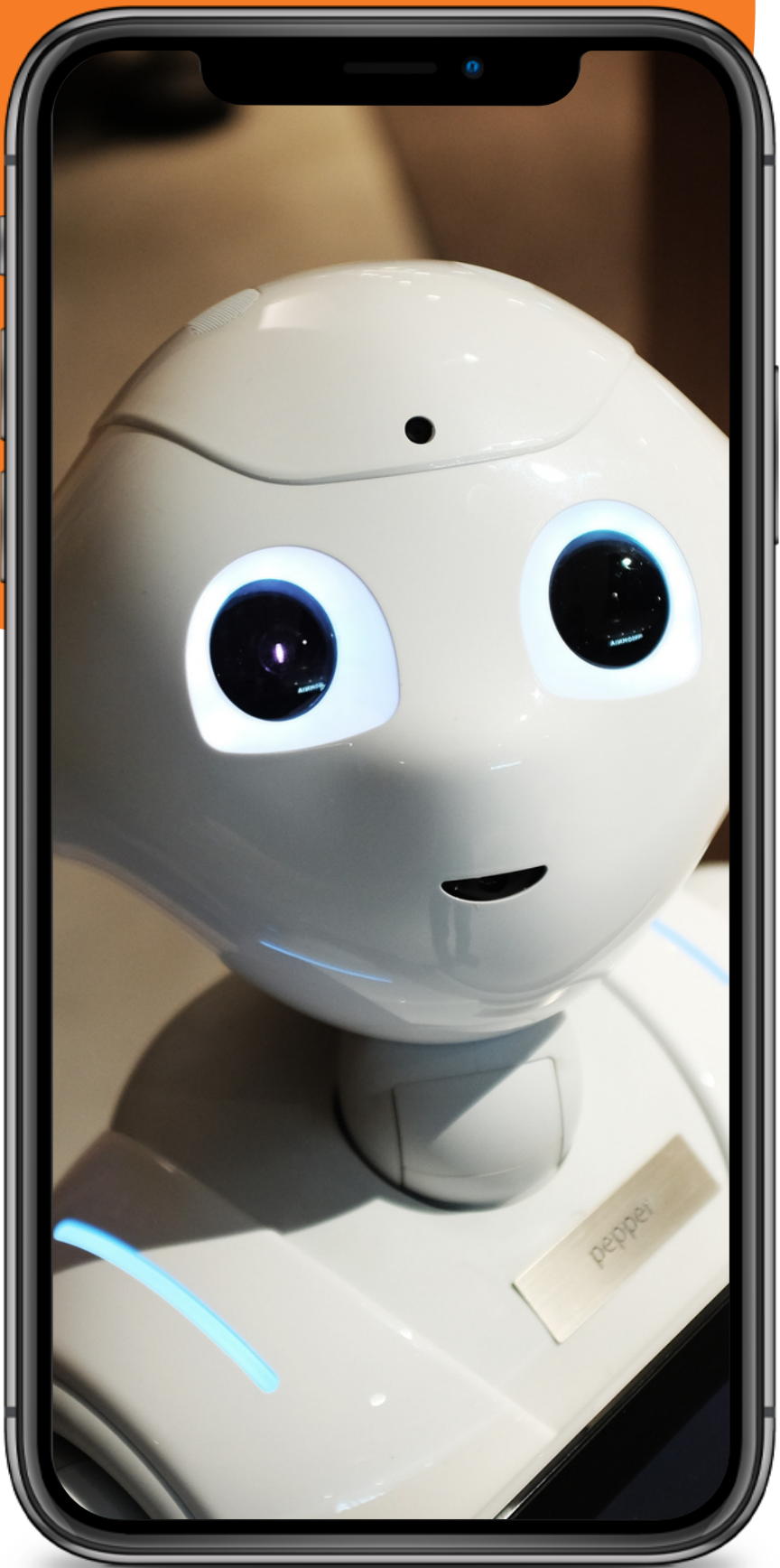


A large orange circle is positioned on the left side of the image, partially overlapping the text 'Hextech' and 'трек Data Engineering'.

Hextech

трек Data Engineering

НАША КОМАНДА



Александр Зайцев
капитан, Data scientist
@Soloeagh



Дарья Лютова
Data scientist
@DariaLyutova
gromovad@mail.ru



Анастасия Павлова
Junior тестировщик
ghipopotamus@mail.ru



Виктор Писарев
Data scientist
@Victor_Pisarev,
pisarev.victor2017@yandex.ru

ПРОБЛЕМА

ДАННЫЕ АБОНЕНТОВ "БИЛАЙН"

01

Трансформация и
обогащение данных

02

Скорость и
производительность

03

Хранение и дальнейшее
использование

Решение

01

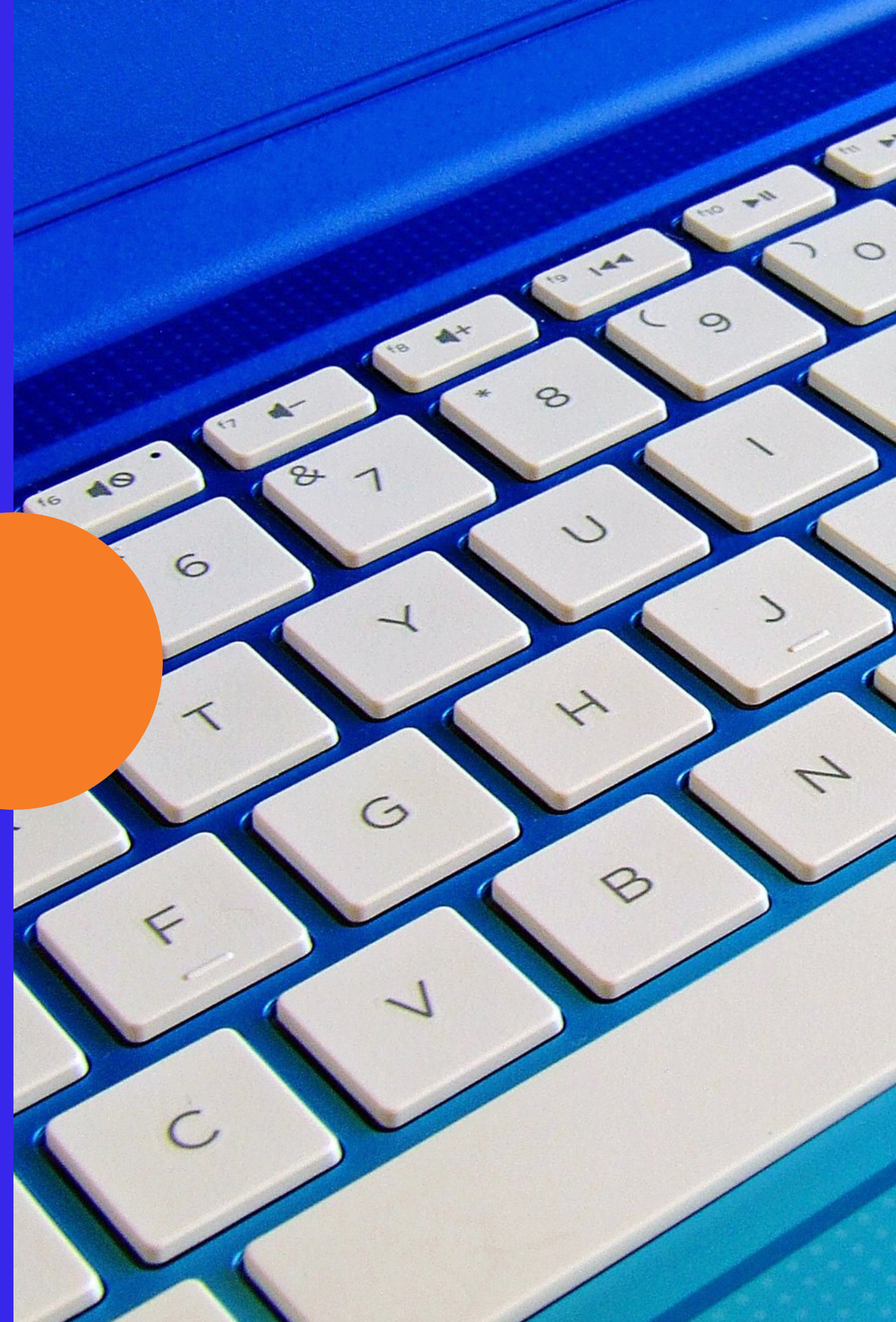
Строгая схема обработки данных

02

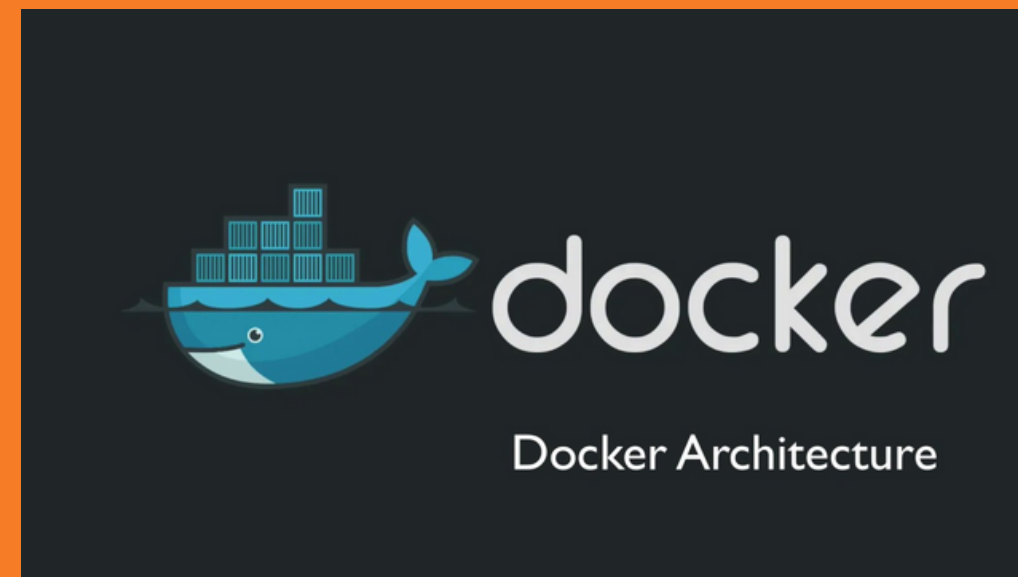
Использование высокопроизводительных технологий

03

Упаковка в Docker



Стек технологий



APACHE FLINK VS SPARK

надежность, масштабируемость и отказоустойчивость

режим работы с данными:

- микропакеты - SPARK
- поточный и пакетный режим - FLINK

зрелость, поддержка сообщества и реализация в коммерческих продуктах - SPARK

поддержка SQL - SPARK (больше)



ЭТАПЫ РАБОТЫ

01

Изучение
исходных
данных



02

Пакетная
обработка данных
(Apache Spark)

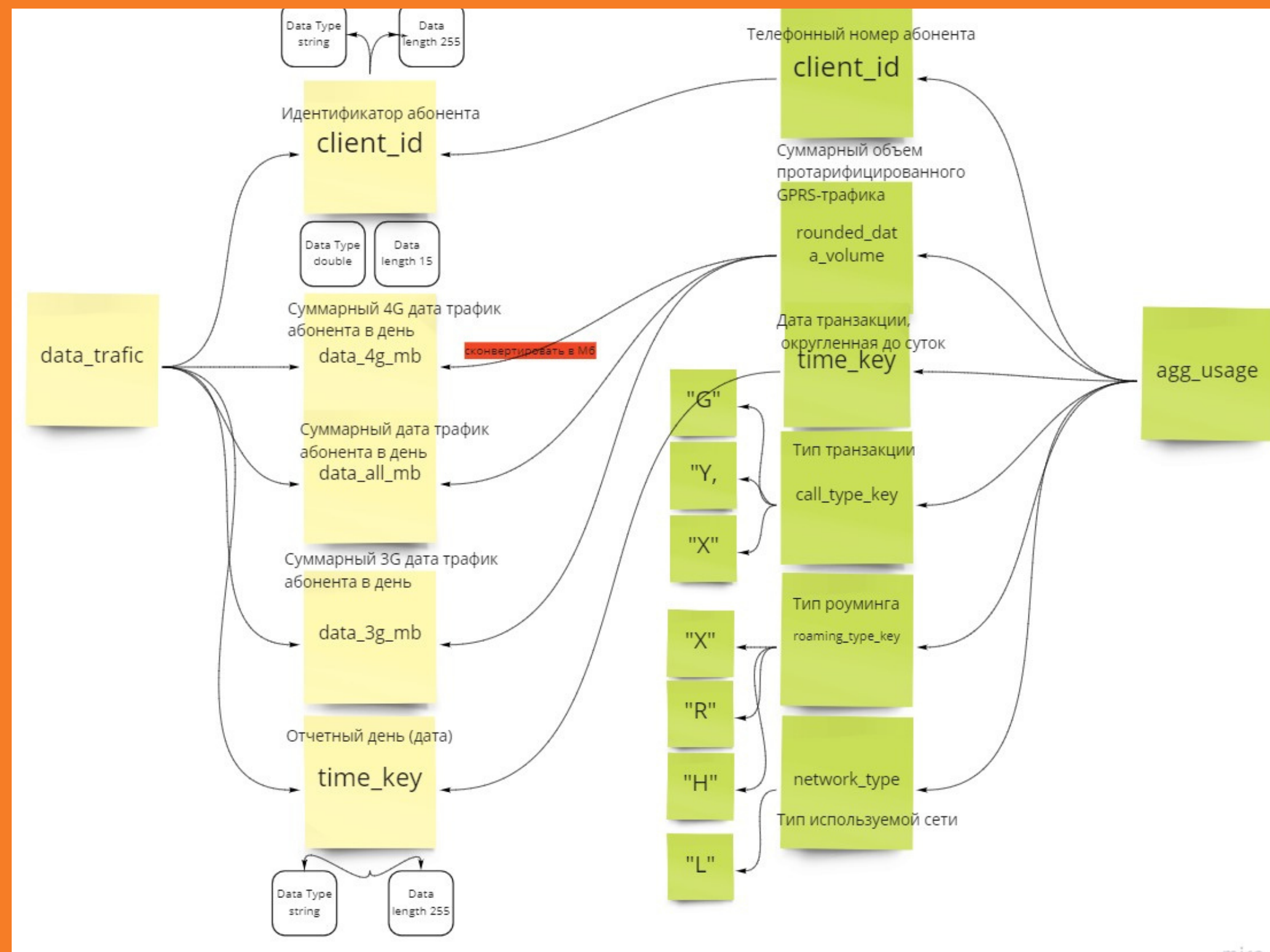
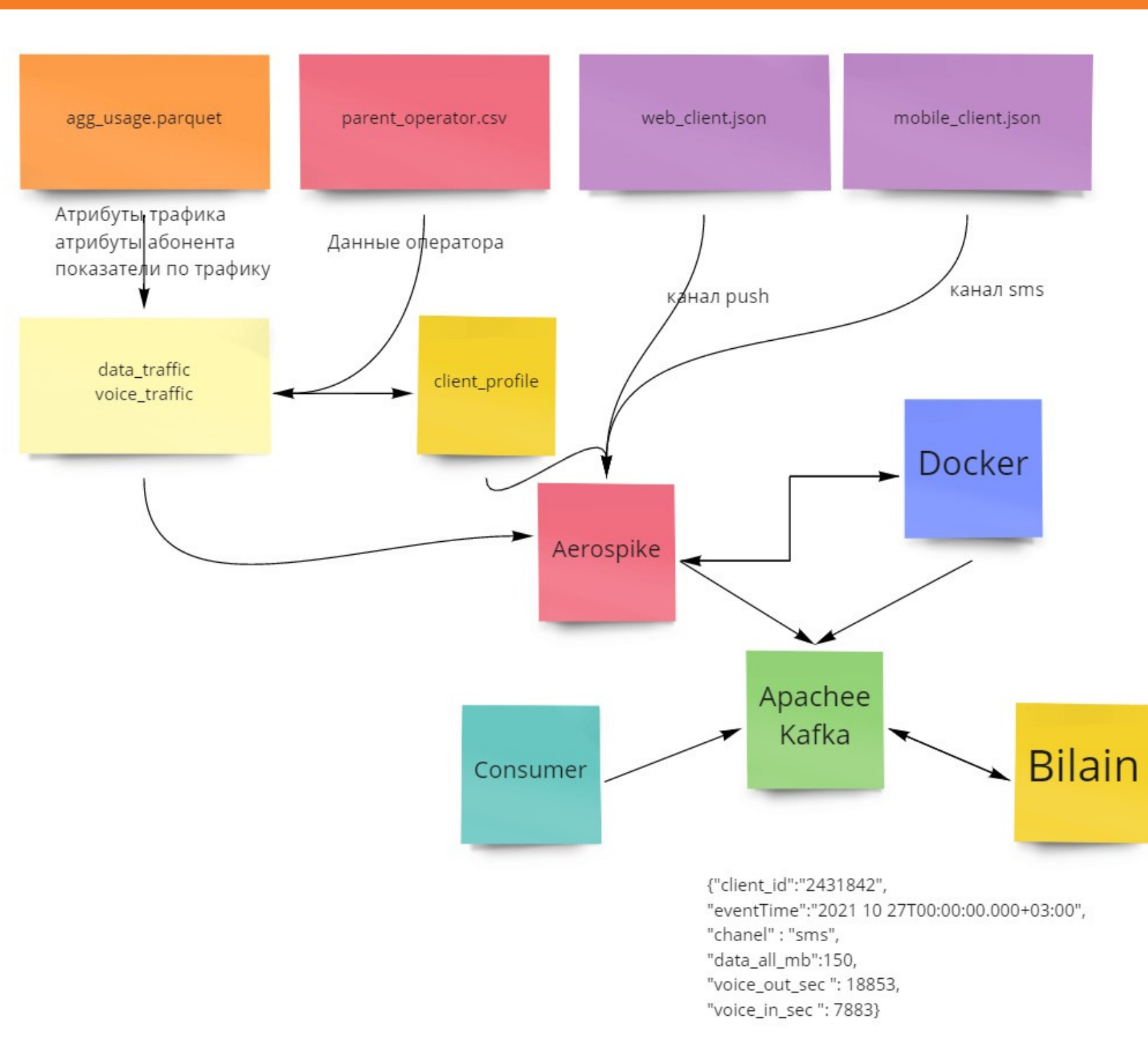


03

Информация о
клиентах



СХЕМА РАБОТЫ



ПОВЫШЕНИЕ ЛОЯЛЬНОСТИ КЛИЕНТА

01

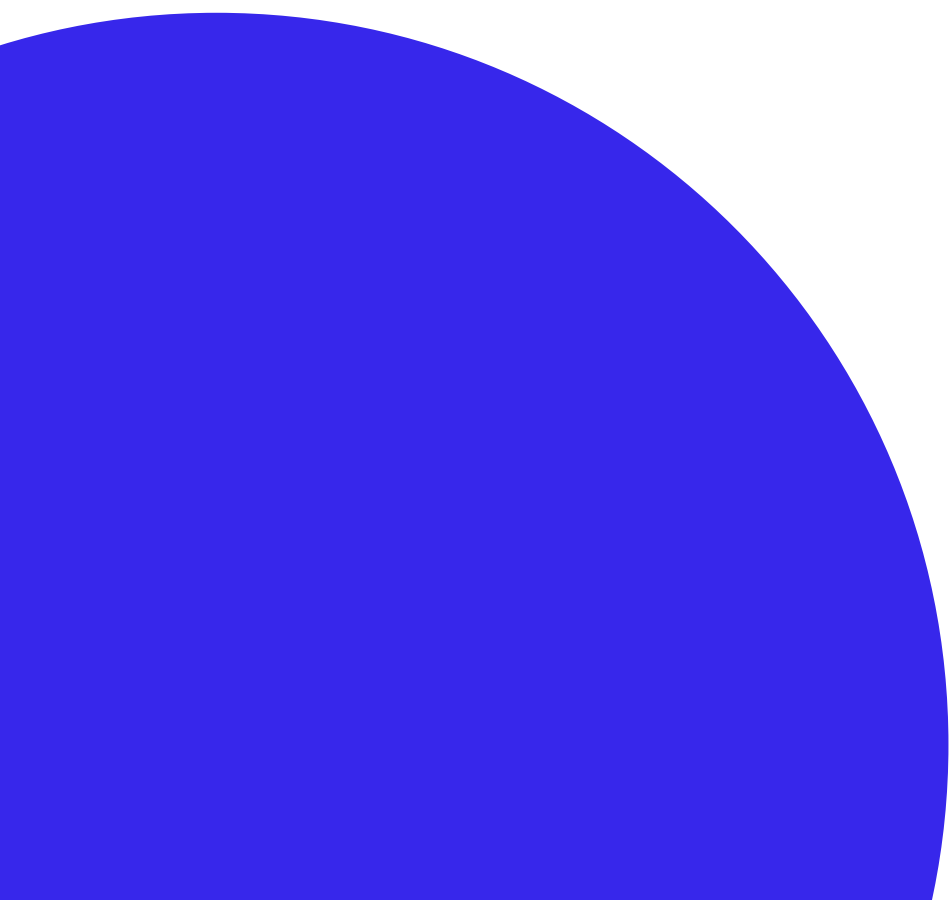
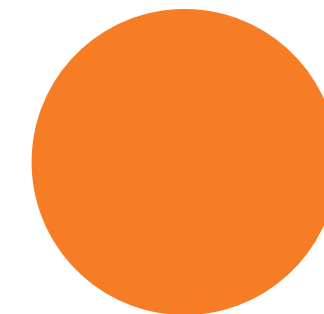
Точность подбора
аудитории

02

Оптимизация коммуникаций
с клиентом:
канал, количество

03

Минимизация оттока
клиентов



Что дальше?

01

Таргетированная реклама

02

Тепловые карты

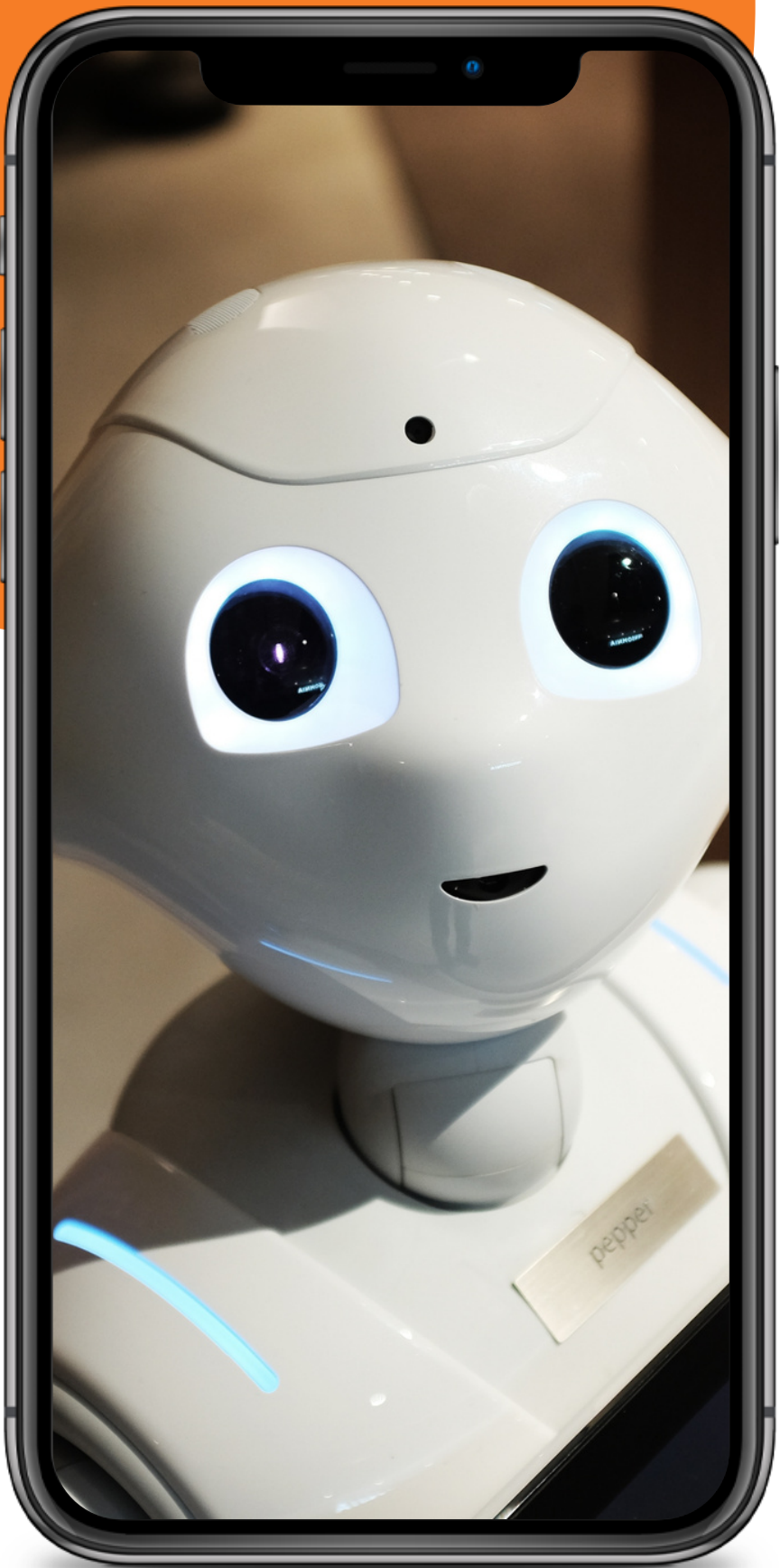
03

Иная информация о
пользователе

- поисковые системы
- информация о путешествиях
- время пользования услугами связи



НАША КОМАНДА



Александр Зайцев
капитан, Data scientist
@Soloeagh



Дарья Лютова
Data scientist
@DariaLyutova
gromovad@mail.ru



Анастасия Павлова
Junior тестировщик
ghipopotamus@mail.ru



Виктор Писарев
Data scientist
@Victor_Pisarev,
pisarev.victor2017@yandex.ru