

Искусственный интеллект: требования к данным

Для алгоритмов искусственного интеллекта важно:



Качество: данные валидированы экспертами в предметной области и тщательно размечены при их участии



Объём: большие данные, по классическому определению – это то, что не помещается в таблице Excel (чуть больше 1 млн. строк)



Структура: для числовых данных важна структура и четкая типизация



Ретроспектива: для временных рядов важны данные за продолжительный период

>1 000 000
строк

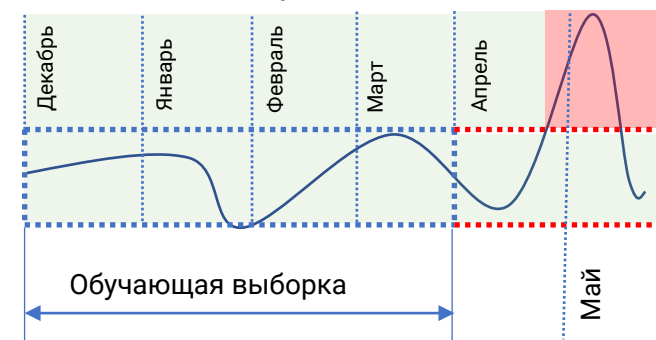
Пример: Качество воды в Калужской области

Собранные объёмные исторические данные о наличии токсических веществ и минеральном составе крайне важны для построения **доверенных** моделей.

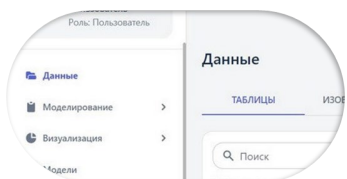
Если такие данные не накоплены, то решение будет безопасным, надёжным, но **неустойчивым (не доверенным)** на больших временных промежутках, например из-за изменения минерального состава из-за сезонных сельскохозяйственных работ или промышленного загрязнения.



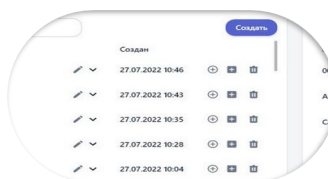
Обработка пестицидами сельхозкультур загрязняет воды



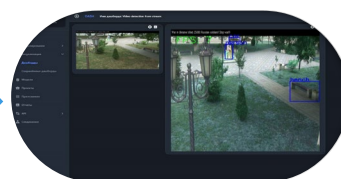
BAUM AI



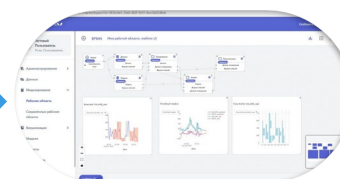
Сбор данных



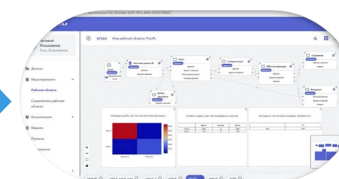
Обработка данных



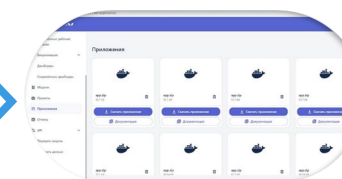
Разметка данных



Ассамблирование алгоритмов



Контроль метрик и версий



Отчуждение приложения

Интеграционная сервисная платформа искусственного интеллекта

