

Как ускорить внедрение ML/AI в энергетике

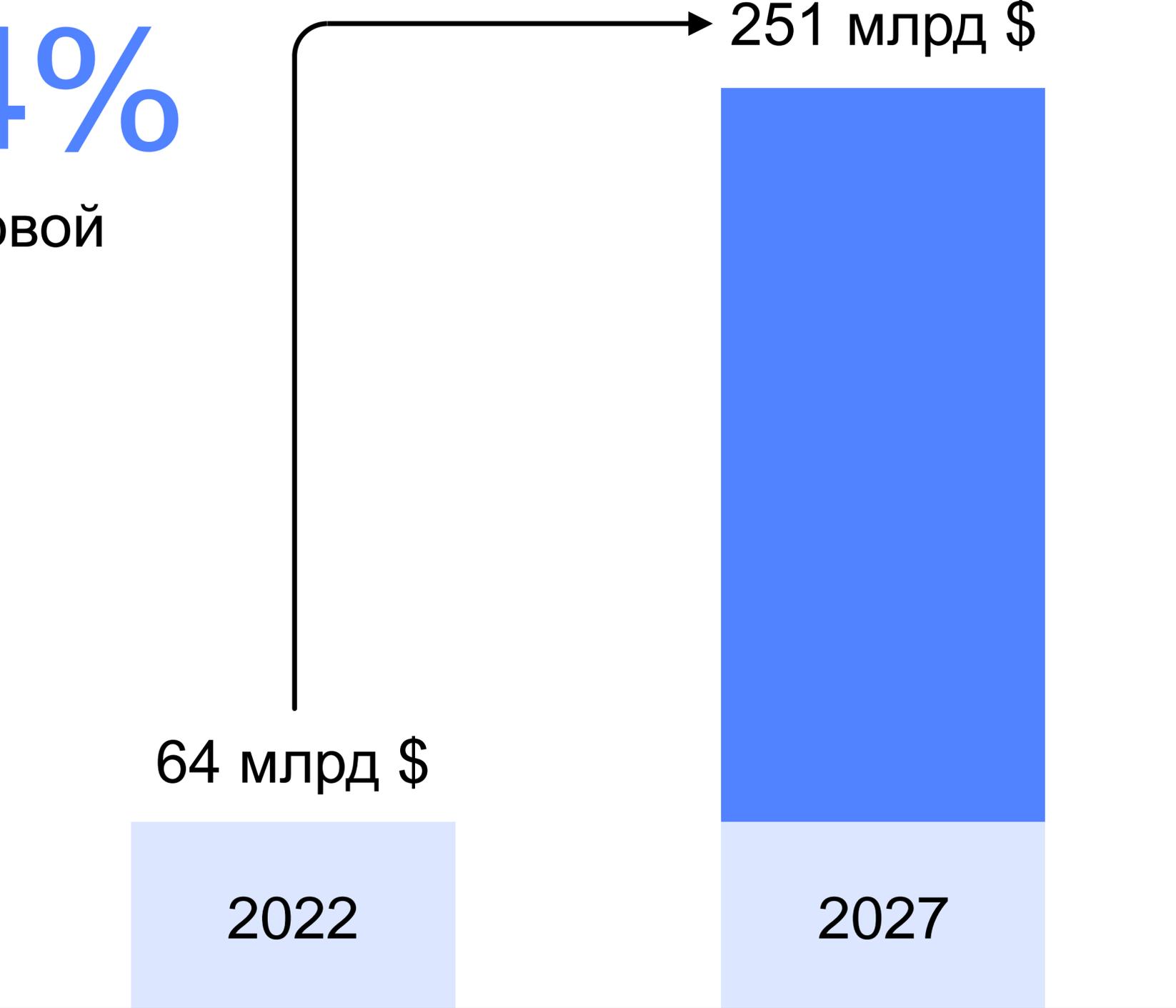
Роман Матусевич,
руководитель индустрии «Нефть,
газ, энергетика» Yandex Cloud

1. ML/ИИ: бурный рост
2. Вызовы и слагаемые успеха проектов AI
3. Технологическое партнерство = ускорение проектов ML/ИИ

Размер рынка ML & AI

31,4%

среднегодовой
темпа роста



Источник: Worldwide Artificial Intelligence Software Forecast, 2023–2027, IDC, сен 2023

ML

Анализ данных

Закономерности

Прогнозы



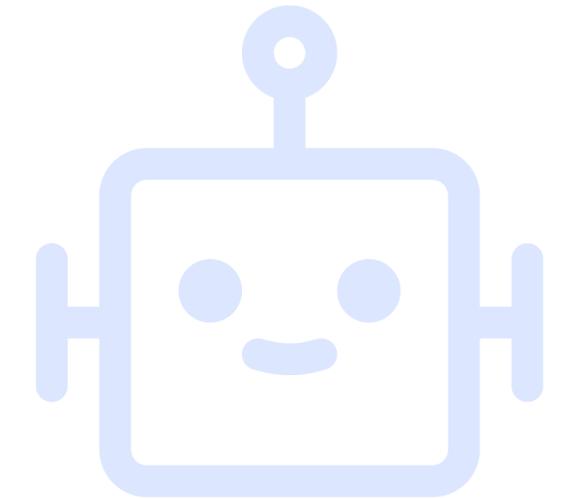
GenAI

Создание контента



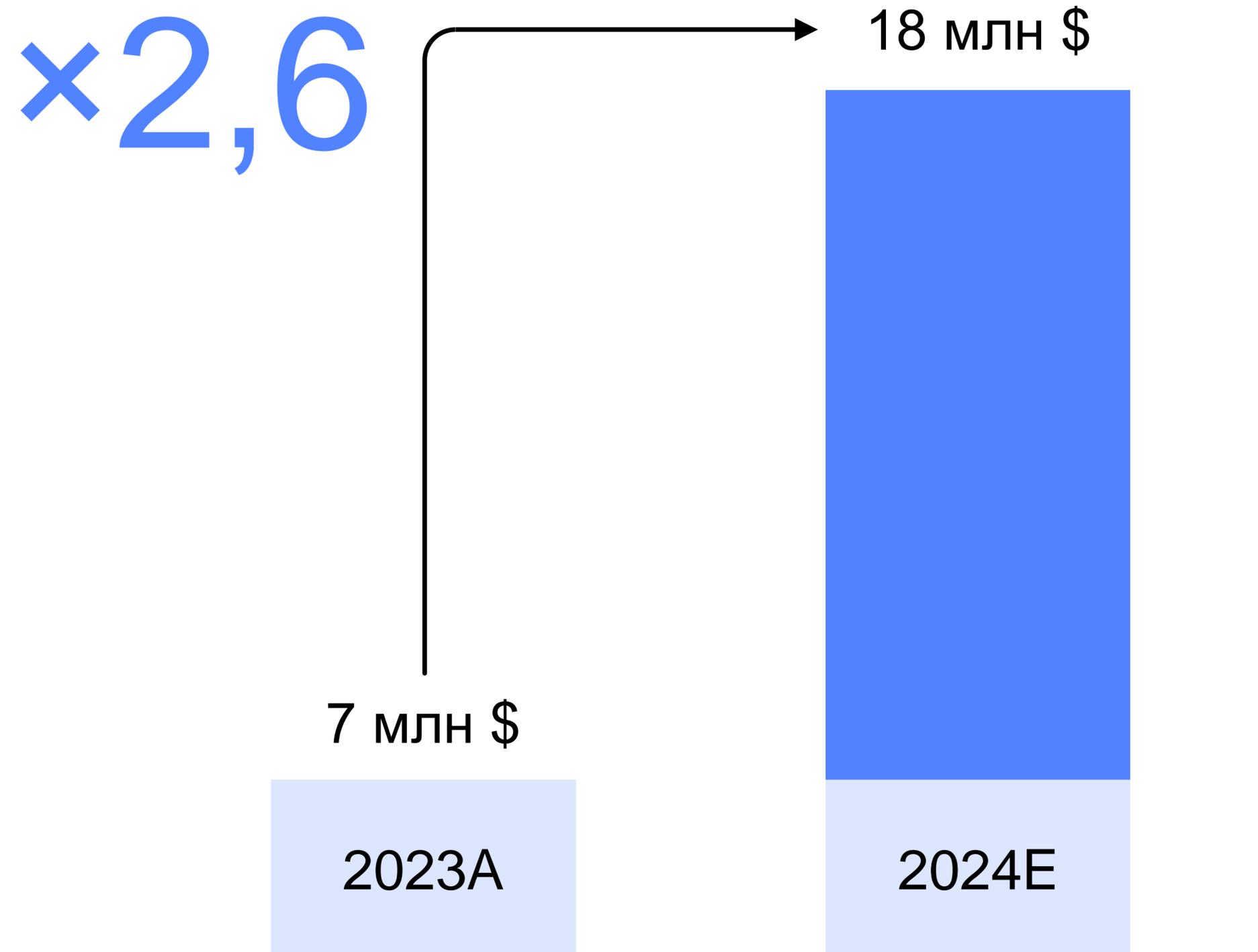
Генеративный AI

Generative AI, GenAI



Растут бюджеты на LLM

Генеративный AI



■ Средние расходы на компанию

10 000 +
КЛИЕНТОВ

ИЗ НИХ

1 500
ежедневно

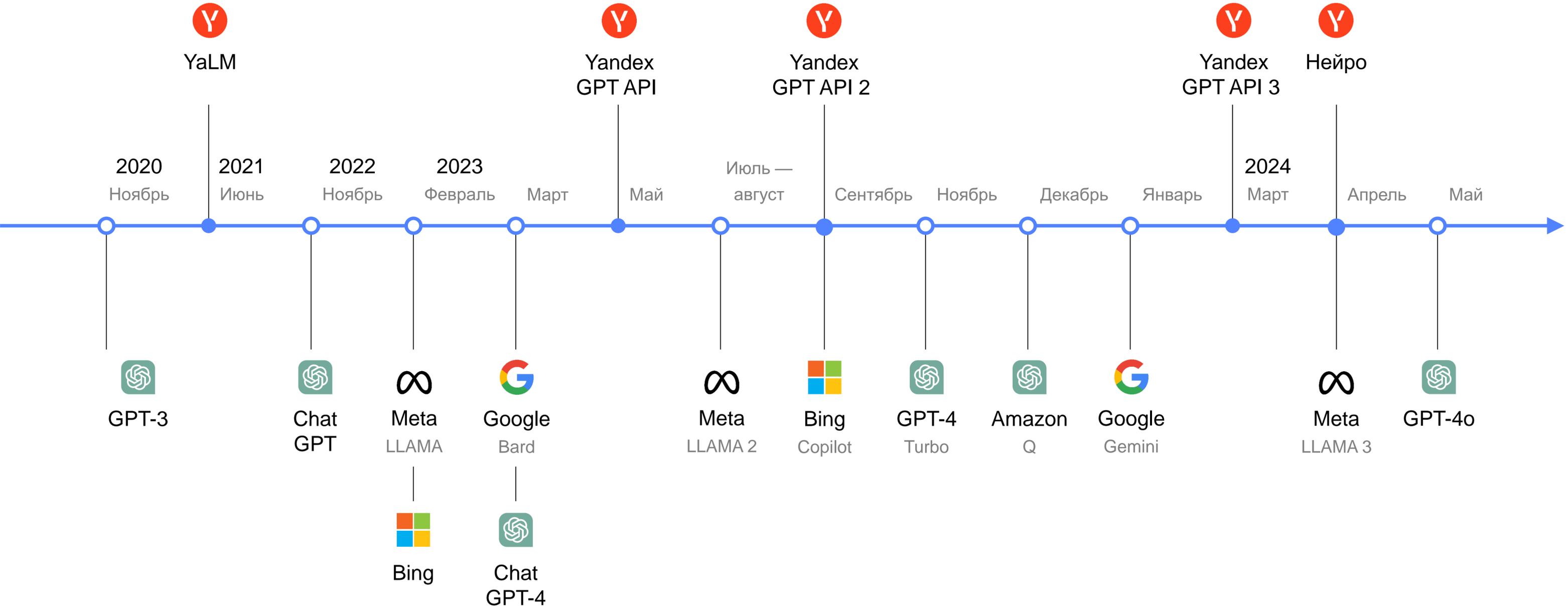
YandexGPT API

31% → Крупный
и средний бизнес

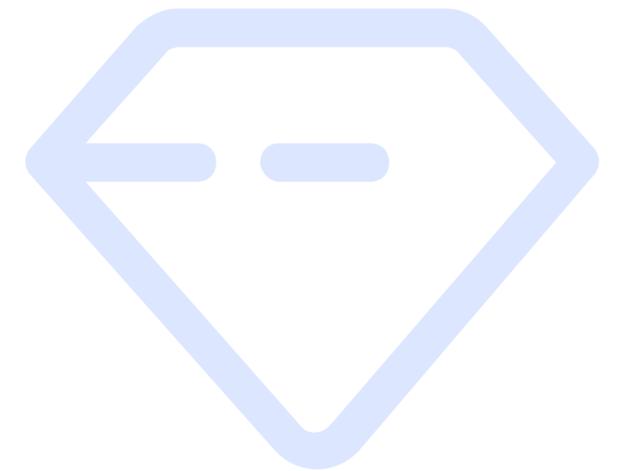
 69% Малый
бизнес

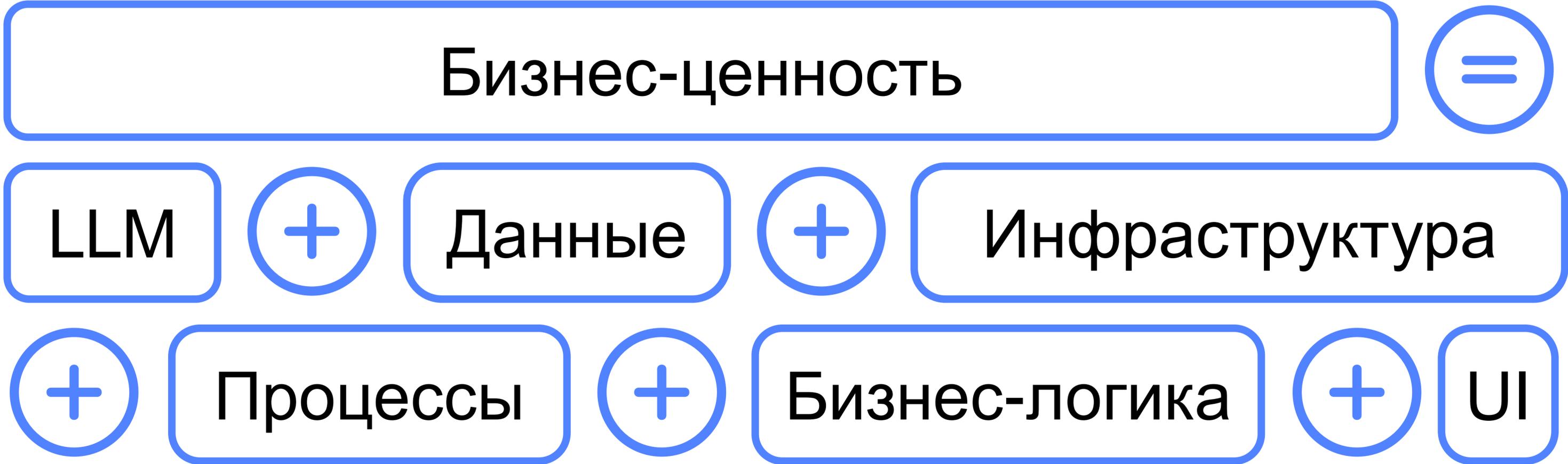
Улучшаются базовые модели

Генеративный AI



Бизнес- ценность





ВЫЗОВЫ

Сложно

Разработка
ML-модели —
это процесс



Дорого

Вычислительные
ресурсы и их
эффективная
утилизация стоят
много



Недостаточно

Нехватка экспертов



Бизнес-логика

ML-модели

Данные

ML-инфраструктура

Технологическое партнерство с Яндексом

Бизнес-логика

ML-модели

Данные

ML-инфраструктура

Вычислительные ресурсы с GPU

Среда разработки и обучения ML-моделей DataSphere

Технологическое партнерство с Яндексом

Бизнес-логика

ML-модели

Данные

Сервисы платформы данных
Apache Spark™, Apache Hadoop®

ML-инфраструктура

Вычислительные ресурсы с GPU
Среда разработки и обучения ML-моделей DataSphere

Технологическое партнерство с Яндексом

Бизнес-логика

ML-модели

YandexGPT API

Классификаторы, промты, эмбединги, RAG, Yandex Search AP
SpeechKit TTS/STT, Translate, Vision OCR, SpeechSense

Данные

Сервисы платформы данных

Apache Spark™, Apache Hadoop®

ML-инфраструктура

Вычислительные ресурсы с GPU

Среда разработки и обучения ML-моделей DataSphere

Технологическое партнерство с Яндексом

Бизнес-логика

Сервисы компаний-партнёров

Продуктовый консалтинг

Обучение

ML-модели

YandexGPT API

Классификаторы, промты, эмбединги, RAG, Yandex Search AP

SpeechKit TTS/STT, Translate, Vision OCR, SpeechSense

Данные

Сервисы платформы данных

Apache Spark™, Apache Hadoop®

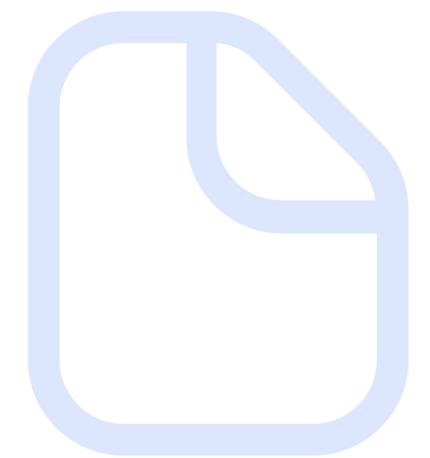
ML-инфраструктура

Вычислительные ресурсы с GPU

Среда разработки и обучения ML-моделей DataSphere

Данные — ключевая часть процесса адаптации и внедрения ML/GenAI

От них зависят качество и эффективность
конечной интеграции



Платформа данных Yandex Cloud

ClickHouse и YDB

Предоставляем большой набор управляемых сервисов на базе open-source-решений, включая разработанные в Яндексе базы данных — ClickHouse и YDB

Проекты с открытым исходным кодом

Разработали и запустили в Open Source YDB, YDS, DataLens, YTsaurus и другие. Регулярно инвестируем ресурсы команды в существующие технологии: Greenplum, ClickHouse, PostgreSQL

Внутренние проекты

Сервисы Yandex Data Platform используются во многих продуктах Яндекса, включая Почту, Такси, Драйв, Диск и другие. Мы понимаем, чего хотят пользователи, и даём им это



YandexGPT API

Доступные модели для всех пользователей

Модели

YandexGPT Pro

YandexGPT Lite

Эмбеддинги

Токенизатор

Суммаризация

Классификация

Инференс

Стриминг

Для real-time-коммуникаций

Синхронно

Стандарт

Асинхронно

Know-how

3 НОВЫХ курса от Yandex Cloud

Курс «AI для бизнеса»

Курс «ML-сервисы:
ключевые решения»

- Модуль 1. Обзор ML-сервисов Yandex Cloud
- Модуль 2. Большие языковые модели

Курс «Основы работы
с Yandex SpeechSense»

8 модулей

Опыт внедрения решений ML/AI

Яндекс	80+ внедрений GenAI		GenAI
Yandex Cloud	Нейросети по мониторингу экологии		ML ВШЭ + ШАД
«Норникель»	Разработка решений для ИЦК Создание ML-прототипов	Ускорение разработки	ML
«Салым Петролеум»	Предиктивный анализ работы оборудования	Сокращение сроков вычислений Снижение нагрузки	ML, платформа данных Glowbyte
«Газпром нефть»	Сервис геонавигации по СМП		ML
«Русал»	Гибридная ML-инфраструктура	Ускорение разработки ML Сокращение затрат	ML
«РусГидро» (ДЭК)	Умный голосовой ассистент	85% — доля автоматически обрабатываемых звонков	ML, SpeechKit
УБРиР	Гибридная ML-инфраструктура	Ускорение разработки ML. Сокращение затрат	ML, GPU
«Т-Банк»	Разработка ML-моделей для рекомендательных систем	Ускорение цикла разработки, обучения ML-моделей	ML, GPU

Опыт внедрения решений ML/AI

«МегаФон»	ТГ-бот по административно-кадровым вопросам. Поиск по базе знаний	Сокращение ручной обработки запросов на 50% Ускорение поиска в 2–3 раза	GenAI
«Инфосистемы Джет»	Чат-бот service-desk	20–30% запросов обрабатывается автоматически 75% — точность классификации запроса 80% — доля корректных запросов LLM	GenAI
ПИК	Суммаризация встреч ВКС	Подготовка саммари сократилась на 75% — до 5–10 минут Совещания ведутся более структурированно Среднее время саммари — 10 минут	GenAI
«МТС Линк»	Суммаризация встреч	Освободить 15–25 минут в день на рутинных задачах	GenAI
«Банки.ру»	AI-ассистент		GenAI

Ускорим получение эффекта от ML/GenAI

Поиск идей

Изучим ваш бизнес и расширим спектр возможностей для роста с помощью GenAI

Интервью с владельцами бизнес-функции

MVP

Совместное проведение MVP «как в Яндексе»

Оценка и фокусировка

Фокусировка на направлениях с наибольшим потенциалом

Паспорта MVP-гипотез, включающие архитектурный дизайн, структуру рабочих групп и расчёт ожидаемого эффекта

Постпроектная оценка

Оценка эффекта для принятия решения о масштабировании

Постпроектное интервью и отчёт о влиянии AI на показатели деятельности

Спасибо!



Роман Матусевич,
руководитель индустрии «Нефть, газ,
энергетика» Yandex Cloud
romatu@yandex-team.ru