



Как платформизация и AI меняют цикл аналитической разработки. Опыт Т-Банка

SmartData 2026

Дмитрий Руднев

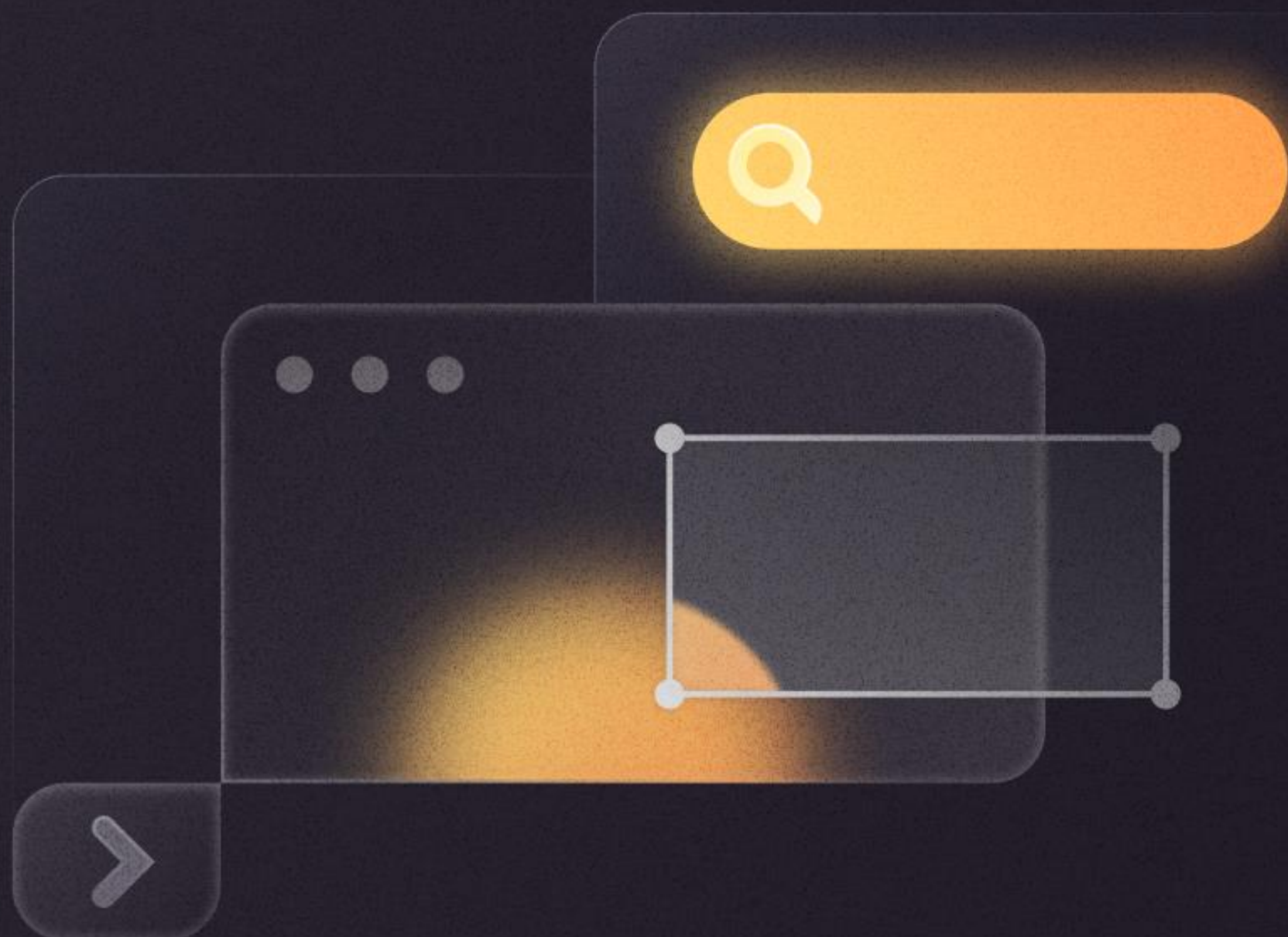
Руководжу разработкой инструментов
для работы
с данными в Data Platform Т-Банка

Telegram: @dorudnev

LinkedIn: dmitryrudnev92



Поговорим о...



- 01 Платформа данных как цикл ADLC
- 02 Тренды в data — почему AI ничего не поменял
- 03 Как меняется платформа данных в Т-Банке
- 04 Цифры, метрики — в поисках эффектов от внедрения AI

Data Platform Т-Банка

~17 000

MAU

~12 000

WAU

~6 780

Пишут код в неделю

52%

Пользователей от всех
digital-профессий

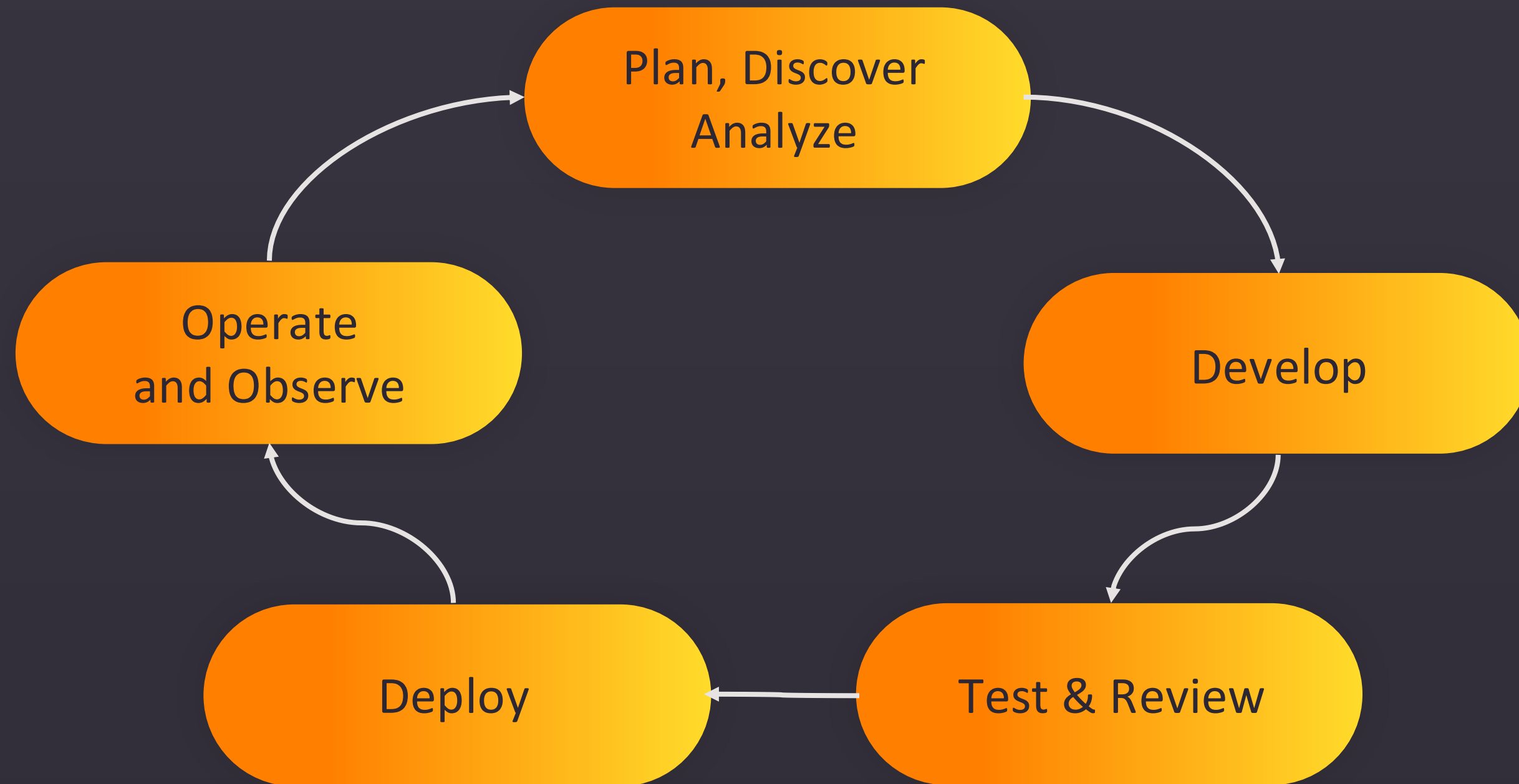
~42 000

Запусков ноутбуков в неделю

~9 000

Новых ноутбуков создаётся
в неделю

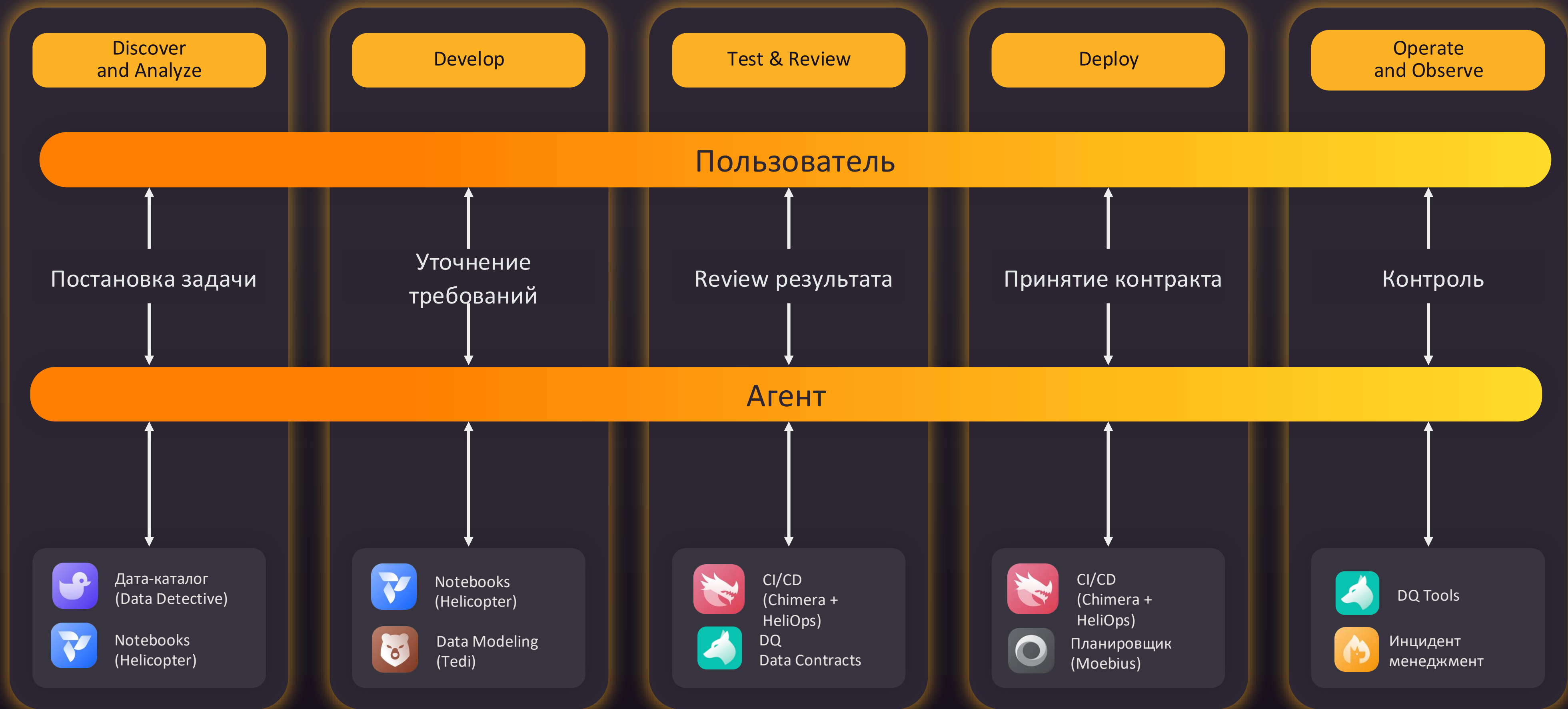
Analytics Development Lifecycle (ADLC)



ADLC в Т-Банке, на примере ETL



ADLC в агентском режиме



AI усиливает и сильные, и слабые стороны систем одинаково

DORA, State of AI-assisted Software Development 2025: «AI — амплификатор.
Он усиливает сильные стороны сильных организаций и дисфункции слабых»

Тренды в Data

- 01 Демократизация и Self-Service
- 02 Продуктивность работы аналитиков
- 03 Описание данных
- 04 Data Contracts и Data Quality
- 05 Платформизация

Ни один из них не появился в 2026-м с приходом AI

Демократизация и Self-Service

До AI

No-Code / Low-Code:

пытались упростить сложную
систему до набора кнопок

В эпоху AI

Agentic AI:

Задача формулируется словами,
агент собирает процесс сам

Повысился уровень абстракций для пользователя

Gartner, Hype Cycle for Agentic AI 2026 — пик завышенных ожиданий: 17% внедрили, 60%+ планируют за два года.
Low-code с Hype Cycle давно съехал в Magic Quadrant (2025) — рынок зрелый, но проблему он не решил.

Продуктивность работы аналитиков

До AI

Скорость брали практиками:
шаблоны, переиспользование, code
review, инженерная культура

В эпоху AI

Код пишется «дёшево»
Дефицитом стало бизнес-знание,
а не руки

Дефицит сместился: не руки, а знание домена

dbt Labs, 2026 State of Analytics Engineering: 72% команд приоритизируют AI-assisted coding
— самый массовый приоритет года

Data Contracts и Data Quality (DQ)

До AI

Контракты считали бюрократией
Data Quality был реактивным: сломался
отчёт — пошли чинить

В эпоху AI

Агент больше доверяет данным
Генерировать проверки стало просто,
но качественный DQ стал необходимостью

Контракт и DQ — необходимость для предсказуемой работы агентов

Gartner, Hype Cycles 2026: data contracts — innovation trigger, при оценке «high benefit».

Рядом Data Quality уже зрелый рынок — Magic Quadrant for Augmented Data Quality Solutions 2026

Описание данных

До AI

Описывали для людей

Эффект — на уровне организации
и не заметен сразу. Мотивации мало

В эпоху AI

Описание стало базовым
условием для работы агента
Эффект виден в тот же день

Метаданные перестали быть документацией. Они стали рантаймом

Top Data & Analytics Predictions 2025: к 2027 семантика в AI-ready данных даёт до +80% точности GenAI.

В Hype Cycles 2026 семантический слой — slope of enlightenment, 20–50% проникновения, 2–5 лет до плато.

Платформизация

До AI

Десятки лучших в своём классе,
но разрозненных инструментов

В эпоху AI

Единый контур вокруг агента
Что под капотом — становится менее
важно

Единый контур собирает не «вендор», а агент

Gartner (2025): к 2028 году 80% GenAI-приложений будут строить не на новом стеке, а поверх дата-платформ, которые в компании уже есть — сложность и сроки при этом падают вдвое.

**AI в платформе
данных Т-Банка
год назад**

Середина 2025 — AI-фичи по месту

Генерация и исправление кода

Визуализация aiChat.sendMessage по неделям (Plotly)2 мин 1 сек▶ Запустить

Type: python

Type '/' for commands and '@' for select table

/ Action +Generate

```
1 from dal_client import query as dal
2 import plotly.graph_objects as go
3 import pandas as pd
4
5 # === Цветовая палитра и константы ===
6 PLOTLY_PALETTE = ['#4C72B0', '#DD8452', '#55A868', '#C44E52', '#8172B3', '#937860']
7 GREY = '#BDBDBD'
8
9 # Получаем данные через DAL с учётом переименования auto → ask/agent
10 df = dal.get_df(
```

Генерация CRON

Конфигурация публикации

Крон

Каждый день в 12:00

✕⚡

Генерация названий параграфов

Сохранение: события по неделям⚡

1 с

Сгенерировать название

Type: python

Конец 2025 — свой агент в каждом продукте

Агент в ноутбуках

Ноут

Разработка

Помощь

Dev

Prod

DWS-AI-TEAM / Helicopter users

Добавьте ваш первый элемент формы

+ Динамическую форму

Данные по пользователям AI и специализациям

1 сек

▶ Запустить

🔍

📄

🔄

⋮

Type:

gp

1 SELECT

2 specialization,

3 COUNT(DISTINCT username) as total_users,

4 SUM("/ai_chat_users") as ai_chat_users,

5 SUM("/find_tables_users") as find_tables_users,

6 SUM("/fix_users") as fix_users,

7 SUM("/text2sql_users") as text2sql_users,

8 SUM(ai_in_par) as ai_in_par_users,

9 SUM(use_chat_at_code) as use_chat_at_code_users

10 FROM usr_hl.helicopter_ai_users_x_specializations

11 WHERE specialization IS NOT NULL

12 AND specialization != ''

13 GROUP BY specialization

14 ORDER BY total_users DESC

15 LIMIT 100

16

usr_hl.helicopter_ai_users_x_specializations

LN 1, COL 1 CHARS: 513

Text

Table-1

specialization

total_users

ai_chat_users

find_tables_users

fix_users

1

Бизнес-аналитик (business)

1,201

7,270

241

1,245

Процент использования helicopter в di...

Прочитал вывод: 1 outputs, 1445 символов

Размышления >

Результат анализа

58.01% от всех digital профессий используют Helicopter.

Детальные цифры:

Показатель	Значение
Всего сотрудников с digital профессиями (IT + Digital)	20,412
Пользователей Helicopter с digital профессиями	11,842
% использования Helicopter	58.01%

Распределение по специализациям:

- IT: 14,013 сотрудников → 7,258 используют Helicopter (51.8%)
- Digital: 6,399 сотрудников → 4,584 используют Helicopter (71.6%)
- NonIT: 53,479 сотрудников → 5,249 используют Helicopter (9.8%)

Общее количество пользователей Helicopter: 22,745

Данные получены из:

find_tables

text2sql

Введите ваше сообщение здесь, используйте '/' для команд

+ Agent

Агент в каталоге данных

Data Detective

Tables / ... / chrono_data_detective / chrono_data_detective.entity_dd2_catalog

Nestor Data

data detective

Greenplum

chrono_data_detective.entity_dd2_catal...

Сущность DD, карточка DD - ключевой элемент, который вы видите в Data...

DLH

chrono_data_detective.entity_dd2_catal...

Сущность DD, карточка DD - ключевой элемент, который вы видите в Data...

Greenplum

chrono_data_detective.entity_attribute...

Ресурс атрибута

DLH

chrono_data_detective.entity_attribute...

Ресурс атрибута

DLH

chrono_data_detective.entity_summary...

Ресурс описания

Greenplum

chrono_data_detective.entity_summary...

Ресурс описания

DLH

chrono_data_detective.entity_dd2_catal...

Сущность DD, карточка DD - ключевой элемент, который вы видите в Data...

Greenplum

chrono_data_detective.entity_dd2_catal...

Сущность DD, карточка DD - ключевой элемент, который вы видите в Data...

DLH

chrono_data_detective.relation_dd2_ca...

Связь сущностей

Greenplum

chrono_data_detective.relation_dd2_ca...

Связь сущностей

DLH

chrono_data_detective.entity_ownership...

Связь сущности с владельцем

DLH

chrono_data_detective.entity_markdown...

Ресурс markdowna

chrono_data_detective.entity_dd2_catalog

Сущность DD, карточка DD - ключевой элемент, который вы видите в Data Detective, и с чем работаете. В данной таблице хранятся базовые ключи и атрибуты. Вся детализация хранится в элементах Ресурсами: ресурс кода, ресурс метрик и тд.

DLH

chrono_data_detective.entity_dd2_catalog

Edited 6 days ago

data-contract

Structure

Physics

Relations

Lineage & Usage

Attributes

Data product	Объекты хранения данных
Tenant	https://devplatform.tcsbank.ru/tenants/mg
Domain	data_detective
Data Contract UI	https://devplatform.tcsbank.ru/tenants/mg/data-contracts/detail/a2ce472a-1bcf-4b9e-a97c-1df07473dee7
Data Contract Schema	https://gitlab.tcsbank.ru/mg/dd-processor/-/tree/master/datacontracts/entity/entity.avdl
Data Contract Version	1
Owner Email	data-detective-adm-adm@tbank.ru
Owner Time	data-detective-contract
Owner AD Group	data-detective-adm
Links	https://wiki.tcsbank.ru/display/DW/Data+Detective
TTL	Не ограничен
Delete flag	False

Table structure 18

Keys	Sensitivity	Name	Description	Jira	Sour
------	-------------	------	-------------	------	------

"Searching for Data Detective Ta...

Qwen 3 with 235 billion parameters — LLM for working with code and dialogs in the bank.

find_tables найти таблицу с сущностями Data Detective

Qwen 3

dds.dwh_rep_logical_entity

chrono_data_detective.entity_dd2_catalog

dwh_rep.mg_entity

dds.dwh_rep_entity

dds.web_application

Didn't find the tables you need? Tell us about it!

find_tables

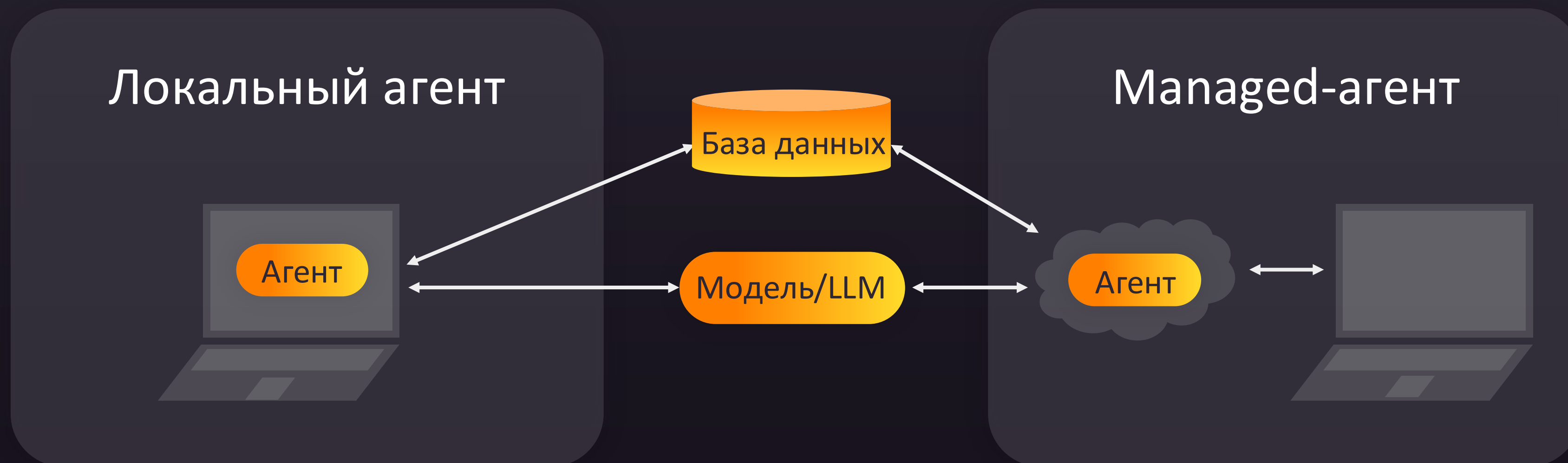
Type your message here, use '/' for command

/ Action

Как AI меняет платформу данных Т-Банка сейчас

Ставка: managed-агент, а не локальный

- Требования ИБ выполняются платформой, а не каждым пользователем
- Длинные задачи не привязаны к сессии пользователя
- Работает из коробки, без настройки плагинов и интеграций



Агент как интерфейс платформы

01

Агент — центральная точка входа. Остальные интерфейсы собираются вокруг него

02

UI не исчезает, пока есть human in the loop

03

Меняется роль UI: с «где я делаю работу» на «где я смотрю и вмешиваюсь»

ADLC через агента



The screenshot shows the Helicopter Data Detective interface. The top navigation bar includes 'Helicopter', 'Администрирование', 'Data Detective', and 'DataWorkspace AI Team'. The left sidebar contains navigation items like 'Nessy Data', 'Недавние', 'Избранное', 'Мои ноуты', 'Все ноуты', 'Коллекции', 'Скилы', 'Переменные', 'Мониторинг', 'Архив', and 'Корзина'. The main workspace is titled 'Решение задачи DWSAI-1569'. It contains a chat history on the left, a search bar, and a list of tasks. A task titled 'Решение задачи DWSAI-1569' is selected, showing a description and a list of steps. A yellow box highlights a specific step: 'Как определять «активного пользователя» Helicopter для расчёта WAU?'. The step includes instructions on how to determine user activity and a 'Введи свой ответ' (Enter your answer) prompt. The right sidebar shows 'Артефакты' (Artifacts) and 'Использованные' (Used) items.

Планирование
и доуточнение
требований

ADLC через агента



Tree view

chronos_gtech_gosotv

chronos_gtech_smev3

chronos_harvester

chronos_helicopter_pd

chronos_helicopter_pd

chronos_helicopter_public

collection_contain

collection_contain

collection_public

collection_public

cron_public

cron_public_log

jobs_view_helicopter

job_id

created_at

creator

cron_flag

ended_at

note_id

paragraph_id

previous_job_id

request_id

resource_group

chronos_helicopter_public.jobs_view_helicopter_engine

DLH

chronos_helicopter_public.jobs_view_helicopter_engine

Edited 4 days ago

Structure

Physics

Relations

Lineage & Usage

prod_chronos_helicopter_public.jobs_view_helicopter_engine

master

table

use_to_query

Rows

Data size

Storage Type

Created At

Last Updated

140,597,513

9.42 GB

iceberg

08.09.2025 13:32 +03:00

20.08.2026 08:48 +03:00

Actuality

Severity

Data metrics

DDL

Actuality

Version time: 04.09.2026 19:09 +03:00

Delivery time: 04.09.2026 19:37 +03:00

Severity

SLA: 1

Data metrics

Checks: 0

DDL

Расчёт WAU и анализ ау...

+

🕒

×

не могу выполнить SQL.

Давайте задам вопросы пользователю через ask_user. Это уместно, так как у меня есть неустранимая неопределённость: нет notebook для выполнения SQL, и неясно, что считать активностью.

Сформулирую вопросы.

Что считать «активностью» пользователя Helicopter для расчёта WAU?

Запуски ноутбуков (jobs)

Активным считается пользователь, запустивший параграфы/ноуты за неделю (jobs_view_helicopter_engine)

UI-события (statist)

Активным считается пользователь, совершавший любые действия в интерфейсе Helicopter (события statist)

Готовые WAU-таблицы

Использовать уже посчитанные таблицы

usr_bi_core.helicopter_nestor_wau_* как основу

Запуски + UI-события

Комбинировать оба источника

1/3

<

Skip

>

Type your prompt here

Agent

🗨

📄

Анализ метаданных в каталоге

ADLC через агента



Нессы Data Beta

Недавние

Избранное

Мои ноуты

Все ноуты

Коллекции

Скилы Beta

Переменные

Мониторинг

Архив

Корзина

ИЗБРАННЫЕ КОЛЛЕКЦИИ

Настройки

История

Новый чат

Поиск

Сегодня

Решение задачи только что

Решение задачи DWSAI-1569

План построения SSETL-процесса для расчёта WAU

1. Архитектура процесса

Хранилище: Greenplum (GP) — весь расчёт и публикация выполняются в GP, т.к. все нужные источники доступны в GP как витрины prod_v_*.

Тип загрузки: SCD1 (полная замена, truncate/insert) — витрина WAU пересчитывается целиком за период.

Структура ноутбука (2 параграфа):

1. SQL-параграф (gp): расчёт WAU во временную таблицу tmp_wau_helicopter

2. Python-параграф: публикация tmp_wau_helicopter → целевая таблица через publish_to_gp

2. Логика расчёта — цепочка джойнов

Источник активности (база):

Unknown

prod_v_chrono_helicopter_public.jobs_view_helicopter_engine (j)

- creator = логин пользователя, запустившего параграф

- created_at = время запуска

- cron_flag = false (только ручные запуски)

- creator NOT LIKE 'sa-%' (исключаем служебные аккаунты)

Обогащение профессией (3 LEFT JOIN):

Unknown

i (jobs_view)

Введите ваше сообщение здесь

Agent (beta)

Артефакты

Созданные 3

Расчёт WAU пользователь Ноутбук

SSETL: Расчёт WAU пользователь Ноутбук

DQ-проверки витрины W Ноутбук

Использованные 10

DWSAI-1569 Задача для DEMO агента

prod_chrono_helicopter_r Документация таблицы

prod_chrono_helicopter_r Таблица отражает выполнение параграфов в ви

prod_v_dds.mdm_employe Информация о профессиях сотрудников

prod_v_dds.dic.employee Версионный справочник по профессиям сотрудников из

prod_v_dds.mdm_employe Единичный сотрудник группы компаний Тинькофф. Источн

prod_v_dds.mdm_employe Учетные записи единого сотрудника из AD, Cisco ISE и

prod_v_emark.mdm_employ Витрина с атрибутами сотрудника как физ. лица и а

prod_v_chrono_helicopter

Валидация плана

ADLC через агента



The screenshot shows the Helicopter Data Detective interface. On the left is a sidebar with navigation items like 'Nessy Data', 'Недавние', 'Избранное', 'Мои ноуты', 'Все ноуты', 'Коллекции', 'Скилы', 'Переменные', 'Мониторинг', 'Архив', and 'Корзина'. The main area is titled 'Решение задачи DWSAI-1569' and contains a chat history on the left and a task description on the right. The task description includes a code block for a SQL query. On the right side of the interface, there is a panel titled 'Артефакты' (Artifacts) with two sections: 'Созданные' (Created) and 'Использованные' (Used). The 'Созданные' section contains one artifact, 'Расчёт WAU пользователей Helicopter (DWSAI-1569)', which is highlighted with an orange box and an arrow pointing to the right.

История

Новый чат

Поиск

Сегодня

Решение задачи DWSAI-1569

Решение задачи DWSAI-1569

сначала проверю метаданные ноутбука и доступные интерпретаторы.

Получил метаданные ноутбука "Расчёт WAU пользователей Helicopter (DWSAI-1569)"

Загрузил интерпретаторы

Ноутбук создан (note_id=1643679), is_runnable=true. Доступны интерпретаторы, включая dlh_trino (для DLH таблиц через Trino) и python.

Таблицы Helicopter (prod_chrono_helicopter_public.*) и справочники (prod_v_dds.*) — это DLH таблицы, доступные через Trino. Использую shebang dlh_trino.

Теперь начну исследование данных. Сначала добавлю параграф для исследования данных jobs_view_helicopter_engine: проверю период данных, формат creator, количество пользователей.

Давайте добавлю первый исследовательский параграф. Согласно skill, исследовательские SELECT — с LIMIT 10, агрегаты — без лимита.

Начну с исследования периода данных и формата creator.

Исследование: период данных и формат creator в jobs_view +14

```
-- Исследование данных jobs_view_helicopter_engine
-- 1. Период данных
SELECT
  MIN(created_at) AS min_created,
  MAX(created_at) AS max_created,
```

Параграф создан. Запущу его для исследования данных.

Вызвал тул: run_paragraphs

Введите ваше сообщение здесь

Agent (beta)

Артефакты

Созданные

Расчёт WAU пользователей Helicopter (DWSAI-1569)

Использованные

DWSAI-1569
Задача для DEMO агента

prod_chrono_helicopter_public.*
Документация таблицы

prod_chrono_helicopter_public.jobs
Таблица отражает выполнение параграфов в виде заданий (jobs)

prod_v_dds.mdm_employee_xref
Информация о профессиях сотрудников

prod_v_dds.dic.employee_profession
Версионный справочник по профессиям сотрудников из Оргструктуры

prod_v_dds.mdm_employee
Единый сотрудник группы компаний Тинькофф. Источник Референт

prod_v_dds.mdm_employee_log
Учетные записи единого сотрудника из AD, Cisco ISE и устаревших Slack

prod_v_emart.mdm_employee_skills
Витрина с атрибутами сотрудника как физ. лица и атрибутами юридич.

Переход в ноутбук

ADLC через агента



НоутРазработкаПомощьDevProd

DWS-AI-TEAM / Расчёт WAU пользователей Helicopter (DWSAI-1569)

Запустить всеОпубликовать

Исследование: покрытие пользователей профессиями

6 сек

Запустить

22 FROM prod_chrono_helicopter_public.jobs_vie

23 LEFT JOIN prod_v_ddm_employee_login mel

24 ON lower(mel.login) = lower(j.creator)

25 AND mel.login_system_cd = 'ADR'

26 AND mel.deleted_flg = 0

27 LEFT JOIN prod_v_ddm_employee_x_profess

28 ON mxp.mdm_employee_rk = mel.mdm_employ

29 AND mxp.deleted_flg = 0

30 LEFT JOIN prod_v_ddm_dic.employee_professio

31 ON ep.employee_profession_dk = CAST(mxp

32 AND ep.deleted_flg = 0

33 WHERE j.created_at >= DATE '2026-06-13'

34 AND j.cron_flg = false

35 GROUP BY ep.employee_profession_nm

36 ORDER BY users DESC

37 LIMIT 15;

31+ WHERE deleted_flg = 0

32+)

33+SELECT p.employee_profession_nm, COUNT(DISTINCT

34 FROM prod_chrono_helicopter_public.jobs_view_he

35 LEFT JOIN prod_v_ddm_employee_login mel

36 ON lower(mel.login) = lower(j.creator)

37 AND mel.login_system_cd = 'ADR'

38 AND mel.deleted_flg = 0

39 LEFT JOIN prod_v_ddm_employee_x_profession i

40 ON mxp.mdm_employee_rk = mel.mdm_employee_r

41 AND mxp.deleted_flg = 0

42+LEFT JOIN prof p

43+ ON p.employee_profession_code = mxp.profess

44+ AND p.rn = 1

45 WHERE j.created_at >= DATE '2026-06-13'

46 AND j.cron_flg = false

47+GROUP BY p.employee_profession_nm

48 ORDER BY users DESC

49 LIMIT 15;

CancelAccept

LN 1, COL 1 CHARS: 1552

TextTable-1Table-2

	total_creators	creators_with_profession
1	16 886	13 851

Решение задачи DWSAI-1569

Шаг 3 — Промежуточная аналитика и графики (Python, matplotlib/plotly)

Шаг 4 — SSETL-процесс (skill ssetl-autocomplete-python) + расписание

Шаг 5 — DQ-проверки (отдельный ноутбук)

Уточню пару моментов перед расчётом:

Исключать ли служебные аккаунты (префикс 'sa-') из расчёта WAU?

Исключать sa-*

Исключать все creator с префиксом 'sa-' (service account) из WAU — это автоматика, не люди.

Не исключать

Не исключать служебные аккаунты, считать их как обычных пользователей.

Введи свой ответ

1/3

Skip

find_tablestext2sql

Изменено 2 параграфовОтклонитьПринять

Введите ваше сообщение здесь, используйте '/' для команд

+ Agent

Анализ кода в процессе разработки

ADLC через агента



НоутРазработкаПомощьDevProdDWS-AI-TEAM / Расчёт WAU пользователей Helicopter (DWSAI-1569)Запустить всеОпубликоватьDR

Графики: динамика WAU и разбивка по профессиям1 секЗапуститьhelicopter-python-lsp LN 1, COL 1 CHARS: 1548

TextImage-1Image-2

WAU пользователей Helicopter (ручные запуски, без sa-*)

Неделя (начало)	WAU (уник. пользователей)
2026-06-15	9574
2026-06-22	9499
2026-06-29	9684
2026-07-06	9537
2026-07-13	9460
2026-07-20	9393
2026-07-27	9449
2026-08-03	9595
2026-08-10	9490
2026-08-17	9602
2026-08-24	9715
2026-08-31	9810

Решение задачи DWSAI-1569

Неделя	WAU
2026-06-15	9574
2026-06-22	9499
2026-06-29	9684
2026-07-06	9537
2026-07-13	9460
2026-07-20	9393
2026-07-27	9449
2026-08-03	9595
2026-08-10	9490
2026-08-17	9602
2026-08-24	9715
2026-08-31	9810

WAU стабилен ~9400-9800, лёгкий рост к концу периода.

Топ-15 профессий по среднему WAU:

check query for sens datafixexplainfind_tables text2sql

Изменено 3 параграфовОтклонитьПринять

Введите ваше сообщение здесь, используйте '/' для команд

+ Agent ↩

Review данных

ADLC через агента



НоутРазработкаПомощьDevProd

SSE_IDWH_D... / HelicopterWAU

DR 🔍 📄 🔔 ▶ Запустить все ☁ Опубликован ↺ DR

▼ Create temp table: WAU by profession4 мин 31 сек ✓ ▶ Запустить ⚙ 📄 🔍 ⋮ ✨

45 ORDER BY week_start DESC, wau DESC;

prod_v_emart.mdm_employee_structure prod_sdp_statist_analytics.eventLN 1, COL 1 CHARS: 1739

▶ Text ⬇

▼ Publish to DLH4 сек ✓ ▶ Запустить ⚙ 📄 🔍 ⋮ ✨

Type: python ▼

1 from ssetl import publish_to_dlh
2
3 # Публикация данных из временной таблицы в целевую (DLH SCD1)
4 publish_to_dlh(
5 source='tmp_helicopter_wau_profession',
6 target='prod_sse.helicopter_wau_profession'
7)

helicopter-python-lsp LN 1, COL 1 CHARS: 205

▶ Text ⬇

/usr/local/lib/python3.12/site-packages/dwh_etl_tools/common.py:11: UserWarning: pkg_resources is deprecated as an API. See https://setuptools.pypa.io/en/latest/pkg_resources.html. The pkg_resources package is slated for removal as early as 2025-11-30. Refrain from using this package or pin to Setuptools<81.
import pkg_resources
[2026-08-26 10:57:44] [INFO] Successfully loaded to ssetl API

Ревью ноутбука перед публикацией+ 🕒 ⚙ ✕ ✨

TABLE tmp_helicopter_wau_profession — без схемы (временная таблица). INSERT INTO tmp_helicopter_wau_profession — без схемы. DROP TABLE IF EXISTS tmp_helicopter_wau_profession — без схемы. Никаких schema.table DML/DDI → PASS

CHECK-006 · ForbiddenSource — параграф #30062713 (dlh_trino): источники prod_sdp_statist_analytics.event и prod_v_emart.mdm_employee_structure. Оба — продуктовые схемы, не входят в запрещённые (usr_wrk*, prod_wrk*, bi_v_*, test_v_*). → PASS

Все проверки пройдены. Сформирую итоговый отчёт.

📄 Notebook Review Summary

HelicopterWAU (1625298) sourceTable: tmp_helicopter_wau_profession Findings: 0 (🔴 Blockers: 0 ⚠ Warnings: 0) Verdict: ✓ SAFE TO PUBLISH

Таблица статусов:

<> Unknown 📄 ✕

Check	Name	Status	Det
CHECK-001	TinkoffPyGpService	✓ PASS	—
CHECK-002	TargetAsSelect	✓ PASS	—
CHECK-003	EmptyParagraphText	✓ PASS	—
CHECK-004	ForbiddenFunctions	✓ PASS	—
CHECK-005	DMLSecurityViolations	✓ PASS	—
CHECK-006	ForbiddenSource	✓ PASS	—

⚡ find_tables ⚡ text2sql

Введите ваше сообщение здесь, используйте '/' для команд
+ Agent ▼ 🔍 ⬆

Подготовка кода к публикации

ADLC через агента



Ноут

Разработка

Помощь

Dev

Prod

SSE_IDWH_D... / HelicopterW...

Опубликован

DR

Create temp table: WAU by profession4 мин 31 сек

```
42 SUM(wau) OVER (PARTITION BY week_start) as total_wau_week,
43 CAST(wau AS DOUBLE) / SUM(wau) OVER (PARTITION BY week_start) * 100 as pct_of_week
44 FROM wau_by_profession
45 ORDER BY week_start DESC, wau DESC;
```

prod_sdp_statist_analytics.eventprod_v_emart.mdm_employee_structureLN 1, COL 1CHARS

Text

Publish to DLH4 сек

Type: python

```
1 from ssetl import publish_to_dlh
2
3 # Публикация данных из временной таблицы в целевую (DLH SCD1)
4 publish_to_dlh(
5     source='tmp_helicopter_wau_profession',
6     target='prod_sse.helicopter_wau_profession'
7 )
```

helicopter-python-lspLN 5, COL 44CHARS

Text

```
/usr/local/lib/python3.12/site-packages/dwh_etl_tools/common.py:11: UserWarning: pkg_resources is deprecated as an API. See https://setuptools.pypa.io/en/latest/pkg_resources.html. The pkg_resources package is slated for removal as early as 2025-11-30. Refrain from using this package or pin to Setuptools<81.
```

Опубликовать

Опубликовать какDLH ETL

Конфигурация публикации

Таргет-таблица*prod_sse.helicopter_wau_profession

Крон0 15 * * MON,WEDВ 15:00, только в понедельник и среду, MSK

Уведомлять о незапуске50 23 * * *В 23:50, MSK

Ключи дистрибуции*week_start

Использовать инициализирующую загрузку

Инициализирующая таблица

Тип инициализирующей загрузки

ОтменитьОпубликовать

Настройки публикации

ADLC через агента



Ноут

Разработка

Помощь

Dev

Prod

SSE_IDWH_D... / HelicopterWAU

DR 🔍 📄 🔔 ▶ Запустить все ☁ Опубликован ↺

▼ Create temp table: WAU by profession 4 мин31 сек 🟢 ▶ Запустить

```
40 legal_position_nm,
41 wau,
42 SUM(wau) OVER (PART
43 CAST(wau AS DOUBLE)
44 FROM wau_by_profession
45 ORDER BY week_start DES
```

prod_sdp_statist_analytics.event

Text

▼ Publish to DLH

Type: python ▼

```
1 from ssetl import publi
2
3 # Публикация данных из
4 publish_to_dlh(
5     source='tmp_helicopter
6     target='prod_sse.he
7 )
```

Text

> /usr/local/lib/python3.12/site-packages/dwh_etl_tools/common.py:11: UserWarning: pkg_resources is deprecated as an API. See https://setuptools.pypa.io/en/latest/pkg_resources.html. The pkg_resources

Лог ревью

2026-08-26 10:58 system
Test run completed successfully. Waiting for the review phase to complete.

Копировать

[2026-08-26 10:58:20] Review started with pipeline dlhReview
[2026-08-26 10:58:20] TargetCheck: Passed
[2026-08-26 10:58:20] NonMatchPublishTargetNameCheck: Passed
[2026-08-26 10:58:20] InvalidTableNameDlhCheck: Passed
[2026-08-26 10:58:20] NoteExecutionTimeCheck: Passed
[2026-08-26 10:58:20] ParagraphsForbiddenFunctionsCheck: Passed
[2026-08-26 10:58:20] ParagraphsAllowedShebangsDlhCheck: Passed
[2026-08-26 10:58:20] ParagraphsSameTenantCheck: Passed
[2026-08-26 10:58:20] EmptyParagraphTextCheck: Passed
[2026-08-26 10:58:20] DMLSecurityViolationsDlhCheck: Passed
[2026-08-26 10:58:20] SparkDropTablePurgeCheck: Passed
[2026-08-26 10:58:20] WikiExistenceCheck: Passed
[2026-08-26 10:58:21] WikiDdlConsistencyDlhCheck: Passed
[2026-08-26 10:58:21] WikiColumnsTypeCodesCheck: Passed
[2026-08-26 10:58:21] SensLabelHubCheck: Passed
[2026-08-26 10:58:21] ForbiddenSourceDLHCheck: Passed
[2026-08-26 10:58:21] MetahubDistributionKeysCheck: Passed
[2026-08-26 10:58:21] ProcessedDttmColumnExistenceDlhCheck: Passed
[2026-08-26 10:58:21] ...

Закреть

Решение задачи DWSAI-1569

траф

ица не существует, я могу создать DQ-ой таблицы tmp_wau_helicopter (P). Давайте проверю её метаданные.

ег — временная таблица, созданная в может быть недоступна из ноутбука

тупна ли tmp_wau_helicopter из а мне нужно понять, где она создана.

е я создам DQ-проверки для витрины ру, которую я знаю из расчёта: fession_nm (varchar), wau (bigint).

ица ещё не существует, я подготовлю DQ-т проверить:

start, profession_nm) — составной ключ ofession_nm, wau льный)

существующая таблица. Давайте я tmp_wau_helicopter из ноутбука DQ.

ска метаданные tmp_wau_helicopter data ⚡ fix ⚡ explain ⚡ find_tables

Проверки и автотесты при деплое

ADLC через агента



Helicopter

Администрирование

Data Detective

DataWorkspace AI Team

Очереди

DR

Нессы Data Beta

Недавние

Избранное

Мои ноуты

Все ноуты

Коллекции

Скилы Beta

Переменные

Мониторинг

Архив

Корзина

ИЗБРАННЫЕ КОЛЛЕКЦИИ

Настройки

История

Новый чат

Поиск

Сегодня

Решение задачи только что

Решение задачи DWSAI-1569

2	Missing	missing_count(week_start) = 0	NULL в ключевом поле
3	Missing	missing_count(legal_position_nm) = 0	NULL в ключевом поле
4	Missing	missing_count(wau) = 0	NULL в метрике
5	Numeric	min(wau) ≥ 0	WAU не может быть отрицательным
6	Validity	invalid_count(pct_of_week) = 0 (valid 0-100)	Доля недели в диапазоне 0-100%
7	Numeric	row_count(week_start) = 0 (filter: week_start > current_date)	Неделя не в будущем
8	Numeric	row_count > 0	Витрина не пустая

Создать эти проверки?

Создать эти DQ-проверки для prod_sse.helicopter_wau_profession?

Создать все 8

Создать SodaCL параграф с 8 DQ-проверками для prod_sse.helicopter_wau_profession.

Только базовые

Создать только базовые проверки (уникальность, NULL, wau>=0).

Отмена

Не создавать проверки.

Введи свой ответ

1/1

<

Skip

>

Введите ваше сообщение здесь

Agent (beta)

Артефакты

Созданные

Расчёт WAU пользователей Ноутбук

SSETL: Расчёт WAU пользователей Ноутбук

DQ-проверки витрины Ноутбук

Использованные

DWSAI-1569

Задача для DEMO агента

prod_chrono_helicopter_wau_profession

Документация таблицы

prod_chrono_helicopter_wau_profession

Таблица отражает выполнение параграфов в витрине

prod_v_dds.mdm_employee

Информация о профессиях сотрудников

prod_v_dds_dic.employee

Версионный справочник по профессиям сотрудников из AD

prod_v_dds.mdm_employee

Единый сотрудник группы компаний Тинькофф. Источник: AD

prod_v_dds.mdm_employee

Учетные записи единого сотрудника из AD, Cisco ISE и др.

prod_v_emarket.mdm_employee

Витрина с атрибутами сотрудника как физ. лица и его родственников

prod_v_chrono_helicopter_wau_profession

Документация таблицы

Разработка проверок качества данных

ADLC через агента



Ноут

Разработка

Помощь

Dev

Prod

DWS-AI-TEAM / DQ-проверки витрины WAU Helicopter (DWSAI-1569)

Запустить все

Опубликовать

DQ-проверки prod_sse.helicopter_wau_profession

7 сек

Запустить

```
5 ---
6 checks for prod_sse.helicopter_wau_profession:
7   # Уникальность составного ключа (неделя + профессия)
8   - duplicate_count(week_start, legal_position_nm) = 0:
9       name: sse_helicopter_wau_profession_primary_keys_uniqueness
10      tags: [autodq]
11   # Отсутствие NULL в ключевых полях
12   - missing_count(week_start) = 0:
13       name: sse_helicopter_wau_profession_week_start_not_null
14       tags: [autodq]
15   - missing_count(legal_position_nm) = 0:
16       name: sse_helicopter_wau_profession_legal_position_nm_not_null
17       tags: [autodq]
18   - missing_count(wau) = 0:
19       name: sse_helicopter_wau_profession_wau_not_null
20       tags: [autodq]
21   # WAU не может быть отрицательным
22   - min(wau) >= 0:
23       name: sse_helicopter_wau_profession_wau_non_negative
24       default value: 0
25       tags: [autodq]
26   # Доля недели в диапазоне 0-100%
27   - invalid_count(pct_of_week) = 0:
28       name: sse_helicopter_wau_profession_pct_of_week_valid_range
29       valid min: 0
30       valid max: 100
31       tags: [autodq]
32   # Неделя не может быть в будущем
33   - row_count(week_start) = 0:
34       name: sse_helicopter_wau_profession_week_start_not_in_future
35       filter: week_start > current_date
36       tags: [autodq]
```

Решение задачи DWSAI-1569

3. Запуск — все 8 проверок PASSED (datasource: DLH, debug: true).

Проверки:

Проверка	Результат
duplicate_count(week_start, legal_position_nm) = 0	0 дубликатов
missing_count(week_start) = 0	0 NULL
missing_count(legal_position_nm) = 0	0 NULL
missing_count(wau) = 0	0 NULL
min(wau) ≥ 0	min = 1
invalid_count(pct_of_week) = 0 (0-100)	0 вне диапазона
row_count(week_start) = 0 (filter: будущее)	0 строк в будущем
row_count > 0	4209 строк

Параграф помечен тегом autodq, имена проверок по маске sse_helicopter_wau_profession_*. Для production-запуска достаточно переключить debug: false в конфигурации параграфа.

find_tables text2sql

Введите ваше сообщение здесь, используйте '/' для команд

+ Agent



Автоматизация Data Quality: как мы изменили подход к нашим инструментам

Разработка проверок качества данных

ADLC через агента



☰

🔒

Editing

Original

Change requests / Контакт для таблицы prod_sse.helicopter_wau_profession

↺

Close

Request review

?

🔄

DR

Содержимое

▼ sse.helicopter_wau_profession

Service Requirements

• Quality Metrics

▼ Service Requirements

Criticality level ?

Business Operational

Update frequency ?

Build time: at 10:00, every day

Actuality boundary: No earlier than 1 day before

▼ Quality Metrics

🔍 Search by path or ID

+

🔄

1

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_primary_keys_uniqueness

Added

🗑

2

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_wau_not_null

Added

🗑

3

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_no_missing_weeks

Added

🗑

4

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_week_start_not_null

Added

🗑

5

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_pct_of_week_not_null

Added

🗑

6

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_row_count_positive

Added

🗑

7

sandbox.sse_helicopter_wau_profession.sse_helicopter_wau_profession_total_wau_week_not_null

Added

🗑

Details

Контакт для таблицы prod_sse.helicopter_wau_profession

Первое заведение контракта

Руднев Дмитрий September 4, 2026

Reviewers

Виктория

Михаил

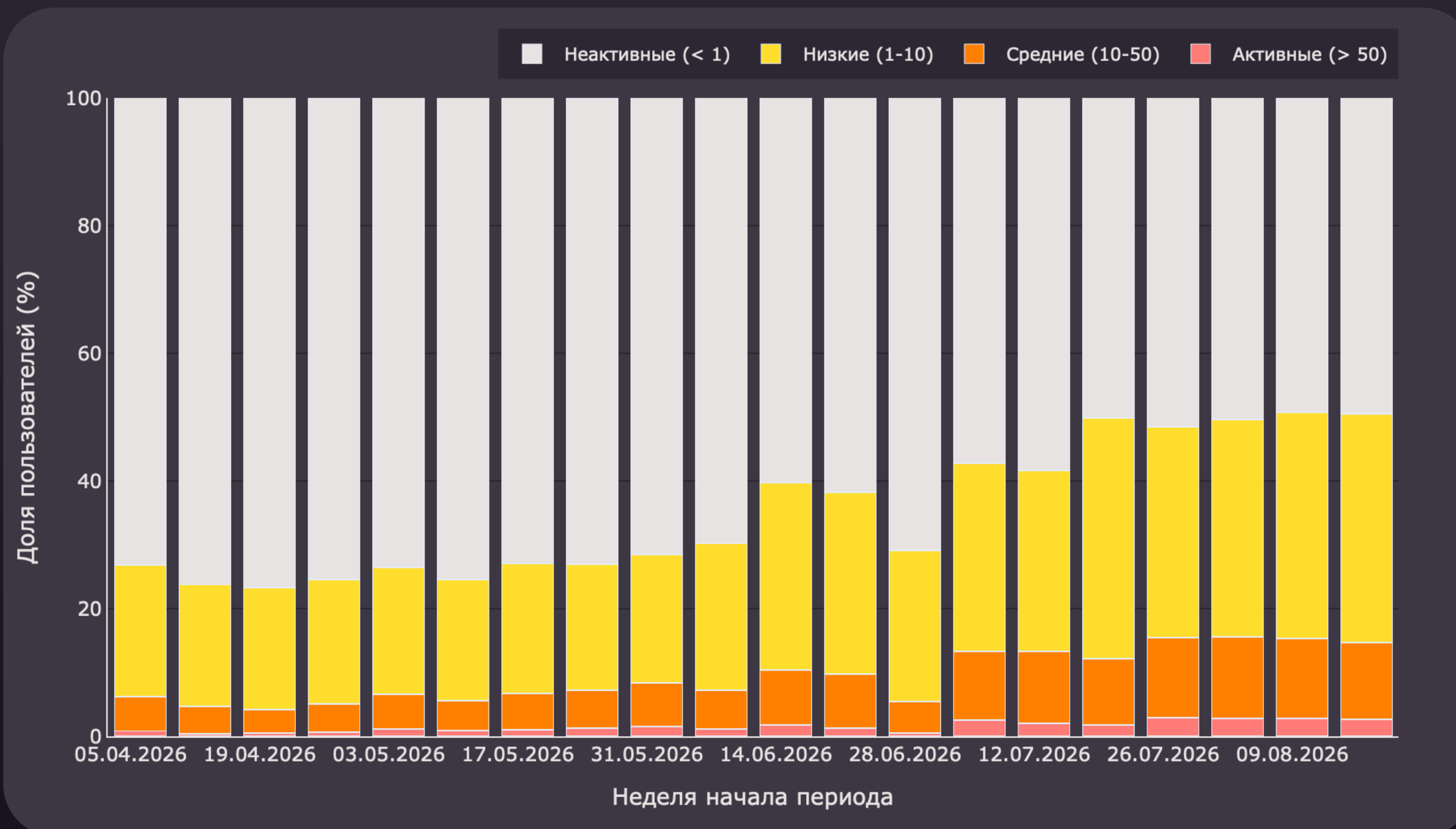
Objects

sse.helicopter_wau_profession

Разработка контракта данных (пока без AI)

**Метрики.
А пользуются ли AI?**

Пользователи AI-фичей в Data Platform



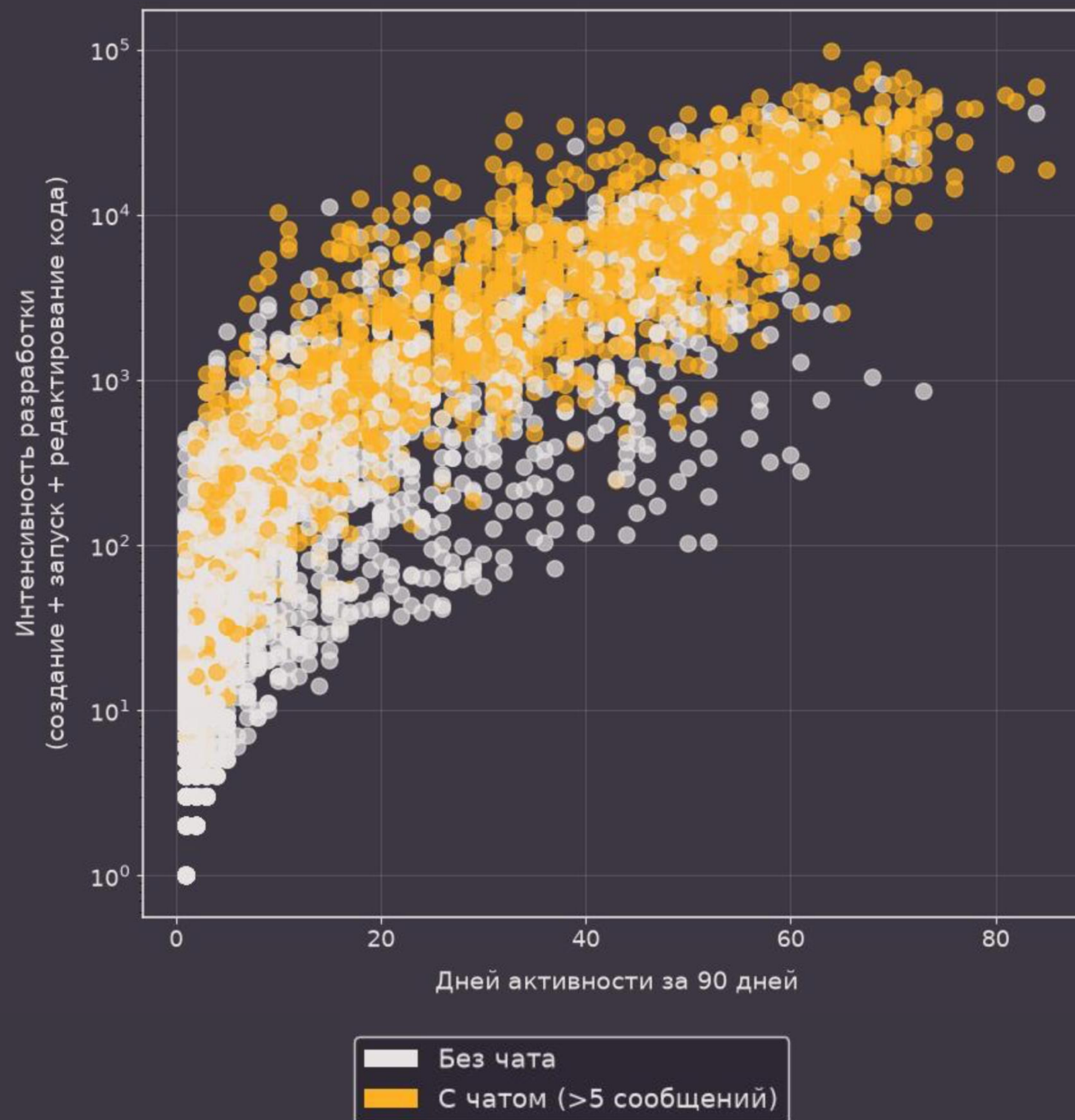
~77%

пользователей продукта
используют AI-фичи

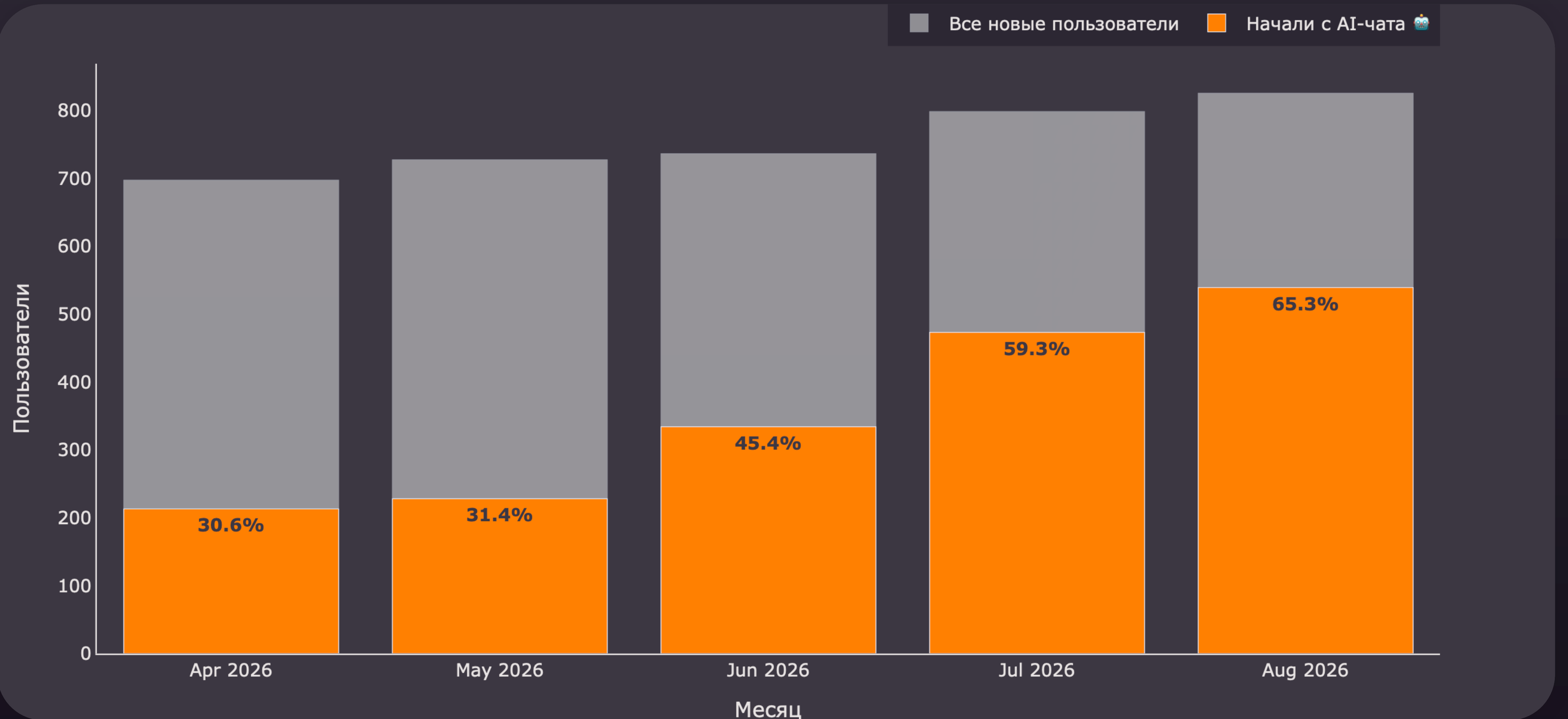
~51%

пользователей
используют AI-агента

Активность в продукте связана с использованием AI



Новые пользователи часто начинают с AI

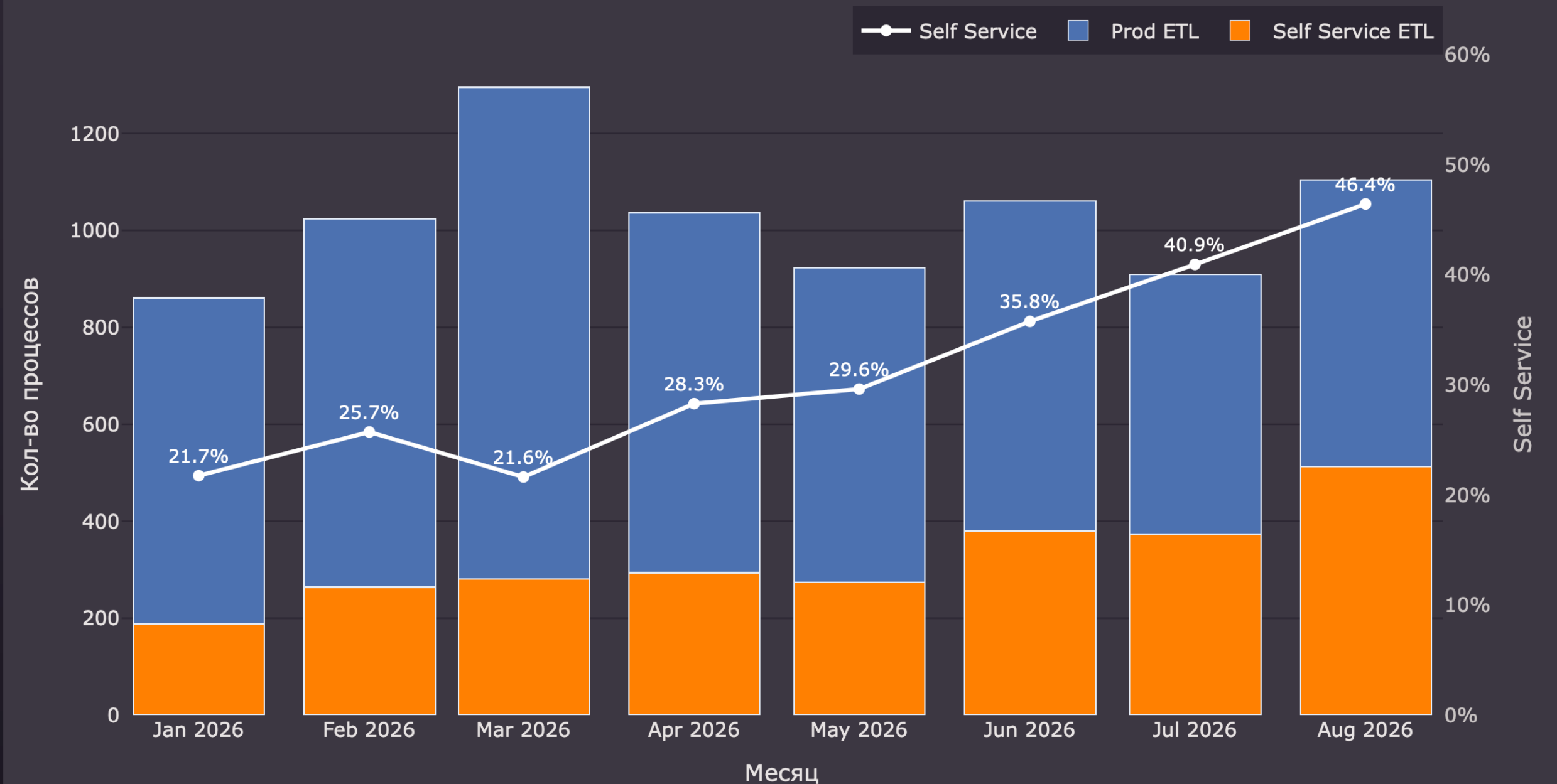


Переезд в self-service

46% новых ETL в месяц — в self-service инструменте

≈35% всех существующих ETL процессов в self-service инструменте

Тренд начался до AI.
AI его ускоряет



**Метрики.
А есть ли
эффект от AI?**

Измерение продуктивности команд

DORA — Скорость +
стабильность доставки

- Lead Time → часы / дни
- Deployment Frequency → deploy / неделю
- Change Failure Rate → %
- Recovery Time → часы

SPACE — Результат +
процесс + люди

- Performance → достигнутый результат
- Activity → PR / задачи / commits
- Efficiency → время / ожидание
- Collaboration → взаимодействия
- Satisfaction → score

Измерение продуктивности команд в Data

ETL / регулярный процесс

- Целевое действие → релиз
- Единица → команда
- Критерий падения → падение пайплайна/DQ
- В неделю ~300 релизов
- WAU: ~700

Ad-hoc

- Целевое действие → код/инсайт
- Единица → сотрудник/команда
- Критерий падения → ?
- В неделю ~9 000 новых ноутбуков
- WAU: 6 780

Ставка на throughput

Throughput — количество
ad-hoc/ETL-процессов
на пользователя за единицу
времени

Ad-hoc +11%

Q1: 1.29 → Q3: 1.44

ETL +26%

Q1: 1,52 → Q3: 1,92



AI ускоряет работу. Но не всегда доставку результата



Faros AI — AI Engineering Report 2026

Задач на разработчика: +34%

Деплоев в неделю: -12%

Инцидентов на PR: x3,4

Телеметрия 22 тыс. разработчиков

MIT / NBER — Writing Code vs. Shipping Code

Коммиты: +180%

Проекты: +50%

Релизы: +30%

100 тыс.+ GitHub-разработчиков

McKinsey — The State of AI: Global Survey 2026

37% связывают AI с положительным вкладом в EBIT — без роста за год.

Только 6% — AI high performers: значимый эффект и вклад AI не менее 5% EBIT.

глобальный опрос 1 719 респондентов

J-Curve — мы ещё не дошли, чтобы увидеть эффект

J-Curve of AI value realization

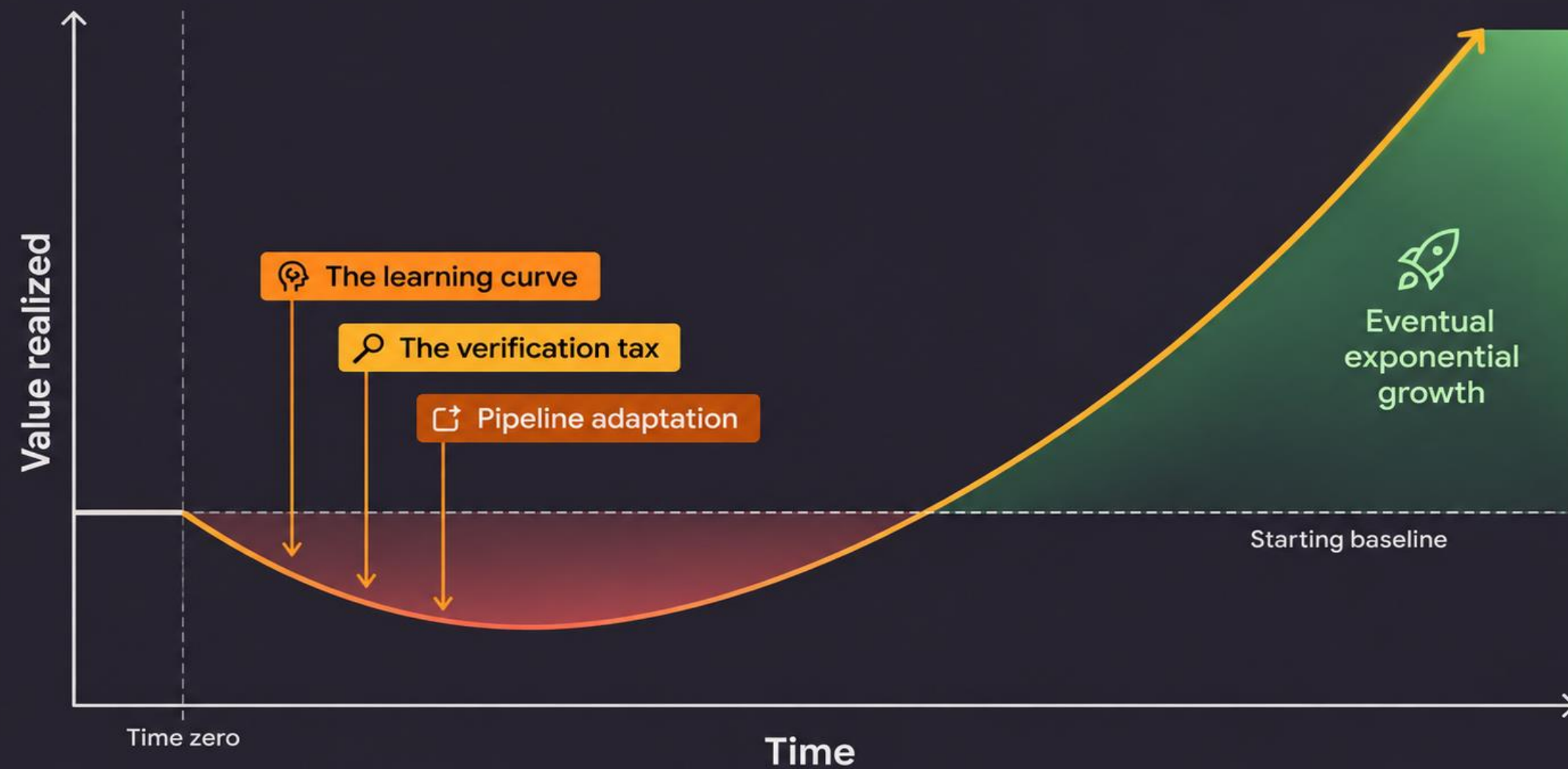


Figure 1: J-Curve of AI value realization

DORA ROI 2026

J-кривая, просадка 15% *
на 3 месяца — «плата за
обучение»

*параметр модели, у каждого цифры свои

Ускоряем создание быстрее, чем контроль

72%

приоритизируют AI-assisted coding

24%

работают над AI-assisted pipeline
management: тесты, observability, качество

66% → 83% за год

Доверие к данным как приоритет

71%

боятся, что неверные данные
доедут до стейкхолдера

Долг понимания

AI сделал за нас

- Снизил порог входа
- Повысил скорость написания кода

AI сделал обязательным, но не принёс

- Описание данных
- Data Contracts & Data Quality
- Платформизация
- Встроенные ограничения

Разрыв между колонками — это и есть долг понимания

Выводы

- ➔ AI не приносит новых свойств. AI усиливает и сильные, и слабые стороны систем одинаково
- ➔ Важно проектировать цикл аналитической разработки целиком, а не отдельными этапами
- ➔ Стандарт задаёт платформа: собранный контекст, границы, проверки
- ➔ Владение = стандарт от платформы + решение от человека

**Скорость дадут агенты.
Задача платформы —
сохранить доверие к данным
на этой скорости.**



Спасибо!

Дмитрий Руднев | Telegram: @dorudnev

Источники и материалы

Официальные страницы, открытые версии и материалы доклада.

Вступление Дмитрий Руднев. «Инструменты Data Quality», SmartData 2024

Страница: <https://smartdataconf.ru/archive/2024/persons/0ed8f7a8898148bdb30a85ab3bd8e41a/>

Видео: <https://youtu.be/cWHW4RgBibw>

Слайд 5 Tristan Handy, dbt Labs. The Analytics Development Lifecycle, 2024

Статья: <https://www.getdbt.com/resources/the-analytics-development-lifecycle>

PDF: <https://8698602.fs1.hubspotusercontent-na1.net/hubfs/8698602/The%20Analytics%20Development%20Lifecycle.pdf>

Слайд 8 DORA / Google Cloud. State of AI-assisted Software Development 2025

Статья: <https://dora.dev/research/2025/dora-report/>

Отчёт: <https://cloud.google.com/resources/content/2025-dora-ai-assisted-software-development-report>

Слайд 10 Gartner. Hype Cycle for Agentic AI, 2026

Статья: <https://www.gartner.com/en/articles/hype-cycle-for-agentic-ai>

Отчёт: <https://www.gartner.com/en/documents/7671861>

Слайд 10 Gartner. Magic Quadrant for Enterprise Low-Code Application Platforms, 2025

Отчёт: <https://www.gartner.com/en/documents/6773234>

Слайды 11, 44 dbt Labs. 2026 State of Analytics Engineering

Статья и отчёт: <https://www.getdbt.com/resources/state-of-analytics-engineering-2026>

Слайд 12 Gartner. Hype Cycle for Data Management, 2026

Отчёт: <https://www.gartner.com/en/documents/8152229>

Слайд 12 Gartner. Magic Quadrant for Augmented Data Quality Solutions, 2026

Отчёт: <https://www.gartner.com/en/documents/7429962>

Слайд 13 Gartner. Top Data & Analytics Predictions, 2025

Статья: <https://www.gartner.com/en/newsroom/press-releases/2025-06-17-gartner-announces-top-data-and-analytics-predictions>

Контекст: <https://www.gartner.com/en/documents/8152229>

Слайд 14 Gartner. By 2028, 80% of GenAI Business Apps Will Be Developed on Existing Data Management Platforms, 2025

Статья: <https://www.gartner.com/en/newsroom/press-releases/2025-06-02-gartner-predicts-by-2028-80-percent-of-genai-business-apps-will-be-developed-on-existing-data-management-platforms>

Слайд 39 DORA / Google Cloud. Software Delivery Performance Metrics

Статья: <https://dora.dev/guides/dora-metrics/>

Слайд 39 Forsgren et al. The SPACE of Developer Productivity, 2021

Microsoft Research: <https://www.microsoft.com/en-us/research/publication/the-space-of-developer-productivity-theres-more-to-it-than-you-think/>

ACM: <https://doi.org/10.1145/3454122.3454124>

Слайд 42 Demirer, Musolff, Yang. Writing Code vs. Shipping Code, NBER Working Paper 35275.

Редакция от 28.05.2026

Страница: <https://www.nber.org/papers/w35275>

PDF редакции: https://www.nber.org/system/files/working_papers/w35275/revisions/w35275.rev0.pdf

Слайд 42 Faros AI. The Acceleration Whiplash: AI Engineering Report 2026

Статья и отчёт: <https://www.faros.ai/research/ai-acceleration-whiplash>

Слайд 42 McKinsey. The State of AI: Global Survey 2026

Статья: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

Слайд 43 DORA / Google Cloud. ROI of AI-assisted Software Development, 2026

Статья: <https://dora.dev/ai/roi/report/>

Отчёт: <https://cloud.google.com/resources/content/dora-roi-of-ai-assisted-software-development>

Слайд 45 Addy Osmani. Comprehension Debt: The Hidden Cost of AI Generated Code, 2026

Статья: <https://addyosmani.com/blog/comprehension-debt/>

Слайд 45 GitClear / GitKraken. The Maintainability Gap: AI Code Quality in 2026

Статья и отчёт: https://gitkraken.gitclear.com/the_ai_code_quality_maintainability_gap

Доп. материал Т-Банк. «Краткий обзор платформы данных Т-Банка»

Статья: <https://habr.com/ru/companies/tbank/articles/926886/>