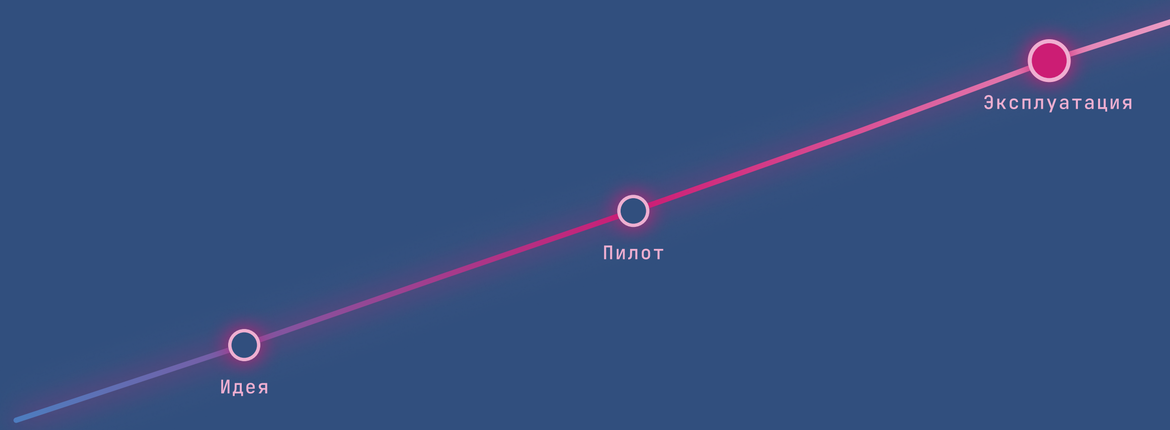
CNEWS · Cloud4Y

Инфраструктура для ИИ

Как облако ускоряет запуск ИИ-проектов в компаниях



Клиентова Зоя
Директор по маркетингу · Cloud4Y

О Cloud4Y

Корпоративный облачный провайдер с собственными ЦОД — и публичным GPU-облаком

2009

облачный провайдер с 2009 года

7 × TIER III

Дата-центров в России и за рубежом

2 000+

корпоративных клиентов

GPU Cloud

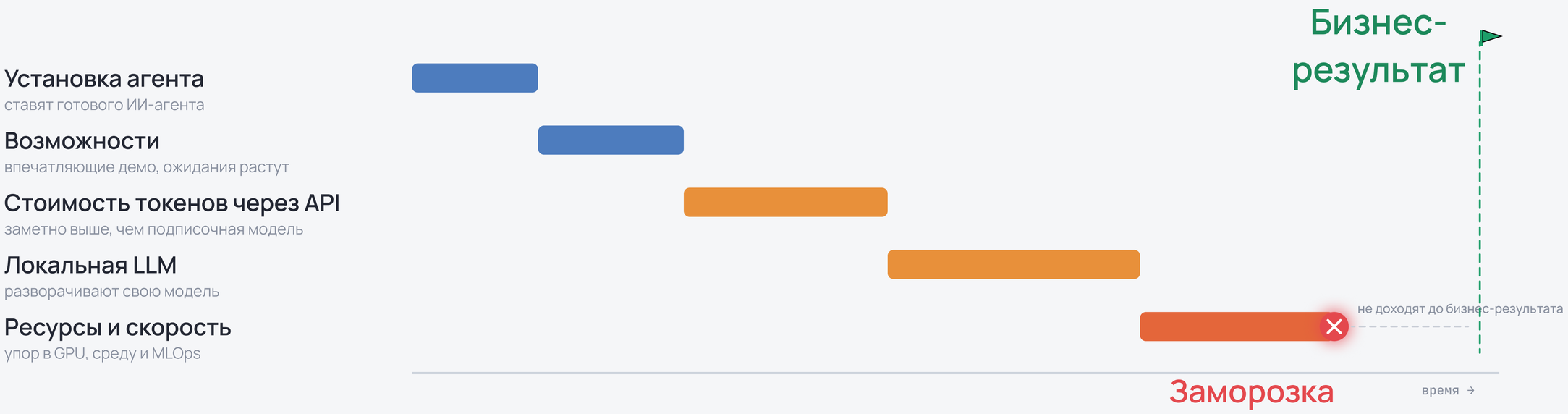
RTX 5090 · RTX 6000 · B200 · B300

- ISO 27001
- PCI DSS
- 152-ФЗ
- TIER III

EST. 2009 CLOUD4Y

Стандартный путь ИИ-проекта в B2B

Команда идёт своими силами — и упирается не в идею, а в инфраструктуру

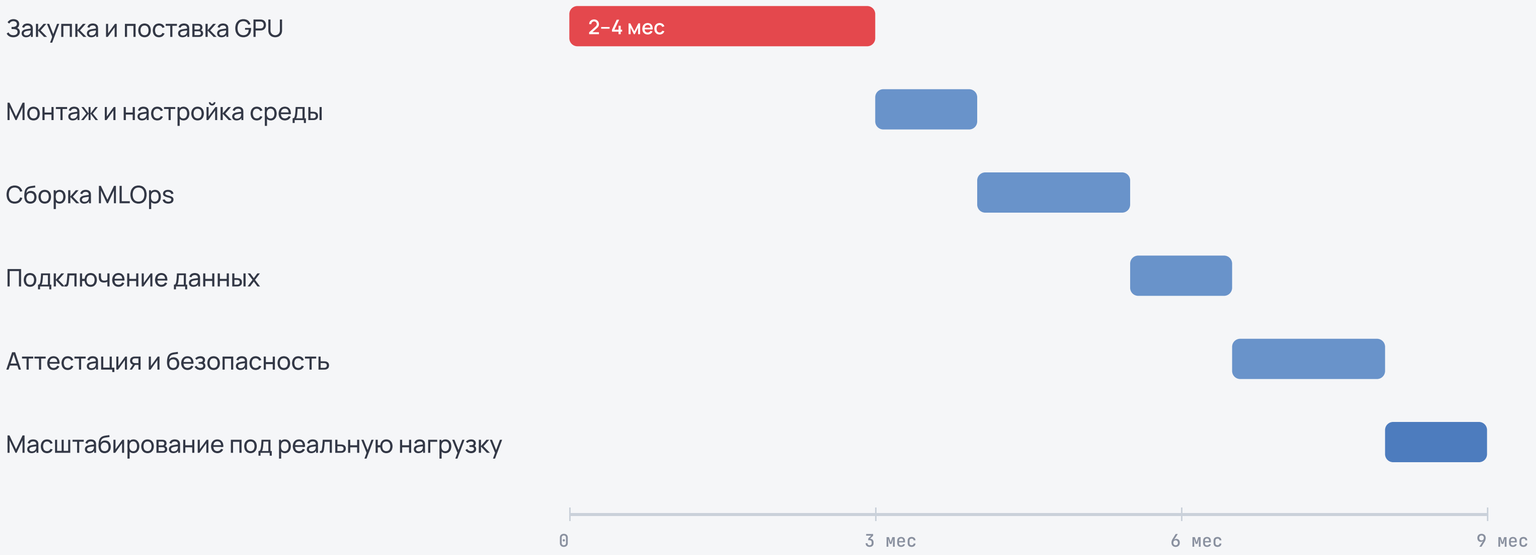


Проект останавливается до бизнес-результата — на среде и скорости работы

EST. 2009 CLOUD4Y

Сроки запуска на собственном оборудовании

На путь от идеи до использования в бизнес-процессах уходят месяцы



≈ 6–9

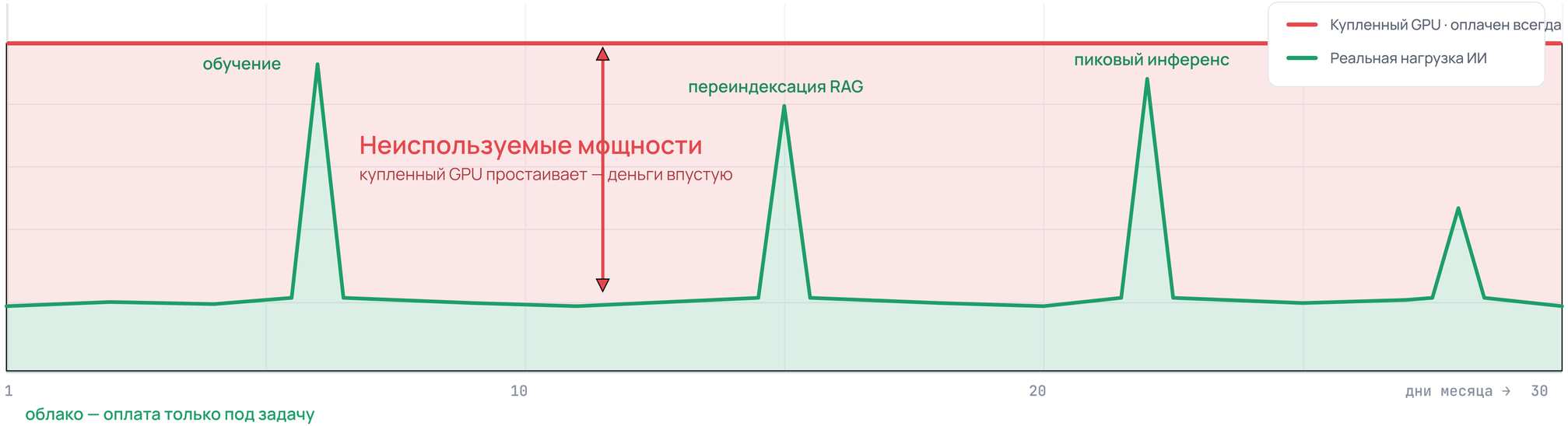
месяцев

и это срок до первого бизнес-результата

EST. 2009 CLOUD4Y

Профиль нагрузок ИИ систем

Нагрузка ИИ — неравномерная. Между пиками купленный GPU простаивает



Особенно на пилотах и тестах: нагрузка низкая и непредсказуемая — закупать GPU под неё невыгодно. Облако — оплата по факту, без простоя.

EST. 2009 CLOUD4Y

Какие барьеры устраняет облако

Облако убирает четыре главных барьера запуска ИИ-проекта.

GPU и мощности

Было

Закупка и инфраструктура для GPU

↓

Стало

GPU за минуты — нужный класс под задачу

Среда и MLOps

Было

Месяцы на среду и MLOps

↓

Стало

Готовая платформа: оркестрация, реестр, мониторинг

Данные

Было

Данные и интеграции вручную

↓

Стало

Управляемые БД и хранилища, S3-совместимость, сети между ЦОД

Безопасность

Было

Долгая аттестация

↓

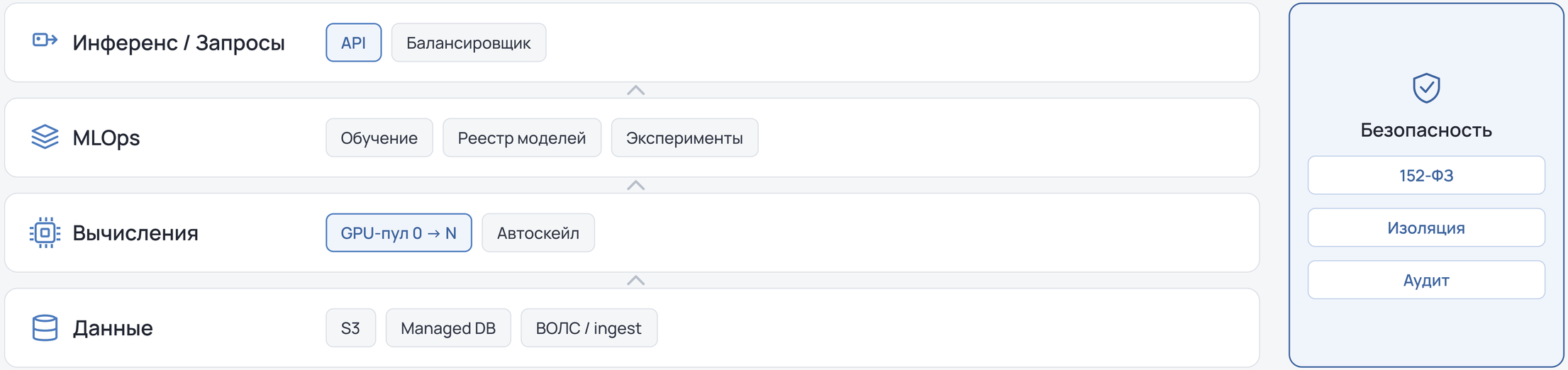
Стало

Уже готовая сертификация: 152-ФЗ, ФСТЭК, ФСБ

EST. 2009 CLOUD4Y

Архитектура ИИ-платформы в облаке

Готовые слои — от данных до serving, с безопасностью поверх всего.



EST. 2009 CLOUD4Y

Квартал против недели

Один и тот же путь до первого результата — но в облаке он короче в несколько раз

Собственное оборудование

≈ 6–9 мес

Закупка GPU доминирует в сроке · далее настройка среды, MLOps, аттестация

× 8–10 быстрее до первого результата

Облако

≈ 2–4 нед

GPU за минуты · готовая платформа, managed-данные, преднастроенный комплаенс

Скорость запуска определяет, кто первым выведем ИИ-функционал на рынок

иллюстративно,
зависит от проекта

EST. 2009
CLOUD4Y

Пример: корпоративный ИИ-ассистент (RAG)

Клиент
B2B-компания: внутренняя база знаний и документы

Задача
ИИ-ассистент с поиском по своим данным (RAG)

Было
Пилот буксовал: нет GPU, среда и данные — вручную

Стало
Cloud4Y: GPU-пул под индексацию + готовая платформа

СРОК ДО ПЕРВОГО ПИЛОТА

В своём контуре

≈ квартал

В облаке Cloud4Y

3 недели

× 4–5 быстрее

3 недели до первого рабочего ассистента

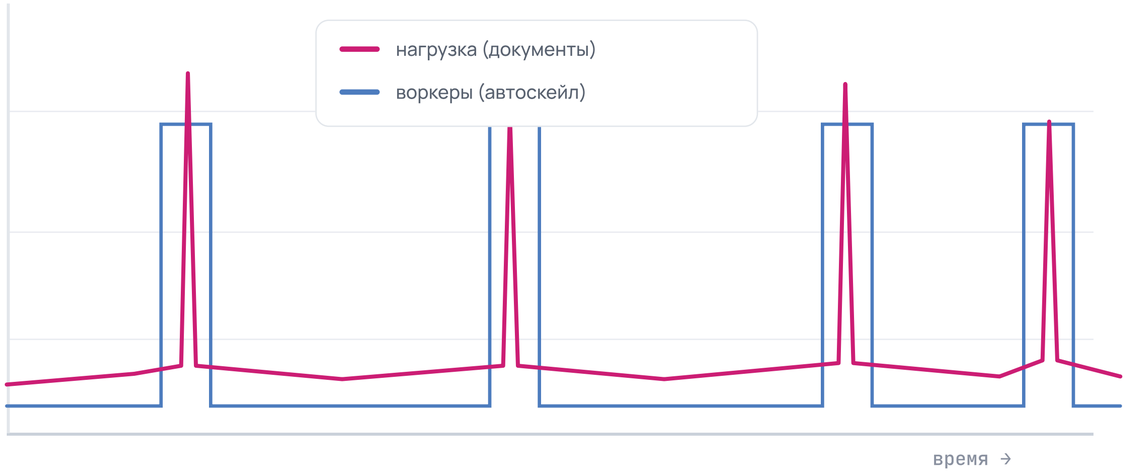
Без закупок железа на старте

Быстрее итерации поиска и промптов

EST. 2009 CLOUD4Y

Из пилота в эксплуатацию без переделки

B2B · пакетная обработка документов под пиковую нагрузку · автоскейл воркеров + Cloud4Y API.



×5 пиковая нагрузка к базовой — поглощается автоскейлом

2–4 нед от пилота в эксплуатацию без переделки архитектуры

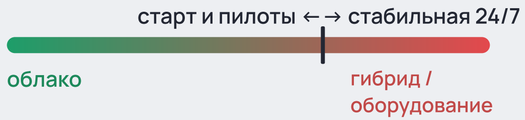
Без закупок на старте

EST. 2009 CLOUD4Y

Сравнение и риски запуска: облако или оборудование

Критерий	Собственное оборудование	Облако
Скорость запуска	месяцы	недели
Стартовые вложения	высокий capex на GPU	без capex, оплата по факту
Гибкость по классу GPU	ограничен закупленным оборудованием	любой класс под задачу
Аттестация по ИБ	аттестация с нуля	преднастроенный контур
Риск нереализации	высокий — долго до результата	низкий — быстрый пилот

Граница применимости: облако — старт, пилоты, переменная нагрузка.
Стабильная массовая нагрузка 24/7 на длинном горизонте — гибрид или свое оборудование



EST. 2009 CLOUD4Y

Собственные ЦОДы для ИИ

Суверенные данные и быстрый ввод под обучение: два TIER IV ЦОД + контейнерные, площадки в РФ и Стамбуле для гео.

DC4Y.1 S33 · Марфино

Готовность в 2026

TIER IV | 2 000 стойко-мест | 30 МВт | 15 кВт/стойка

24 мин от МКАД · полная автономность (сеть, газогенерация, дизель, своя вода) · антидрон-защита

DC4Y.2 V40 · Мытищи

TIER IV | 2 000 стойко-мест | 30 МВт | 15 кВт/стойка

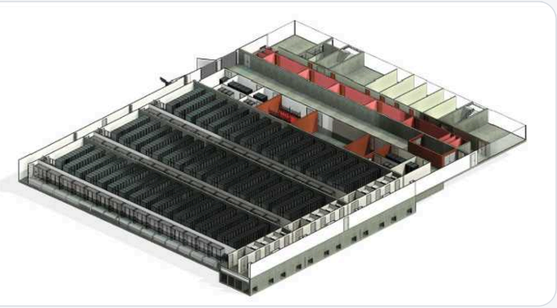
питание напрямую от ТЭЦ-27 · резерв ИБП + дизель · собственные ВОЛС до МКАД

Контейнерные ЦОД (КЦОД) · TIER III

мобильные · 10 стоек до 12 кВт · охлаждение 2N · герметичные коридоры · разворот на площадке

Под ИИ-нагрузки

- **Плотность 15 кВт/стойка** — под плотные GPU-узлы и инференс
- **Энергоэффективное охлаждение** — снимает тепло GPU, низкий PUE
- **TIER IV · резерв N+1…2N** — многодневное обучение без простоев
- **Антидрон + периметры охраны** — защита GPU-железа и данных
- **Собственные ВОЛС между ЦОД** — распределённое обучение, быстрый ввод данных
- **Площадки в РФ · 152-ФЗ** — суверенные данные для ИИ



EST. 2009 CLOUD4Y

Безопасность

Данные для обучения и RAG не покидают аттестованный контур и российскую юрисдикцию.

152-ФЗ

УЗ-1 для персональных данных

ФСТЭК

К1 + аттестованный сегмент

PCI DSS

Для платёжных систем

ISO 27001

Система менеджмента ИБ

ISO 9001

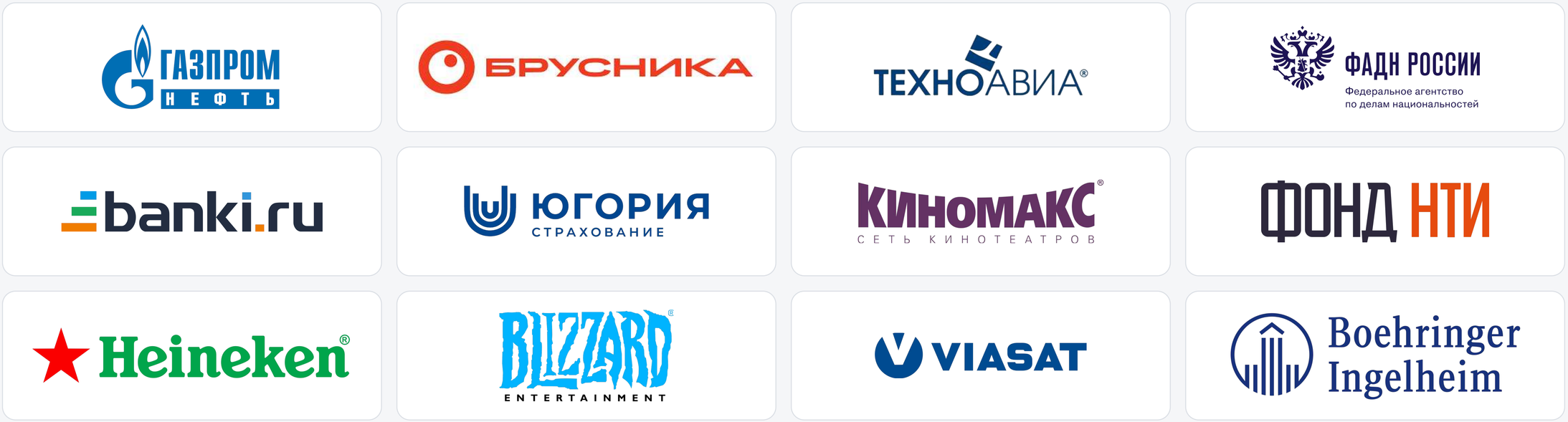
Система менеджмента качества

187-ФЗ

Работа с КИИ

EST. 2009 CLOUD4Y

С нами работают страховые компании, ритейл, промышленность, госсектор



2 000+ корпоративных клиентов · 17 лет на рынке

EST. 2009 CLOUD4Y

ВЫВОД · CLOUD4Y

Запустить ИИ-проект — за недели, а не за кварталы

- Быстрый старт без CAPEX
- Готовый комплаенс под данные ИИ
- Масштабирование пилот → эксплуатация без перепроектирования архитектуры

ДЛЯ УЧАСТНИКОВ CNEWS

Экспресс-аудит готовности к запуску ИИ

+ план инфраструктуры под ваш кейс

cloud4y.ru

Telegram: @cloud4y

+7 (495) 268-04-12

sales@cloud4y.ru

PROD