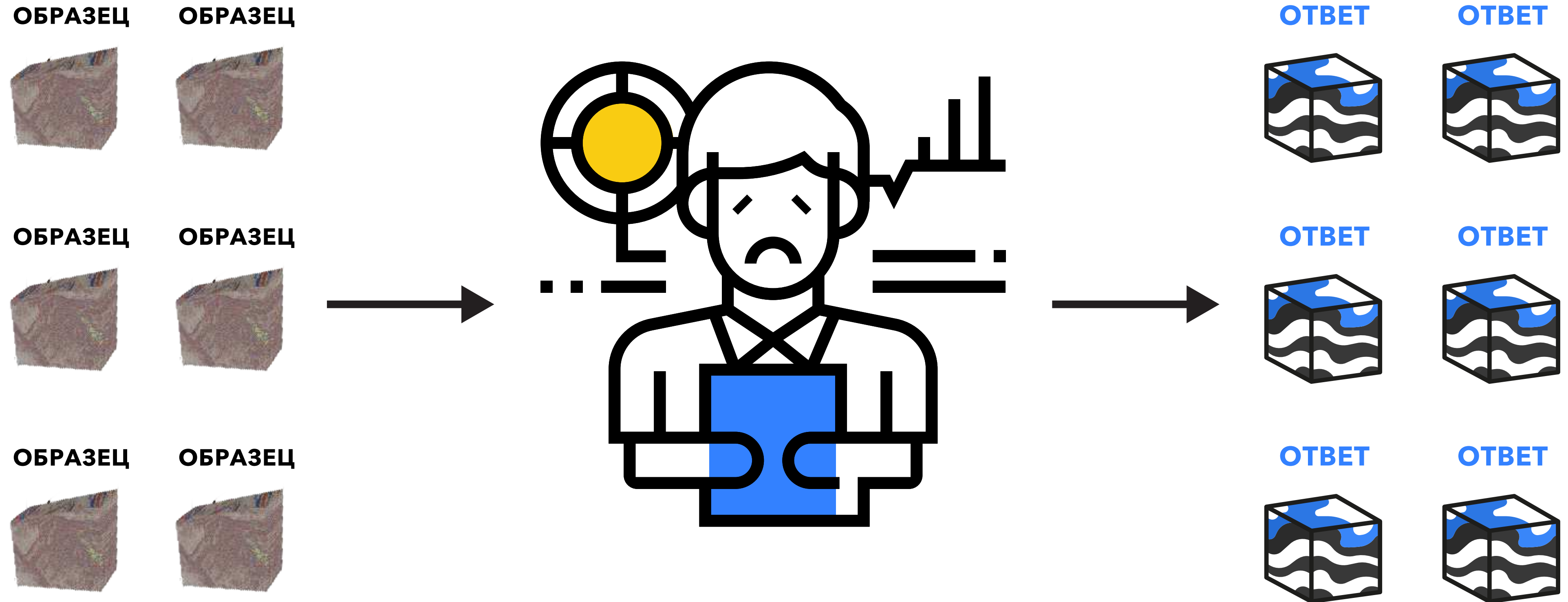
ОБРАЗЕЦ



ОТВЕТ



Зачем нужен датасет?



Зачем нужен датасет?

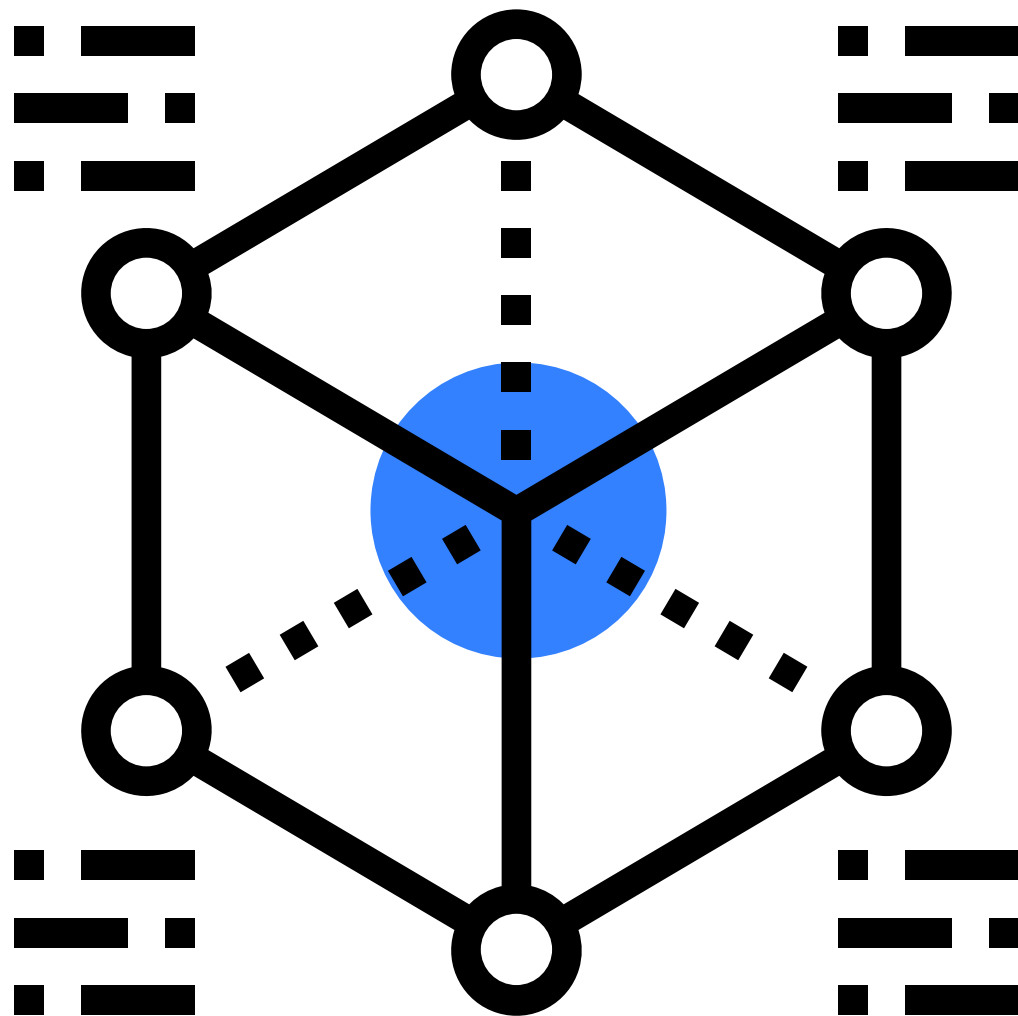
ОБРАЗЕЦ ОБРАЗЕЦ



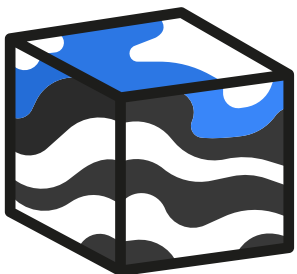
ОБРАЗЕЦ ОБРАЗЕЦ



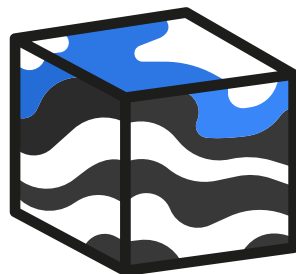
ОБРАЗЕЦ ОБРАЗЕЦ



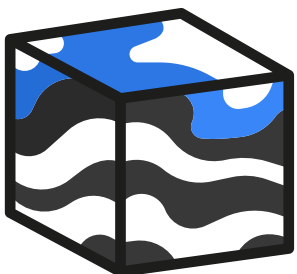
ОТВЕТ



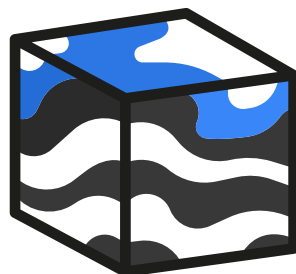
ОТВЕТ



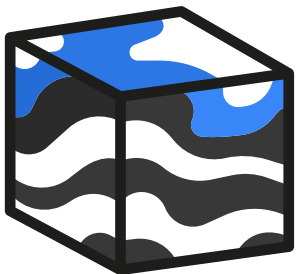
ОТВЕТ



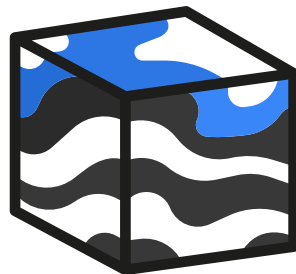
ОТВЕТ



ОТВЕТ



ОТВЕТ

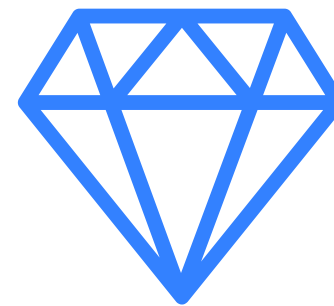
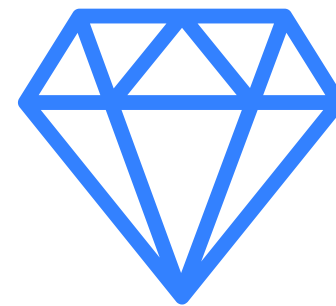
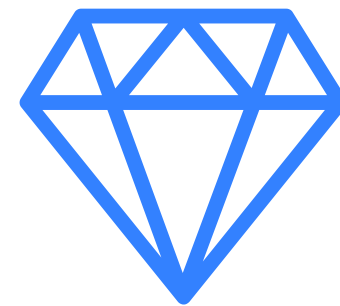
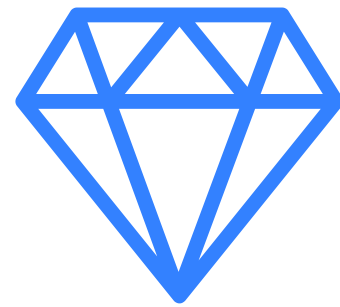
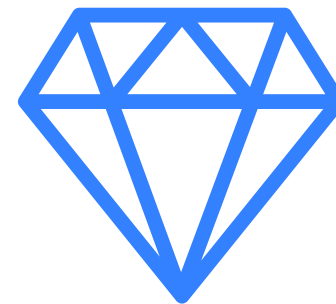


Зачем нужен датасет?

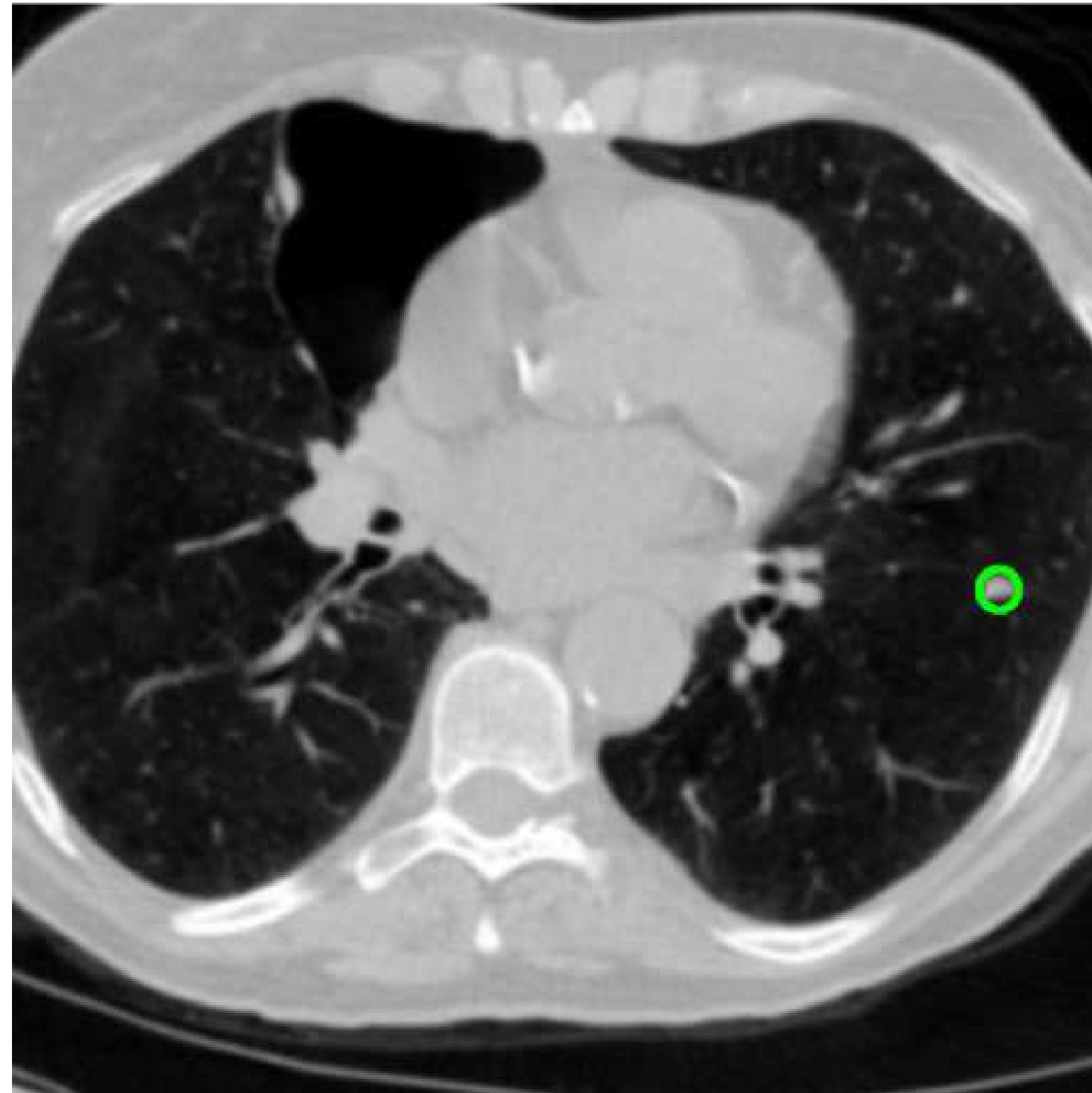
DATA

MODEL

RESULT



Что значит «плохой датасет»?



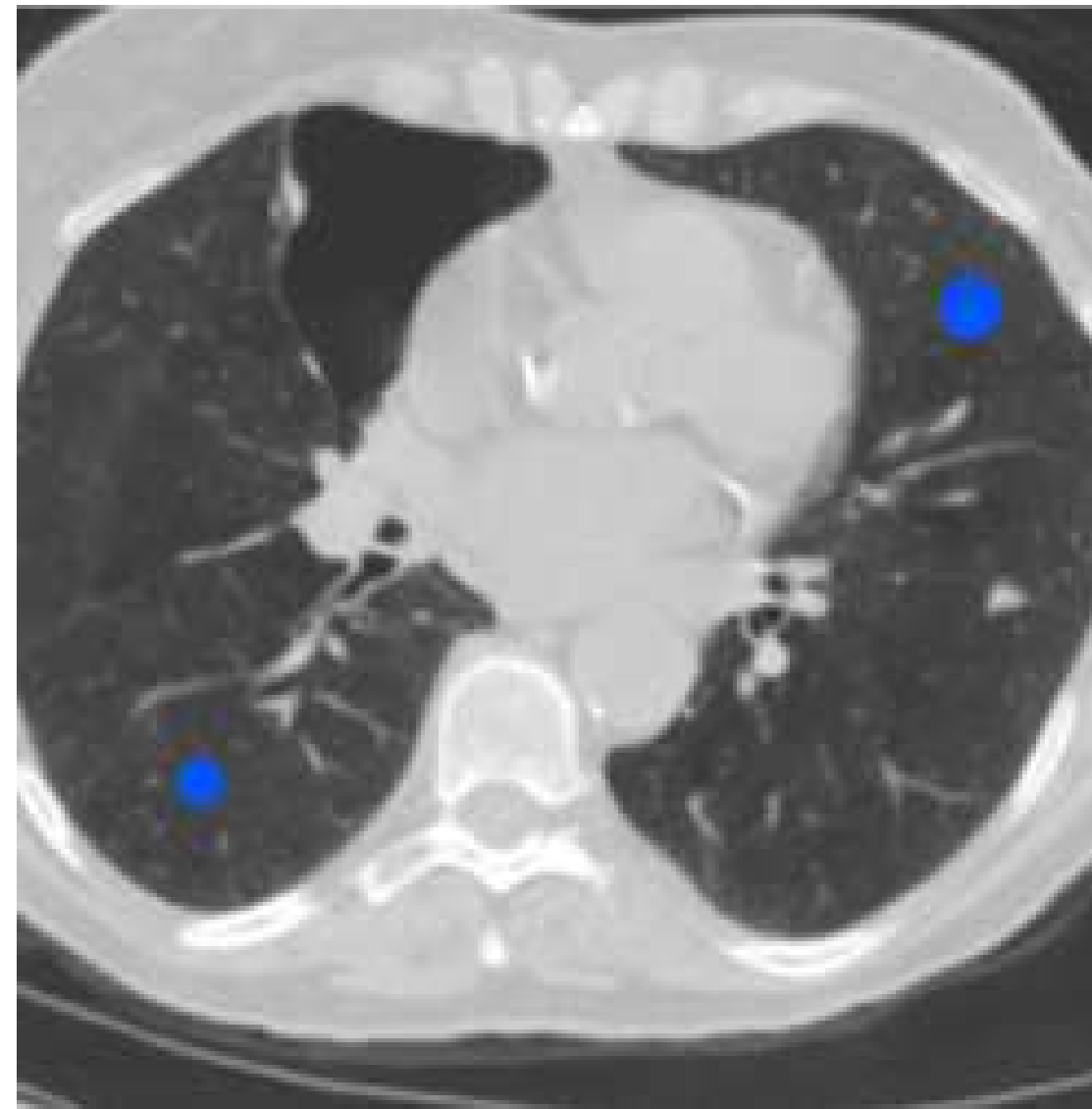
Что значит «плохой датасет»?

Эксперт 1



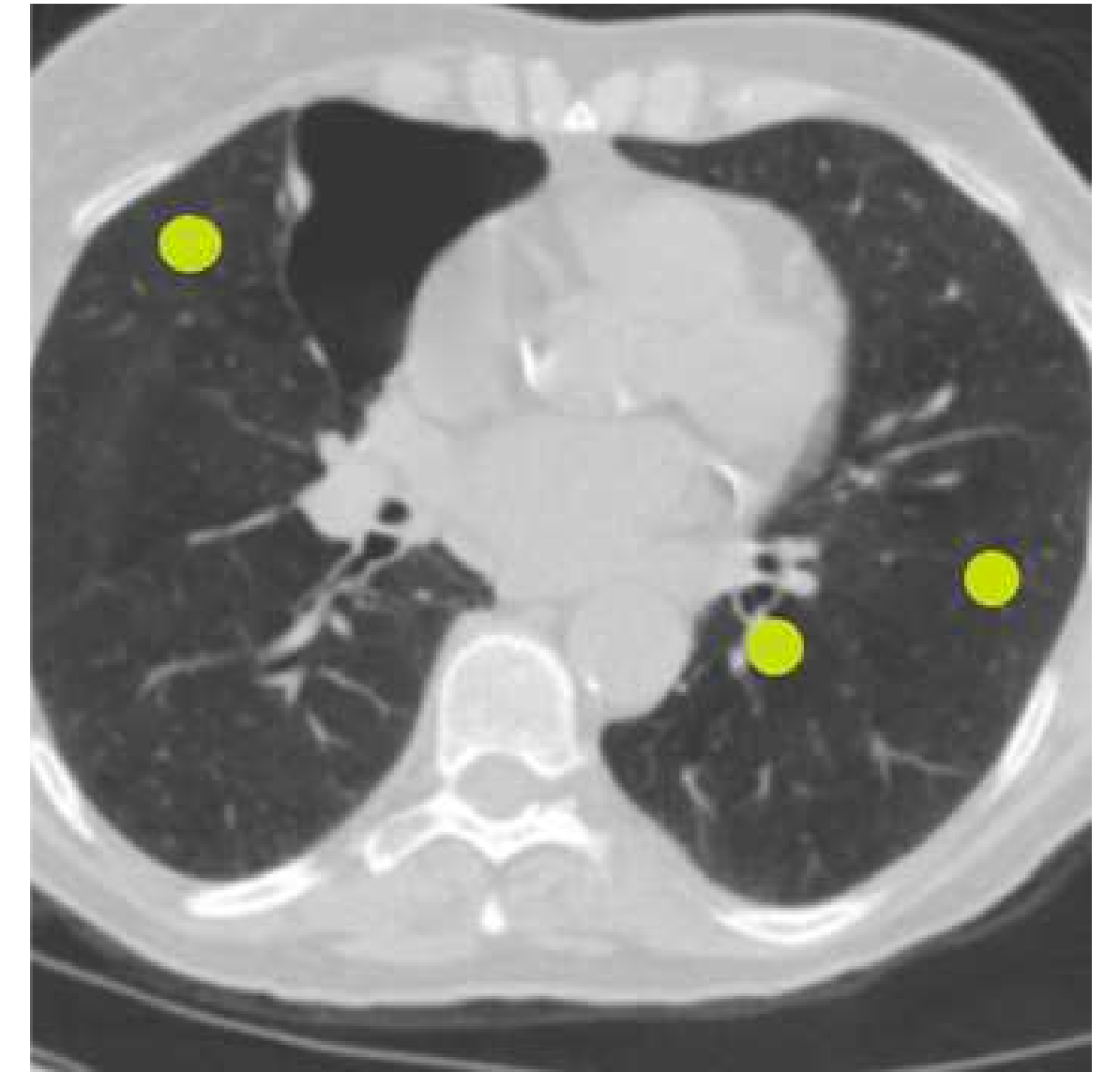
+

Эксперт 2

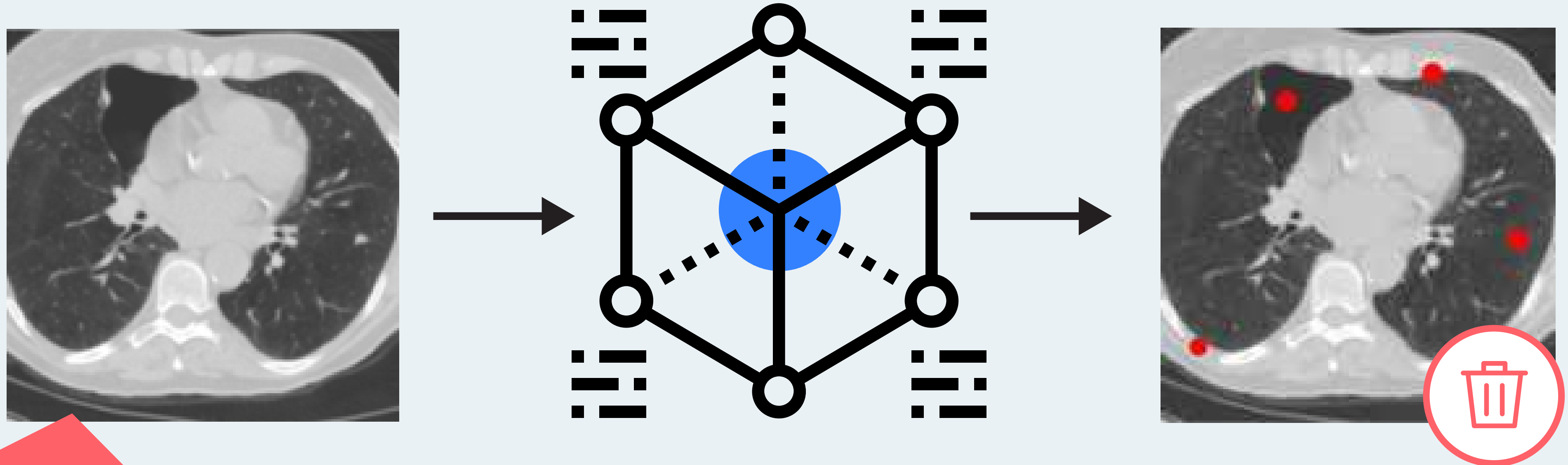


+

Эксперт 3



Что значит «плохой датасет»?



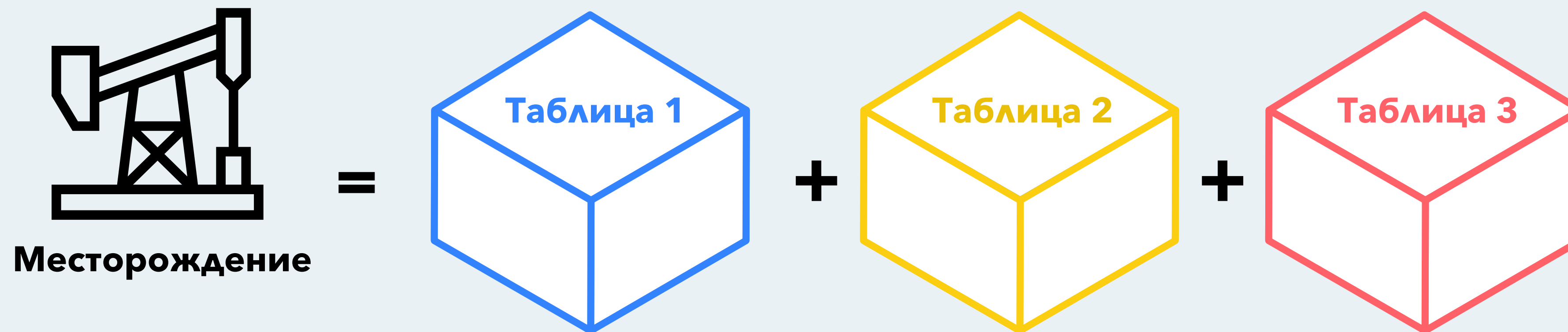
Проблема

Разметка разных экспертов сильно отличается

Решение

Отбор экспертов и валидация итоговой разметки

Что значит «плохой датасет»?



Проблема

В разных таблицах одна скважина может быть под разными названиями

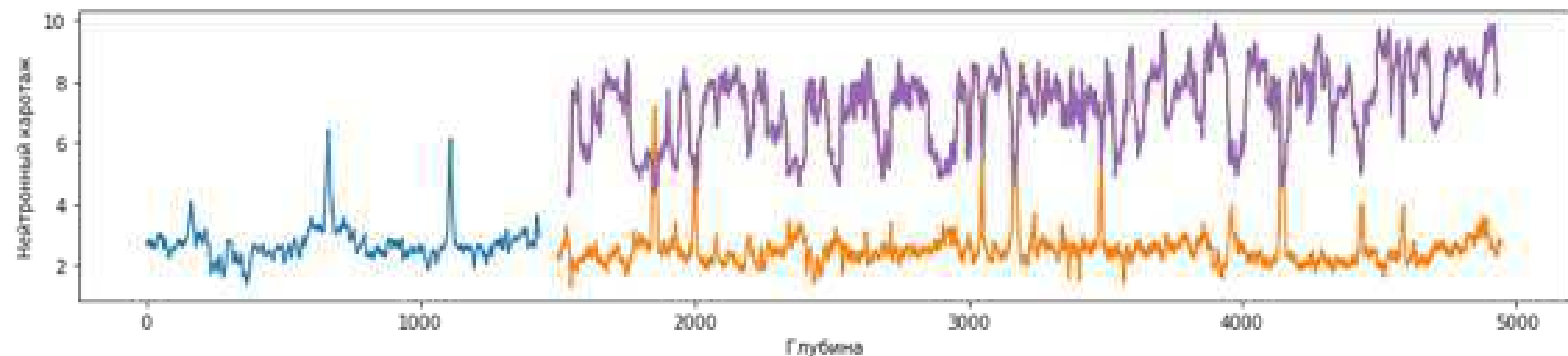
Решение

Стандарты занесения данных в базы данных

Что значит «плохой датасет»?

Проблема

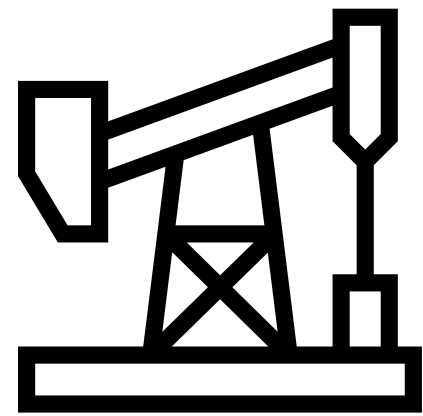
1. Малое количество общих каротажей по скважинам месторождения
2. Отсутствие алгоритма объединения нескольких каротажей одного типа по одной скважине



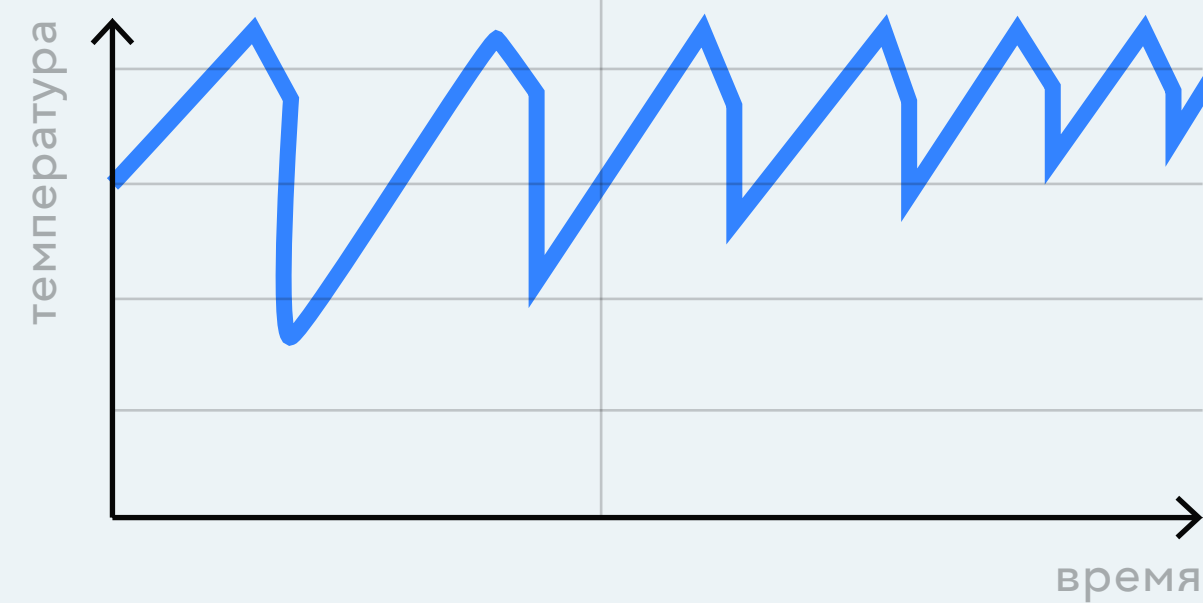
Решение

Создание эталонного датасета с помощью петрофизика

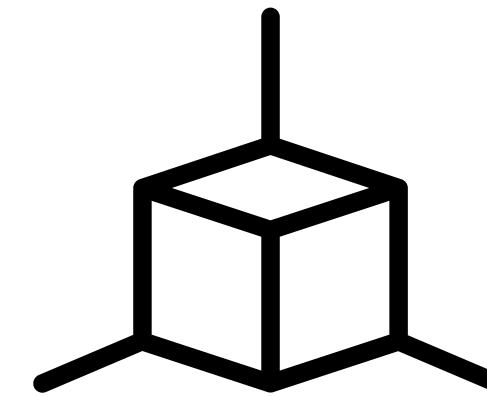
Моделирование: простой путь



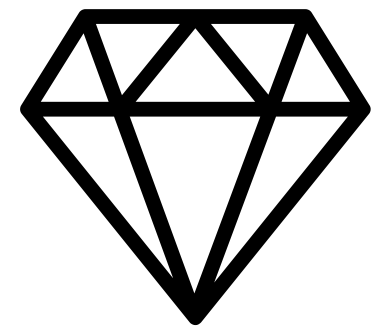
Насос на скважине



Показания датчиков

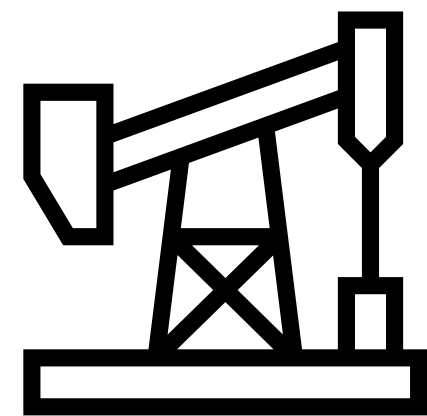


Простая модель

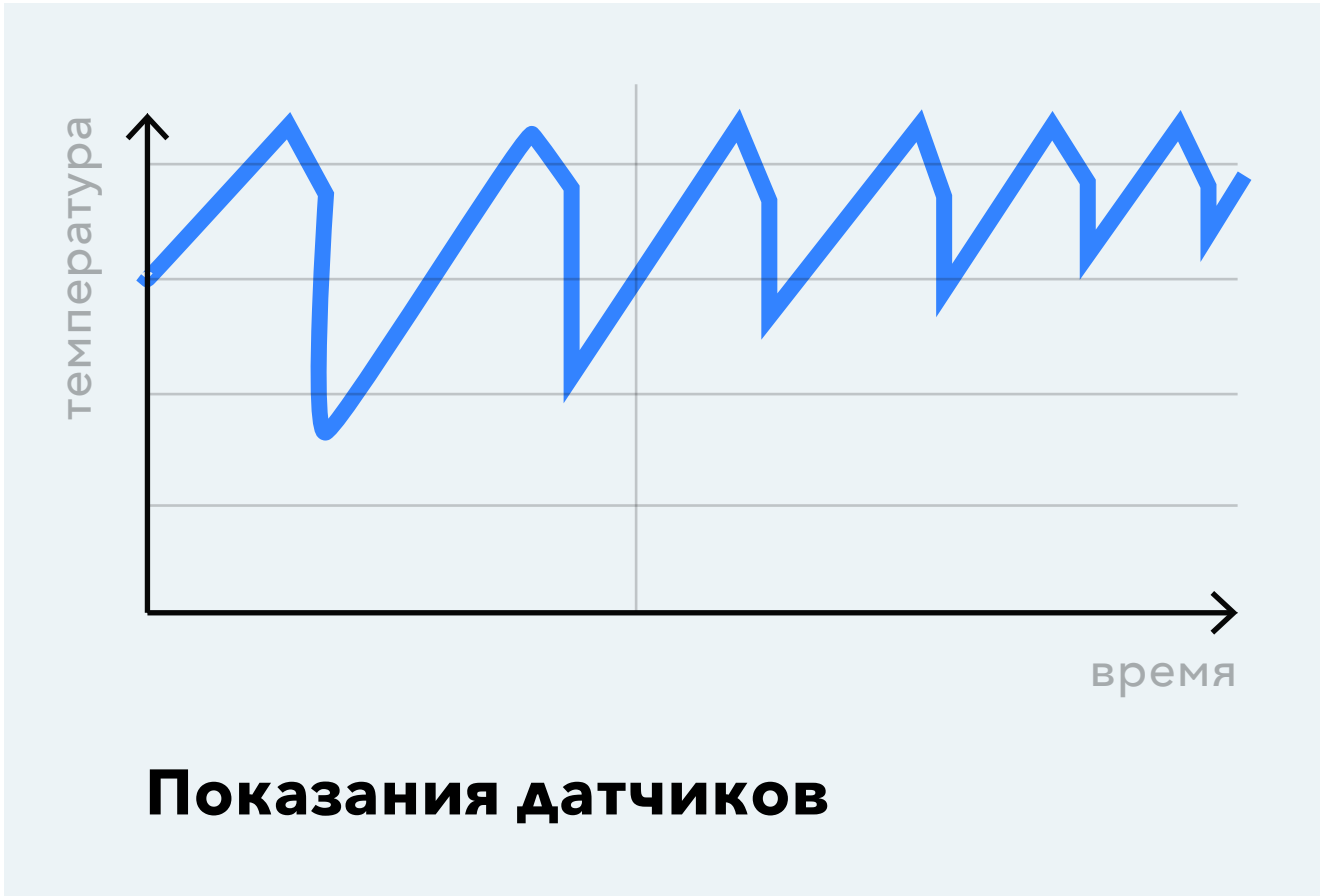


Прогноз поломки

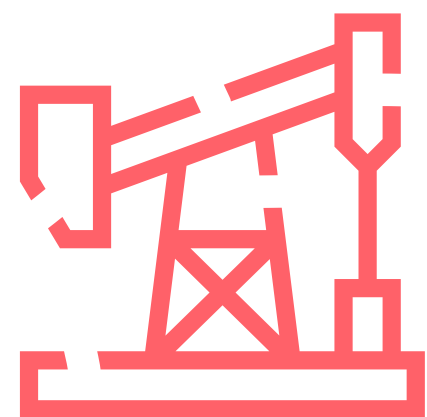
Проблемы простого пути



Насос на скважине



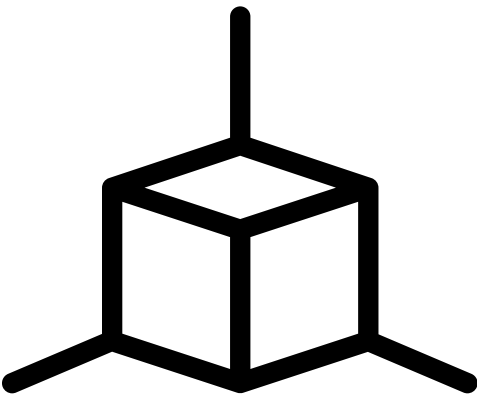
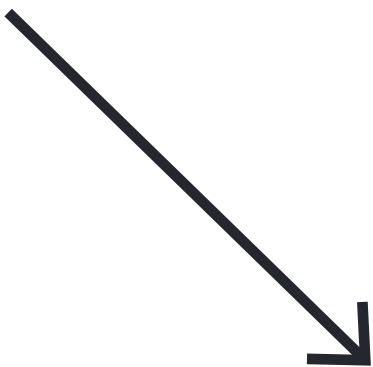
Показания датчиков



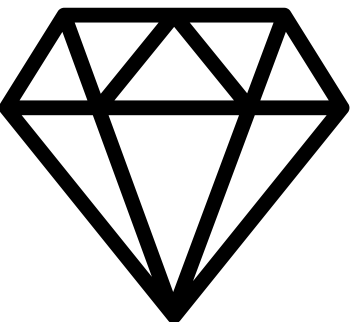
Изношенный насос



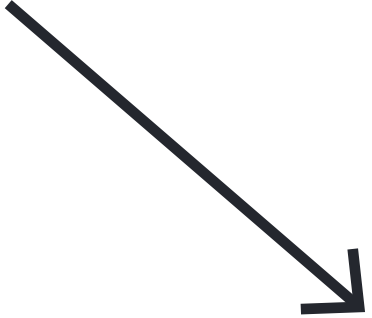
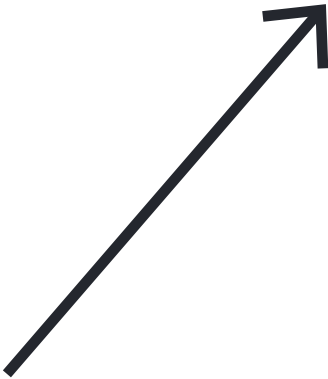
Показания датчиков
Нетипичные для модели данные



Простая модель

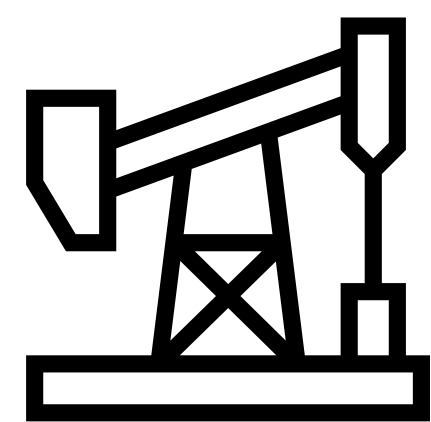


Прогноз поломки

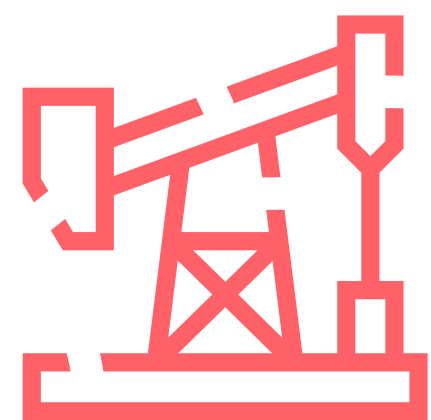
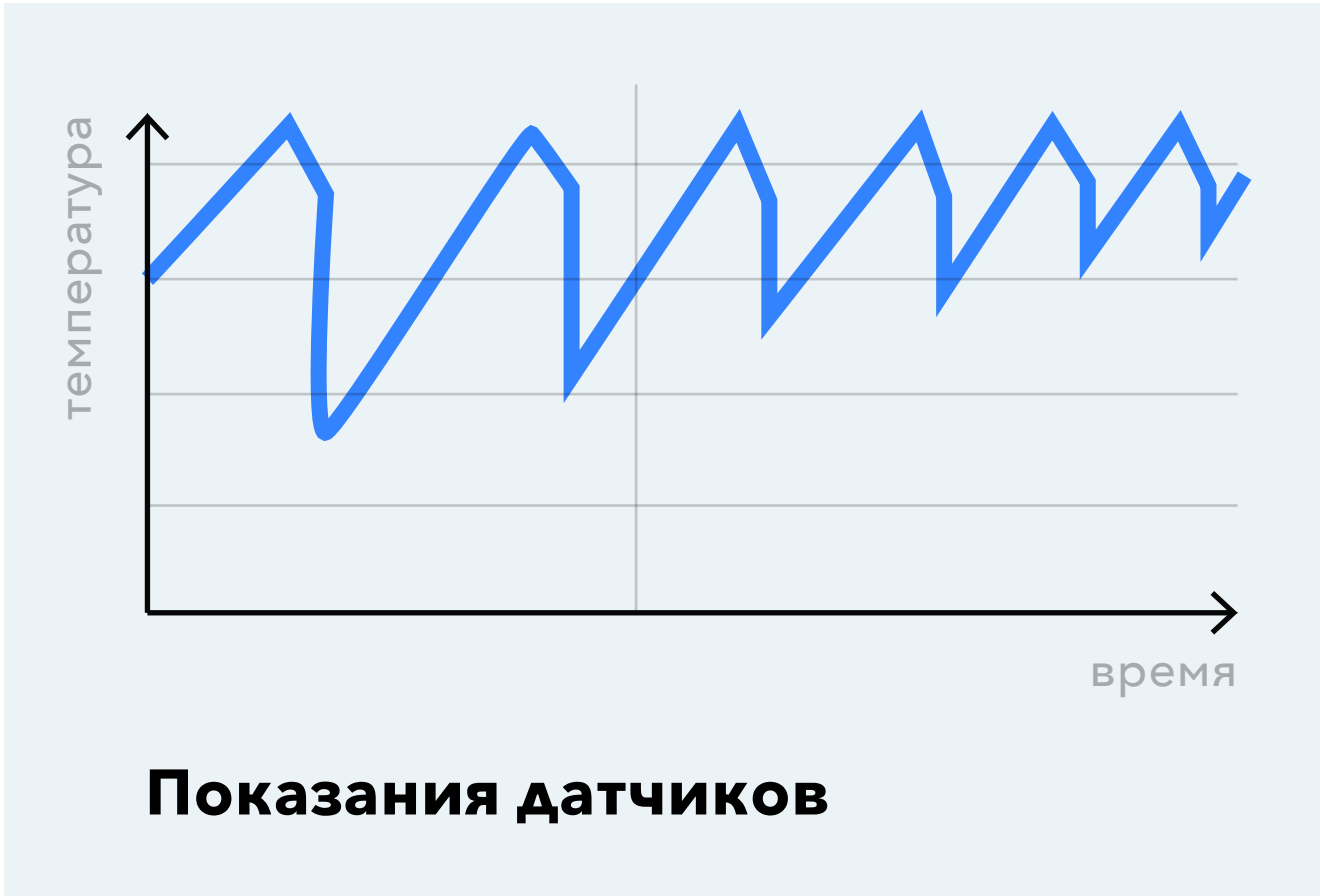


Прогноз поломки

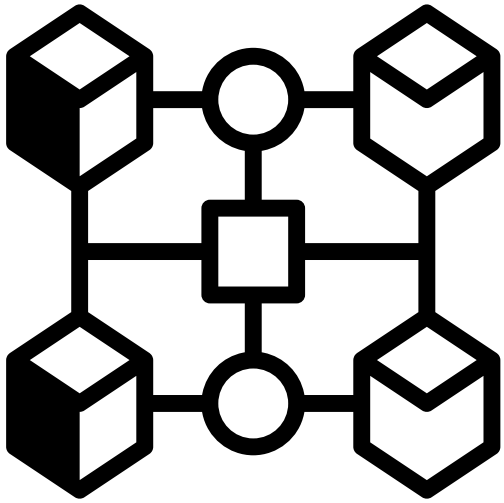
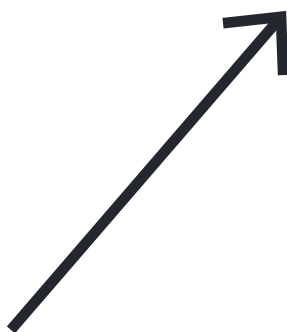
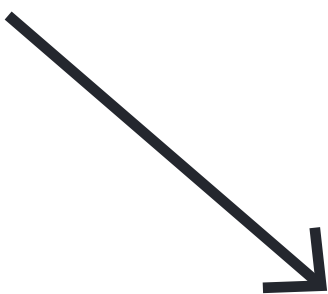
Моделирование: правильный путь



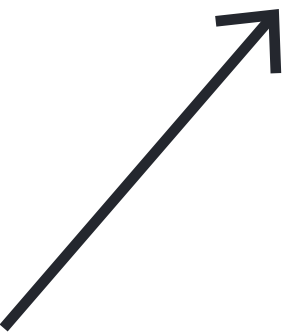
Насос на скважине



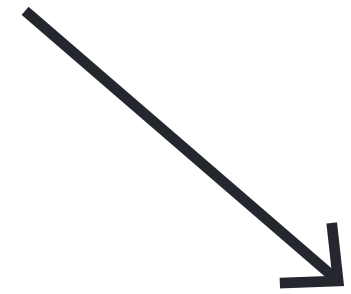
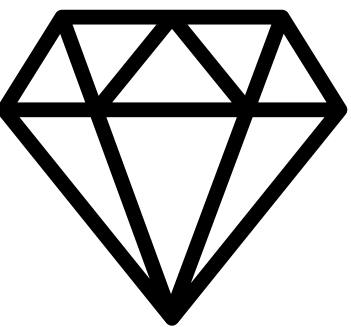
Изношенный насос



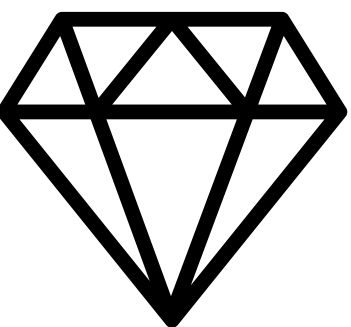
Сложная модель



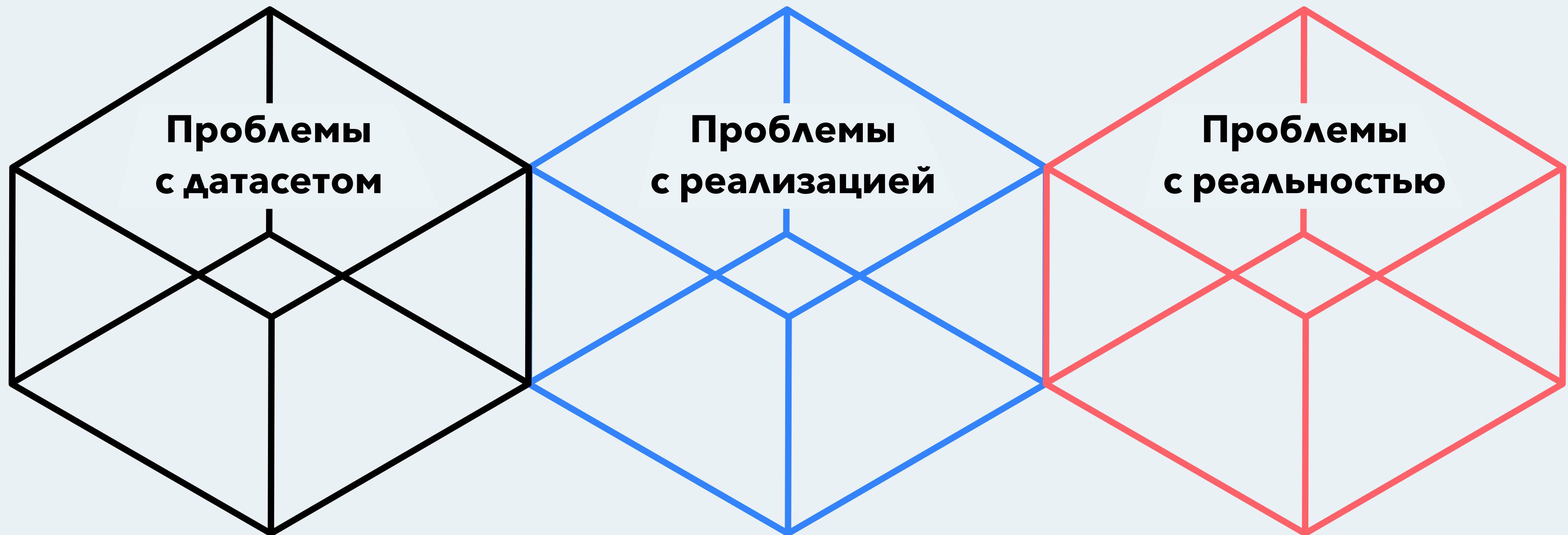
Прогноз поломки



Прогноз поломки



Причины провалов в машинном обучении



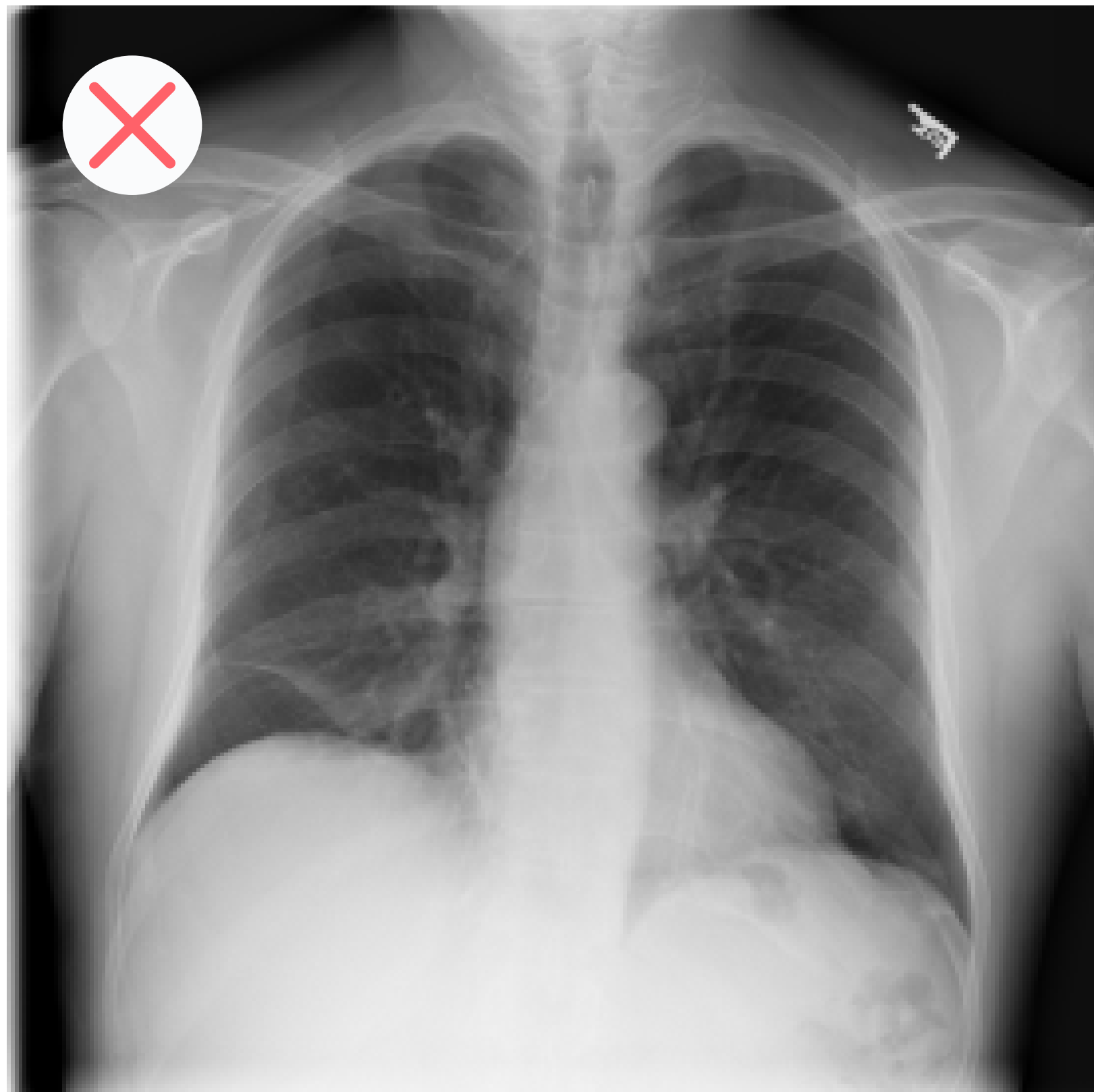
Китайская система распознавания нарушений



Common objects in context



ChestXray14



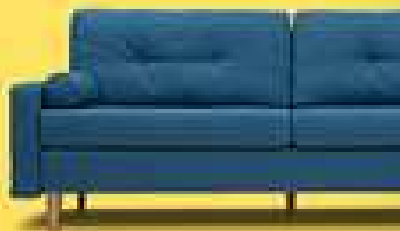
Luna16



Реализация

Контекстная реклама
в интернете часто предлагает
нам то, что мы уже купили,
потому что модель вынуждена
что-то рекомендовать, даже
когда это не нужно

Диван.
интернет-магазин мебели



Диваны от 9 990 руб.

Выбрать

www.divan.ru

ООО «ДИВАН.РУ», ОГРН: 1157744565450, 129110,
г. Москва, пр-т Мира, д. 79, стр. 1, пом. 21, этаж 2


Анатомические диваны АСКОНА! ×

askona.ru

Диван с матрасом АСКОНА внутри. Рассрочка: 0-0-24. Гарантия 10 лет. Купите онлайн!

[Прямые диваны](#)
[Кровати](#)
[Аксессуары к диванам](#)
[Акции](#)
[Адрес и телефон](#)

RG
НЕВЕЩНЫЙ



ДИВАН-КРОВАТЬ RG

Автомобильная
Ленинская Слобода, 26

HOFF.ru - Диваны от 4 290 руб! – Скидки на Диваны!

hoff.ru/купить-диван-hoff Реклама

Угловые и раскладные. Гарантия. Подъем + Доставка 1 день!

HOFF- 70 Угловых Диванов 100 Раскладных Диванов Кредит Дарим 500р.

Контактная информация - +7 (495) 988-57-42 - пн-вс 10:00-22:00

★★★★★ Магазин на Маркете - Москва

Купите диван в Москве от 3000 р. / nedorogdivan.ru

nedorogdivan.ru Реклама

Диваны от 3600 руб. Доставка по Москве, сборка, подъем на лифте Бесплатно!

Доставка за 1 день Диваны-книжки Диваны-аккордеоны Гарантия

Контактная информация - +7 (495) 740-22-91 - пн-вс 8:00-23:45 - Москва

Купите диван в магазине АСКОНА / askona.ru

askona.ru/диван-купить Реклама


Независимый пружинный блок. Гарантия 10 лет. Сборка и доставка - бесплатно!

Прямые диваны Угловые диваны Аксессуары к диванам Акции

Контактная информация - +7 (800) 200-40-90 - пн-вс 9:00-21:00

★★★★★ Магазин на Маркете

Яндекс.Директ



Диван выбирают не сердцем!

Сумасшедший выбор **диванов** в Казани! Заходи!

[Скидки и Акции](#) [Услуги](#) [Рассрочка](#) [Контроль](#)

mgrad.ru Адрес и телефон Казань

IBM Watson Health

**«Маленькие компании
съедают нас живьем»**

Сотрудник IBM WH
spectrum.ieee.org



Когда это было целесообразно, проект был приостановлен. Будучи государственным учреждением, мы решили выйти на рынок для конкурентных предложений, чтобы посмотреть, как развивается отрасль.

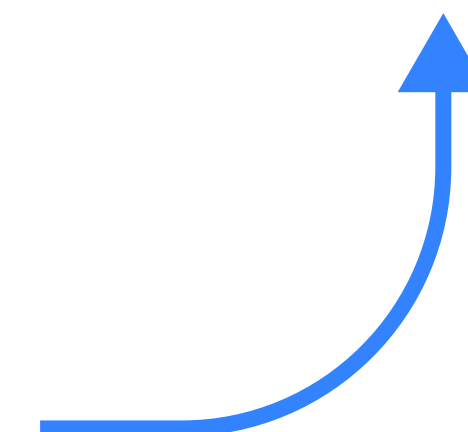
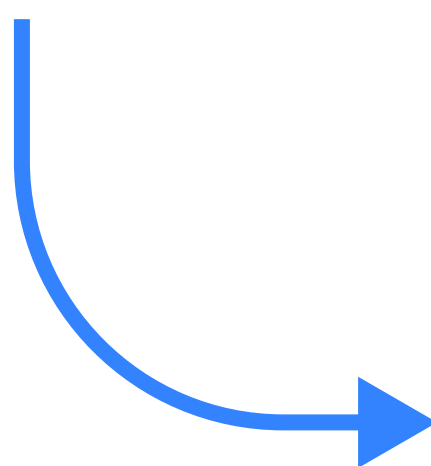
Представитель MD Anderson
forbes.com

IBM проигрывает войну за ИИ-talанты и, вероятно, столкнется с растущей конкуренцией.

Сотрудник IBM WH
spectrum.ieee.org

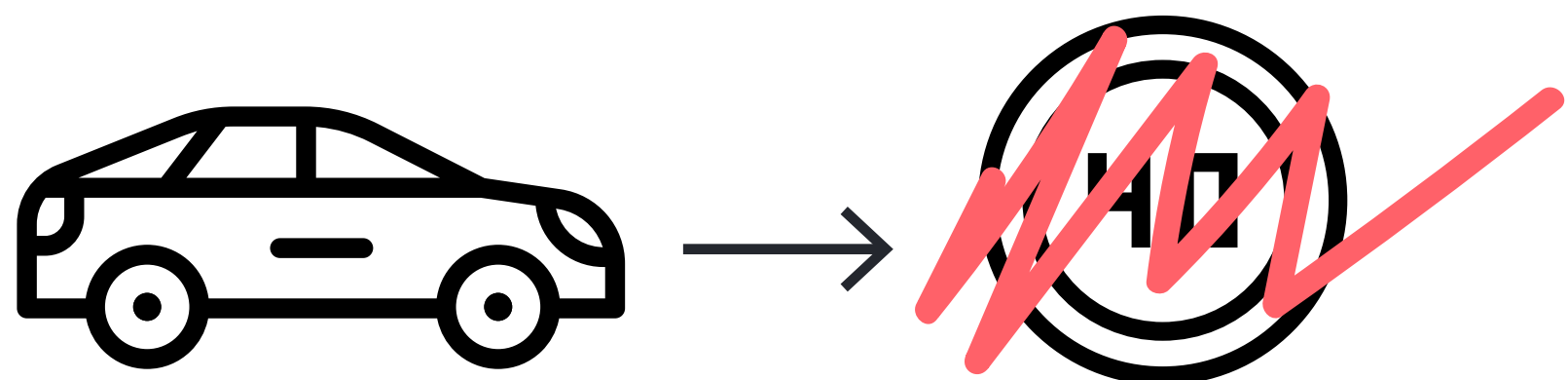


Чат-бот Tay учился общению у пользователей Twitter



Реальность

Если на дорожных знаках
нарисовать что-либо – слова
или рисунки – то системы
распознавания беспилотных
автомобилей начинают
ошибаться



Что мы делаем,
чтобы было все
круто?

