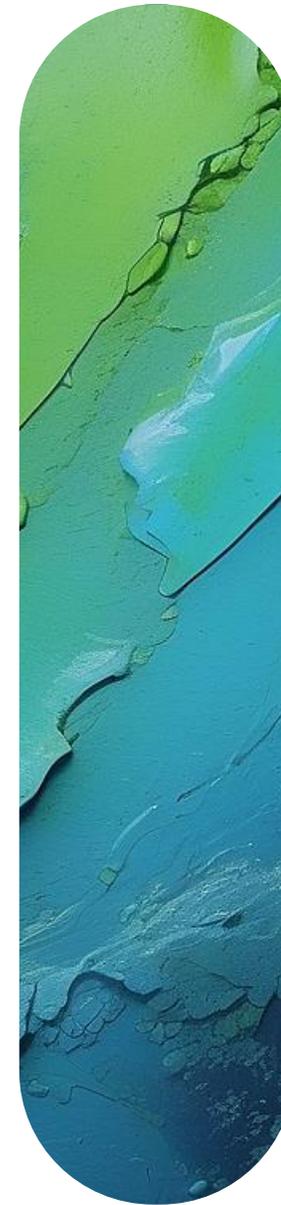


# ML в облаке

Как эффективно управлять процессами машинного обучения на реальных примерах

Никита Гаврилов



# Содержание

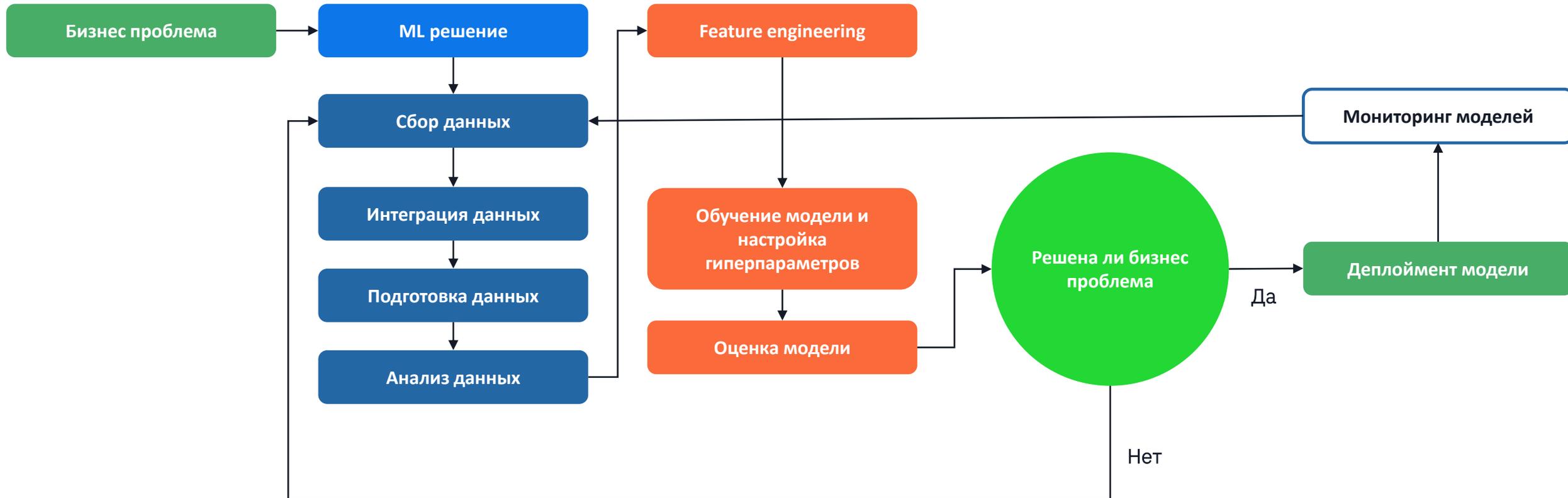
**Какие есть трудности в операционном управлении моделями и почему это вообще важно**

**Почему мы начали внедрять MLOps в облаке**

**Разбор архитектуры на реальной бизнес задаче**

**В чем преимущество и куда идем дальше**

# Машинное обучение это постоянный процесс



# В чем проблема

Чаще всего все эти процессы

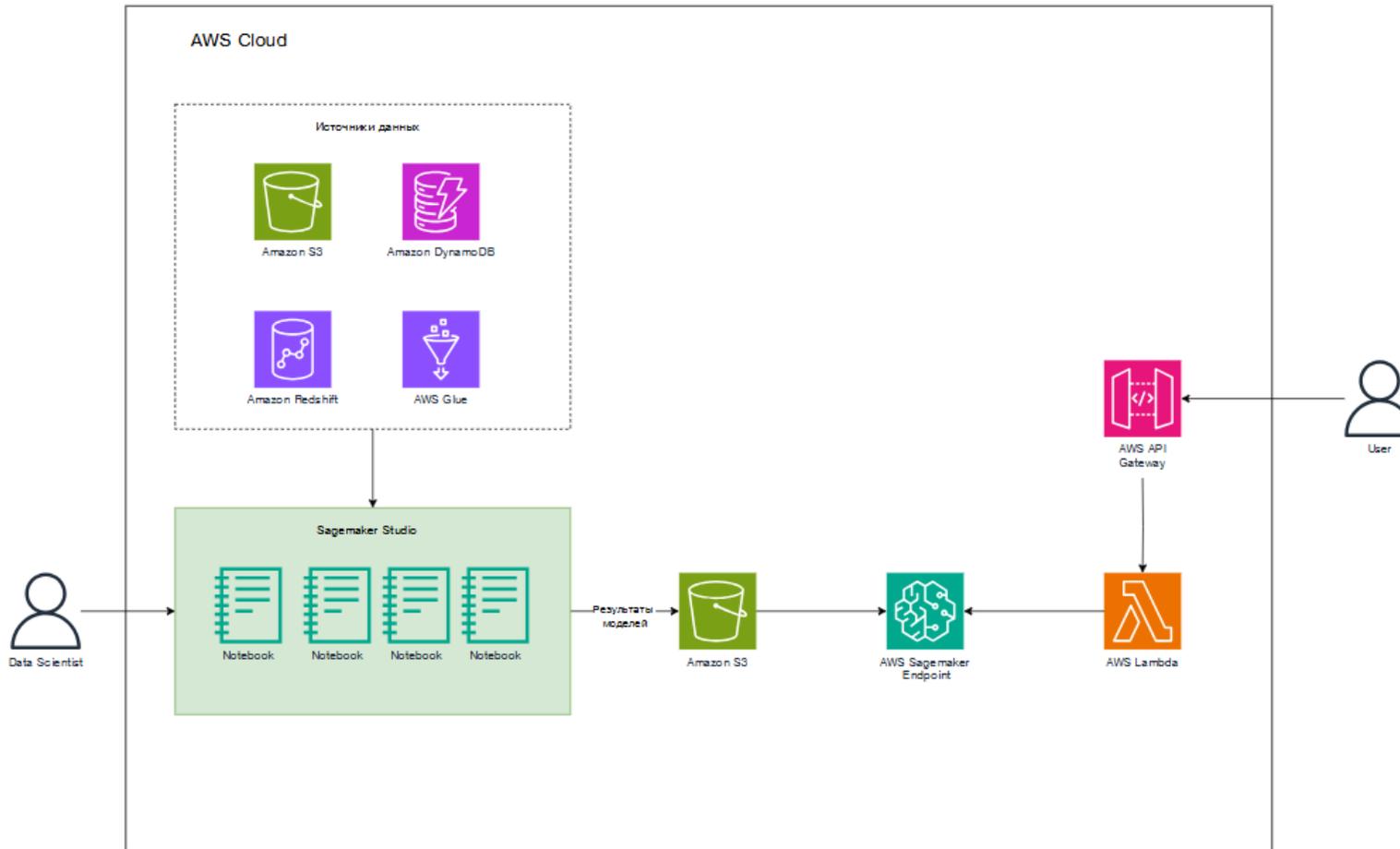
- Работа с данными
- Обучение моделей
- Деплой моделей
- Мониторинг производительности
- Управление моделями
- Переобучение моделей

Делаются в ручном режиме, а некоторые вообще не делаются

При масштабировании это становится реальной проблемой

# Как мы усовершенствовали свой пайплайн

# С чего мы начинали

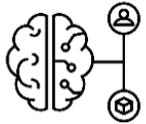


## Базовый подход к ML Flow

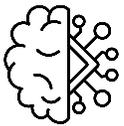
- Весь пайлайн находится в облаке
- Позволяет быстро реализовать доступ к результатам модели
- Бесшовное подключение к данным
- Управляемая среда Jupyter Notebook с предустановленными пакетами Python
- Контроль за расходами на разработку и внедрение моделей
- Множество встроенных алгоритмов ML



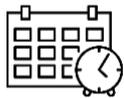
Все изменения делаются руками специалистов DS через изменение Notebook



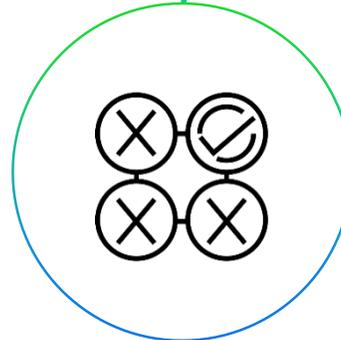
Невозможность автомасштабирования



Очень легко потеряться в Notebook и забыть какая версия модели лучше



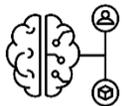
Нет фидбека по перфомансу модели, т.е. мы не понимаем хорошо ли модель работает или нет



## Недостатки

В чем основные минусы такого подхода и почему мы усложнили процесс

# Основные направления проектов по МЛ



Рекомендательная система и  
Next Best offer



Рекомендации  
ассортиментных матриц



Автоматическая модерация  
отзывов на сайте



Рекомендация  
комплементарных товаров



Прогнозирование траффика в  
магазинах



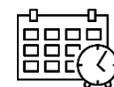
Сопоставление названий  
товаров с товарами  
конкурентов



Прогнозирование продаж



Определение категории опросов  
Net Promotions Score (NPS)



Прогноз количества заказов

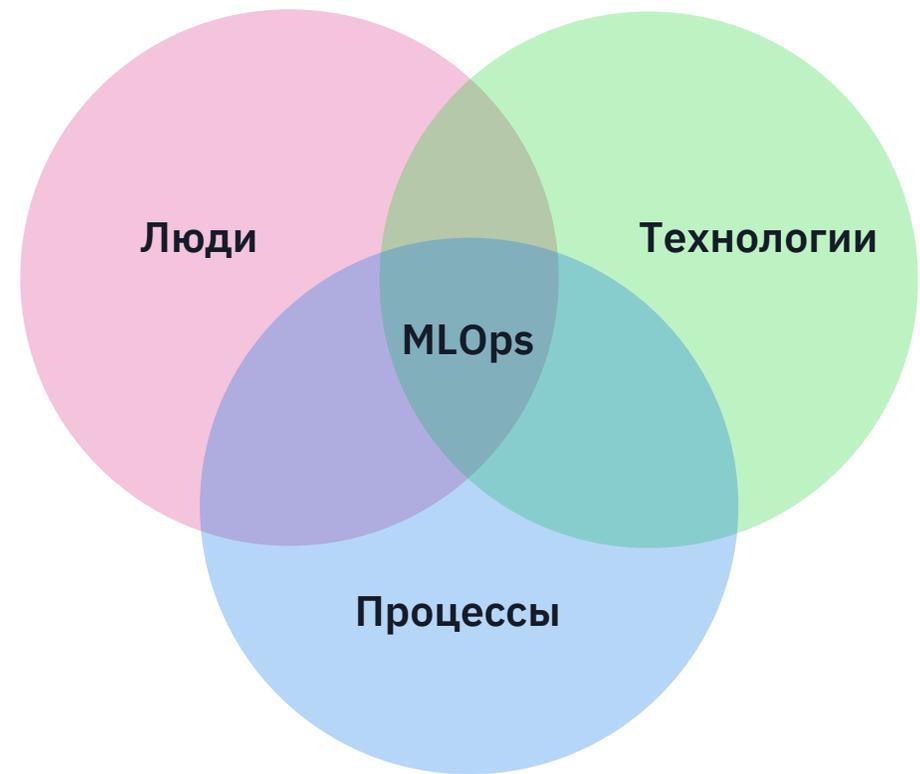


Определение Lifetime value

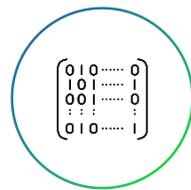


Автоматическое  
распознавание документов

# Внедрение MLOps для end-to-end решения



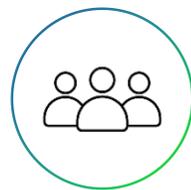
# Внедрение MLOps для end-to-end решения



Набор операционных практик для всего жизненного цикла моделей машинного обучения

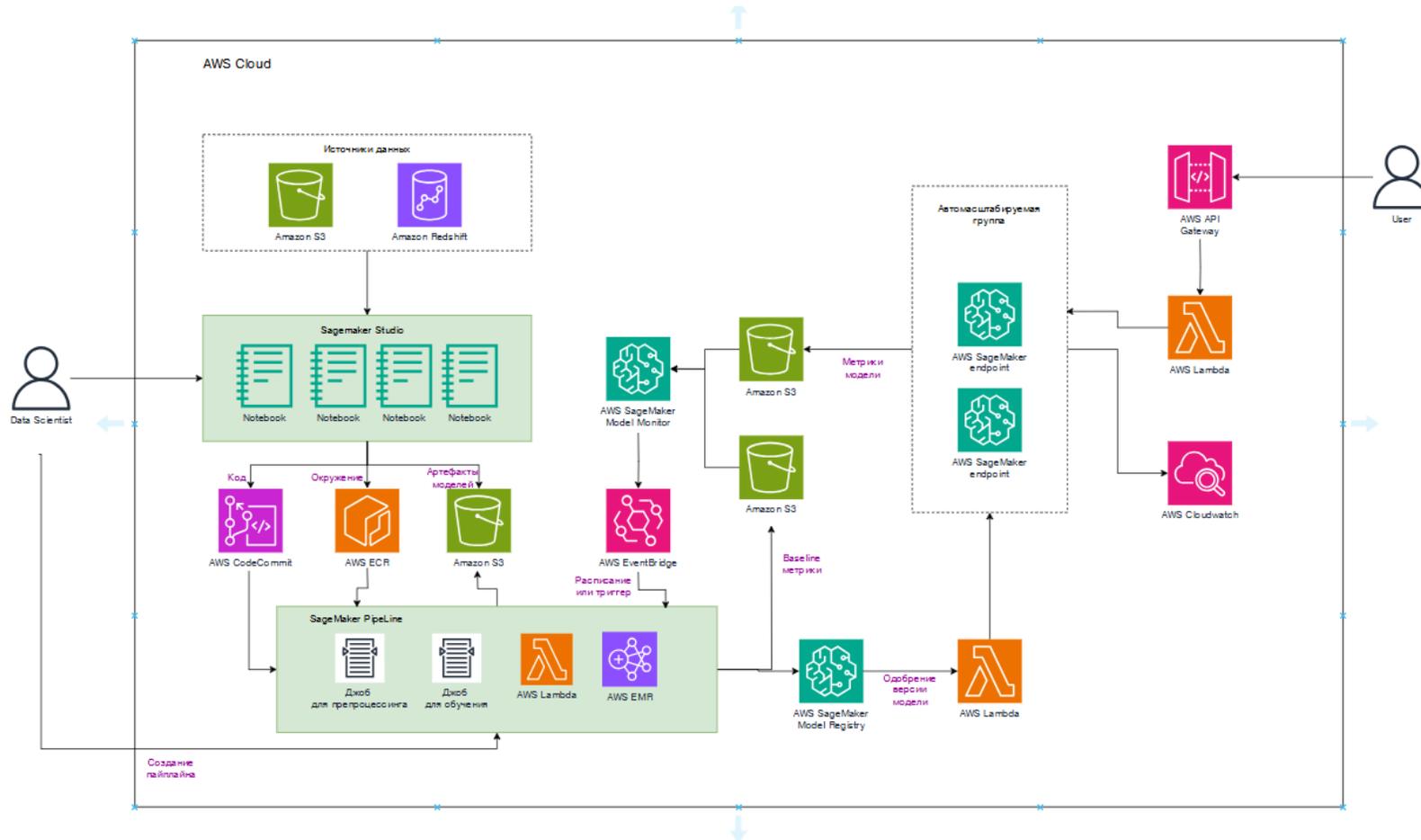


Внедрение MLOps помогло нам увеличить продуктивность и за счет постоянного мониторинга улучшить качество моделей



MLOps помогает обеспечить безопасность всего процесса

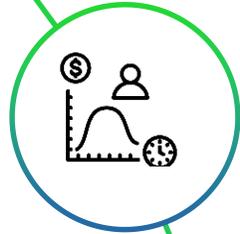
# Обновленная схема ML Flow



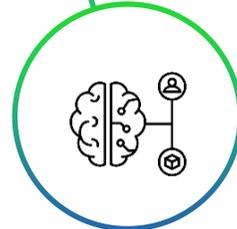
## Чем данный подход стал лучше

- Централизованное управление настройками ML-процессов
- Полный контроль версий (код, данные, модели) на всех этапах
- Автоматическое и гибкое масштабирование инфраструктуры
- Процесс утверждения (апрув) перед деплоем в production
- Постоянный мониторинг качества моделей и производительности системы

# Какой эффект



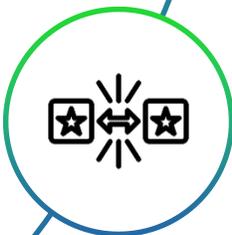
Уменьшили время на переобучение моделей, что позволило легче адаптироваться к изменениям



Автоматизировали процесс **сбора метрик и мониторинга**, что позволяет в моменте реагировать на результаты моделей



Легкое переключение **между версиями моделями** упростило процесс более применимым в разные периоды времени



Детализация процесса **упрощает исправление ошибок**, т.к. легче отслеживается каждый этап

# MLOps это постоянный процесс

О чем мы должны помнить, когда внедряем модели в продакшен

## Воспроизводимость

- Контроль версий кода, данных и моделей
- Четкое документирование этапов обучения и гиперпараметров
- Управление зависимостями
- Автоматизация пайплайнов обучения и развертывания

## Надежность

- Мониторинг производительности моделей и инфраструктуры
- Обработка ошибок и автоматический откат версий
- Алерты и уведомления о проблемах или аномалиях

## Масштабируемость

- Эффективное распределение ресурсов
- Возможность быстрого добавления новых моделей и версий
- Гибкая интеграция с другими системами и сервисами

# Спасибо за внимание

Никита Гаврилов

LinkedIn

