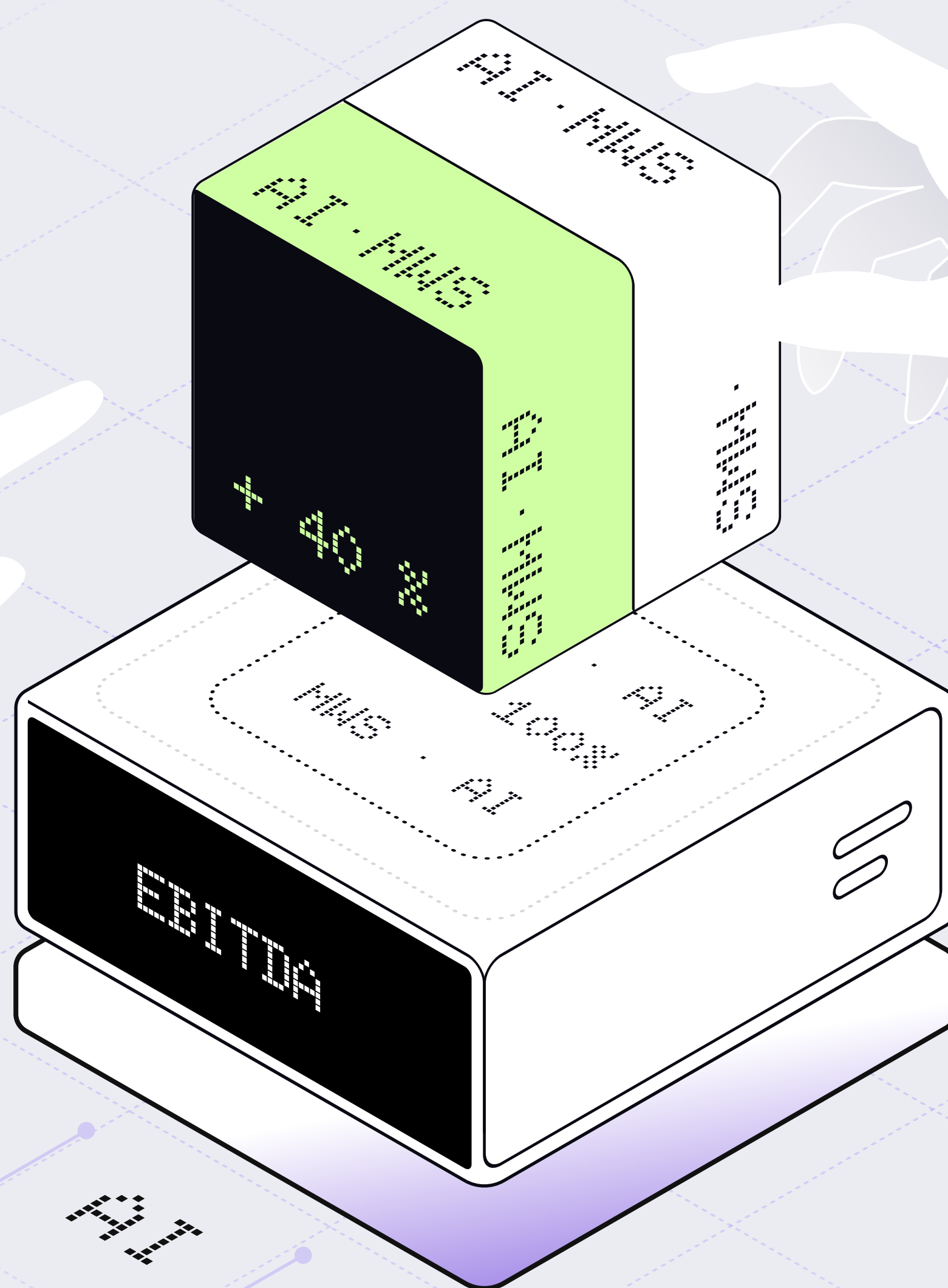


К эффекту от AI приводит команда

Гайд: как собрать команду, которая делает результат



Зачем вам этот гайд

Главный барьер при запуске ИИ-трансформации сегодня — это не выбор моделей, не бюджеты на ИИ, а найм и «сборка» сильной управленческой ИИ-команды. Рынок перегрет, роли гибридизируются, а стандартные методы подбора IT-специалистов не работают: компании либо переплачивают за нерелевантный опыт, либо месяцами не могут закрыть ключевые позиции для ролей с приставкой «ИИ».

Попытки собрать ИИ-команду для реализации ИИ стратегии «вслепую» обычно приводит к раздутым бюджетам, ошибкам в грейдинге и затянутым на месяцы поискам, а когда команда сформирована, она не приносит тот результат, который ожидали акционеры.

Как понять, кто именно нужен под стратегию вашей компании, какими компетенциями должна обладать такая команда, какой реально бюджет необходим на зарплатный фонд такой команды, всё это вы найдёте в нашем гайде.

Мы начали ИИ трансформацию с себя и прожили этот опыт вместе с нашими клиентами и командами.

ЦК ИИ-трансформации MWS AI входит в группу МТС и сопровождает компании группы в процессе реализации ИИ стратегий, ИИ трансформации и внедрения ИИ в продукты и процессы. Это компании с десятками тысяч сотрудников и всеми типичными сложностями большой компании.

Реализуя такие программы ИИ внутри группы, мы обрабатываем воронку из более чем двух тысяч ИИ инициатив и сотен проектов в год. На таком масштабе отчётливо видны системные точки роста и ключевые барьеры: возвратность инвестиций в ИИ определяется правильным орг. дизайном, корпоративной культурой и операционной моделью. Мы прошли этот путь сами, проходили его несколько раз с нашими клиентами. Из этой практики вырос гайд, который вы сейчас читаете.

Между стратегией на слайдах и эффектом в P&L лежит **операционный слой**: кто именно, в каких командах, по каким правилам и с какими навыками ежедневно получает результат от ИИ. Этот слой нельзя купить вместе с лицензией и нельзя утвердить приказом. Его выстраивают люди. Именно поэтому мы уделили особое внимание этой теме и разработали с нашими экспертами гайд «К эффекту от AI приводит команда. Как собрать команду, которая делает результат».

Уверены, что гайд станет настольной книгой по орг. дизайну и найму ИИ команд для эффективной реализации вашей ИИ стратегии.

Введение и методология

Зачем этот гайд, для кого он
и как собраны данные исследования

В ЭТОМ РАЗДЕЛЕ

- 1 Зачем этот гайд · содержание
- 2 Методология · 5 методов
- 3 Глоссарий терминов



01

Введение и содержание

Этот материал создан для компаний, которые стоят перед решением о внедрении AI — независимо от того, на каком этапе они находятся: только формируют стратегию, запускают первый PoC или уже масштабируют AI-функцию. Мы собрали здесь то, что выяснили на практике за несколько лет работы с AI-командами: через собственный найм, через ошибки и через измеримые результаты.

Цель гайдбука — дать руководителям и HR-специалистам **структурированную основу для принятия решений**: кого нанимать, в какой последовательности, сколько это стоит и чего ожидать от рынка. Не абстрактные рекомендации, а данные и выводы, которые можно применить сегодня.

AI-трансформация — это организационный проект, а не технологический.
Главное в нём — люди. Эта методичка о том, как их найти, оценить и удержать.

Содержание · 7 разделов

Разделы гайдбука		Проблема / задача	Что внутри	Ценность
1	Рынок AI-специалистов	Что ждать от рынка ближайшие 3 года	Ключевые тренды 2026–2029: агентный AI, импортозамещение, гибридизация ролей · эффект «сейчас» и прогноз «через 3 года» · взгляды СТО и СРО из первых уст	Готовые инструменты, позволяющие принимать решения о найме с опорой на рынок, а не на интуицию
2	Роли AI-команды	Какие роли нужны, в какой последовательности нанимать	10 профилей: задачи, требования, вилка ЗП · логика: кто нужен на стадии PoC, кто — в Prod	Нанимающий менеджер видит полную картину до начала поиска
3	Зарплатные ожидания · рыночные данные 2026	Бюджет занижен или завышен — непонятно, как обосновать	Как собрать бюджет и обосновать его	Защита бюджета с рыночными ориентирами, а не «на глаз»
4	Орг. структуры + матрицы компетенций	Как выстроить AI-офис и оценить уровень кандидатов	PoC и Prod команды, их структуры с триггерами перехода. Матрицы компетенций по каждой роли от Basic до Expert. Инструмент для оценки кандидатов на интервью и их грейдинга.	Выбор оргструктуры под стадию; единые критерии оценки кандидата
5	Опыт найма · 142 позиции, 2025	Как не наступить на грабли и сохранить деньги при найме	Обзор сроков закрытия вакансий, компаний-доноров, каналов поиска кандидатов.	Понимание, где искать кандидатов и на какой срок рассчитывать
6	Мотивация	Как сделать так, чтобы оффер приняли	Exit-интервью: топ-причины ухода AI-специалистов · что важно инженеру vs исследователю vs руководителю · как продать вакансию кандидату из крупной tech-компании	Оффер формулируется под конкретный профиль, а не универсально
7	Экспертные интервью	Как это устроено у тех, кто уже строил AI-команды	Разбор процессов построения команды с СРО, СТО и CAIO	Практический опыт тех, кто прошёл этот путь — без фильтра и маркетинга

Данные: MWS AI, hh.ru, Forbes, Habr, McKinsey, Gartner, Deloitte, MTS AI · Версия май 2026

Этот гайдбук даёт структуру, данные и опыт, с которыми каждая компания может найти свой путь.

02

Методология исследования

Методология разработана так, чтобы каждый вывод в гайдбук опирался на верифицированные данные, а не на экспертные предположения. Мы сочетали анализ открытых источников с первичными исследованиями — собственной практикой найма, структурированными интервью и количественными данными по рынку.

Методы исследования · 5 этапов

- 1

Кабинетное исследование
Изучение открытых российских и зарубежных источников: отчёты hh.ru, Forbes, Habr, McKinsey, Gartner, Deloitte, MTS AI и отраслевые публикации по рынку AI-специалистов. Анализ зарплатной аналитики, динамики вакансий и резюме, трендов технологического стека 2024–2026.
ВТОРИЧНЫЕ ДАННЫЕ · РЫНОК ТРУДА И ТЕХНОЛОГИЙ
- 2

Анализ собственной практики найма
Структурированный разбор закрытых позиций в AI-командах за 2025 год: сроки закрытия по ролям, конверсия на каждом этапе воронки, эффективность каналов поиска, распределение офферов. Основа — внутренняя HR-аналитика MWS AI.
ПЕРВИЧНЫЕ ДАННЫЕ · 2025 ГОД
- 3

Exit-интервью и количественный опрос
Систематизированные данные exit-интервью сотрудников AI-подразделений: причины смены работы, мотиваторы перехода, оценка условий труда. Данные использованы для построения профилей мотивации по каждой роли и разработки рекомендаций.
ПЕРВИЧНЫЕ ДАННЫЕ · EXIT-ИНТЕРВЬЮ
- 4

Глубинные интервью с экспертами
Структурированные интервью с ключевыми руководителями AI-функций: СТО, СРО и CAIO компаний с реальным опытом построения AI-команд. Каждое интервью охватывает 8 тематических блоков: рынок, роли, скрининг, KPI, мотивация, типичные ошибки и горизонт 5–10 лет.
ПЕРВИЧНЫЕ ДАННЫЕ · ИЮНЬ 2026
- 5

Экспертная разработка рекомендаций
На основе синтеза всех источников — матрицы компетенций для 10 ролей (Basic → Expert), типовые оргструктуры для стадий PoC и Production с триггерами перехода, чек-лист HR-специалиста из 5 шагов, сравнительный анализ экспертных позиций и итоговые выводы для ЛПР.
АНАЛИТИЧЕСКИЙ СИНТЕЗ · ПРАКТИЧЕСКИЕ ИНСТРУМЕНТЫ

Каждый вывод методички подкреплён хотя бы двумя независимыми источниками. Там, где данные расходятся — мы фиксируем оба взгляда и объясняем разницу.

Источники: MWS AI, hh.ru, Forbes, Habr, McKinsey, Gartner, Deloitte, MTS AI · Период исследования: 2025–2026

03

Глоссарий терминов

Англицизмы и профессиональный сленг, встречающиеся в гайдбуке — с короткими пояснениями на человеческом языке.

КАТЕГОРИИ (Всего 83 термина)

- **Модели и обучение**

23 терм.
- **Инфраструктура и MLOps**

20 терм.
- **Продукт и процессы**

15 терм.
- **Роли и грейды**

6 терм.
- **Найм и HR-метрики**

9 терм.
- **Рынок и стратегия**

10 терм.

Модели и обучение23 терминов		
МОДЕЛИ И ОБУЧЕНИЕ LLM Large Language Model Большая языковая модель — нейросеть, обученная на огромных объёмах текста. Предсказывает следующее слово и за счёт этого ведёт диалог, пишет и анализирует тексты. Основа ChatGPT, Claude и подобных.	МОДЕЛИ И ОБУЧЕНИЕ VLM Vision-Language Model Мультимодальная модель, которая работает одновременно с текстом и изображениями.	МОДЕЛИ И ОБУЧЕНИЕ RAG Retrieval-Augmented Generation «Генерация с поиском»: перед ответом модель находит нужные факты в базе документов компании и отвечает на их основе. Снижает выдумки и позволяет работать с внутренними данными.
МОДЕЛИ И ОБУЧЕНИЕ Fine-tuning дообучение Адаптация уже обученной модели под конкретную задачу или предметную область компании.	МОДЕЛИ И ОБУЧЕНИЕ SFT Supervised Fine-Tuning Дообучение «с учителем» — донастройка модели на размеченных примерах «запрос → эталонный ответ».	МОДЕЛИ И ОБУЧЕНИЕ LoRA / QLoRA Low-Rank Adaptation Экономный метод дообучения: вместо переобучения всей модели обучается небольшая «надстройка». Резко удешевляет и ускоряет процесс.
МОДЕЛИ И ОБУЧЕНИЕ DPO Direct Preference Optimization Метод дообучения по человеческим предпочтениям — модель учат на парах «какой из двух ответов лучше».	МОДЕЛИ И ОБУЧЕНИЕ RLHF Reinforcement Learning from Human Feedback Обучение с подкреплением на оценках людей. Так модель приучают отвечать полезно и безопасно.	МОДЕЛИ И ОБУЧЕНИЕ Квантизация quantization «Сжатие» модели: снижение точности её чисел, чтобы она занимала меньше памяти и работала быстрее — почти без потери качества.
МОДЕЛИ И ОБУЧЕНИЕ Дистилляция distillation Перенос знаний большой модели в меньшую, более дешёвую в эксплуатации.	МОДЕЛИ И ОБУЧЕНИЕ Токен / токенизация token Токен — кусочек текста (слово или его часть), которым оперирует модель. От числа токенов зависят стоимость запроса и длина ответа.	МОДЕЛИ И ОБУЧЕНИЕ Эмбеddинг embedding Представление текста в виде набора чисел (вектора), по которому модель измеряет смысловую близость фрагментов. Основа поиска в RAG.
МОДЕЛИ И ОБУЧЕНИЕ Промпт prompt Запрос-инструкция, которую человек даёт модели. От формулировки сильно зависит качество ответа.	МОДЕЛИ И ОБУЧЕНИЕ Prompt engineering промпт-инжиниринг Дисциплина составления, тестирования и оптимизации промптов.	МОДЕЛИ И ОБУЧЕНИЕ Chain-of-Thought (CoT) цепочка рассуждений Приём, при котором модель пошагово рассуждает «вслух», повышая точность на сложных задачах.
МОДЕЛИ И ОБУЧЕНИЕ Few-shot / zero-shot Приёмы запроса: дать модели несколько примеров (few-shot) или ни одного (zero-shot) перед самой задачей.	МОДЕЛИ И ОБУЧЕНИЕ Агентные системы agentic AI, который не просто отвечает, а сам планирует и выполняет последовательность действий — вызывает инструменты, ищет данные — ради достижения цели.	МОДЕЛИ И ОБУЧЕНИЕ Мультиагентная оркестрация multi-agent orchestration Координация нескольких AI-агентов, каждый из которых решает свою часть общей задачи.

МОДЕЛИ И ОБУЧЕНИЕ Function calling вызов функций Способность модели обращаться к внешним инструментам и API прямо в ходе ответа.	МОДЕЛИ И ОБУЧЕНИЕ Галлюцинации hallucinations Уверенно звучащие, но недостоверные ответы модели. Ключевой риск, который контролируют через RAG и eval.	МОДЕЛИ И ОБУЧЕНИЕ Ground truth эталонные данные Размеченные людьми «правильные» данные, с которыми сравнивают ответы модели при обучении и оценке.
МОДЕЛИ И ОБУЧЕНИЕ SOTA State of the Art Текущий лучший результат или передовой уровень технологий в области.	МОДЕЛИ И ОБУЧЕНИЕ Eval / Evaluation оценка качества Измеримая проверка качества модели по заранее заданным метрикам и тестовым наборам.	

Инфраструктура и MLOps20 терминов

ИНФРАСТРУКТУРА И MLOPS MLOps Machine Learning Operations Практики и инструменты для надёжного вывода ML-моделей в эксплуатацию и их сопровождения. Аналог DevOps, но для моделей.	ИНФРАСТРУКТУРА И MLOPS LLMOps То же, что MLOps, но для больших языковых моделей: версии промптов, контроль стоимости и качества генераций.	ИНФРАСТРУКТУРА И MLOPS DevOps Культура и инструменты, объединяющие разработку и эксплуатацию ПО ради быстрых и надёжных релизов.
ИНФРАСТРУКТУРА И MLOPS Pipeline / пайплайн Автоматизированная цепочка обработки: от исходных данных до готового результата.	ИНФРАСТРУКТУРА И MLOPS Inference / инференс Режим работы обученной модели «в бою» — получение ответов на запросы, в отличие от обучения.	ИНФРАСТРУКТУРА И MLOPS Deploy / деплой Вывод модели или сервиса в рабочую среду.
ИНФРАСТРУКТУРА И MLOPS Production (prod) продакшен «Боевая» среда, где система работает с реальными пользователями и нагрузкой.	ИНФРАСТРУКТУРА И MLOPS GPU Graphics Processing Unit Графический процессор — основное «железо» для обучения и работы AI-моделей.	ИНФРАСТРУКТУРА И MLOPS Observability наблюдаемость Возможность в реальном времени видеть, что происходит с системой: качество ответов, ошибки, стоимость, деградация.
ИНФРАСТРУКТУРА И MLOPS Latency задержка Время от запроса пользователя до ответа системы.	ИНФРАСТРУКТУРА И MLOPS SLA Service Level Agreement Соглашение о гарантированном уровне сервиса — например, доступность 99,9%.	ИНФРАСТРУКТУРА И MLOPS CI/CD Continuous Integration / Delivery Автоматизация сборки, тестирования и выкладки кода для частых и безопасных релизов.
ИНФРАСТРУКТУРА И MLOPS Docker / Kubernetes Docker упаковывает приложения в контейнеры, Kubernetes управляет их массовым запуском и масштабированием.	ИНФРАСТРУКТУРА И MLOPS Drift / дрейф Постепенное ухудшение качества модели из-за того, что реальные данные со временем меняются.	ИНФРАСТРУКТУРА И MLOPS Retrain / петрейн Повторное дообучение модели на свежих данных.
ИНФРАСТРУКТУРА И MLOPS Cost-per-request стоимость запроса Цена одного обращения к модели. Ключевая метрика экономики LLM-продукта.	ИНФРАСТРУКТУРА И MLOPS vLLM / Triton Системы быстрой «подачи» (serving) моделей в продакшене — обслуживают много запросов с минимальной задержкой.	ИНФРАСТРУКТУРА И MLOPS Векторная база vector database Хранилище эмбеддингов для быстрого смыслового поиска: Milvus, Qdrant, FAISS и аналоги.
ИНФРАСТРУКТУРА И MLOPS Reranking переранжирование Повторное упорядочивание найденных фрагментов по релевантности — шаг повышения качества в RAG.	ИНФРАСТРУКТУРА И MLOPS Chunking / чанкинг Нарезка документов на фрагменты для индексации и поиска.	

Продукт и процессы

15 терминов

ПРОДУКТ И ПРОЦЕССЫ PoC Proof of Concept Проверка гипотезы на прототипе перед полноценной разработкой — «доказательство работоспособности».	ПРОДУКТ И ПРОЦЕССЫ MVP Minimum Viable Product Минимально жизнеспособный продукт: самая простая версия, которой уже можно проверить спрос.	ПРОДУКТ И ПРОЦЕССЫ PMF Product-Market Fit Соответствие продукта реальной потребности рынка — точка, после которой продукт начинает «тянуть» спрос.
ПРОДУКТ И ПРОЦЕССЫ Roadmap дорожная карта План: что и в какой последовательности делает команда.	ПРОДУКТ И ПРОЦЕССЫ Backlog / бэклог Приоритизированный список задач команды.	ПРОДУКТ И ПРОЦЕССЫ Sprint / спринт Короткий рабочий цикл (обычно 1–2 недели) в Agile/Scrum.
ПРОДУКТ И ПРОЦЕССЫ Pivot / пивот Смена направления продукта или гипотезы по итогам проверки.	ПРОДУКТ И ПРОЦЕССЫ Discovery / CustDev Customer Development Этап исследования потребностей пользователей до начала разработки.	ПРОДУКТ И ПРОЦЕССЫ A/B-тест Сравнение двух вариантов на реальных пользователях, чтобы выбрать лучший по метрикам.
ПРОДУКТ И ПРОЦЕССЫ EDA Exploratory Data Analysis Разведочный анализ данных перед построением модели: поиск закономерностей, пропусков, аномалий.	ПРОДУКТ И ПРОЦЕССЫ Feature engineering Создание информативных признаков из сырых данных для повышения качества модели.	ПРОДУКТ И ПРОЦЕССЫ KPI Key Performance Indicator Ключевой показатель эффективности.
ПРОДУКТ И ПРОЦЕССЫ ROI Return on Investment Окупаемость вложений — отдача от инвестиций.	ПРОДУКТ И ПРОЦЕССЫ GTM Go-to-Market Стратегия вывода продукта на рынок.	ПРОДУКТ И ПРОЦЕССЫ Adoption Уровень реального использования продукта внутри компании или пользователями.

Роли и грейды

6 терминов

РОЛИ И ГРЕЙДЫ Грейд grade Уровень квалификации сотрудника, определяющий зону ответственности и зарплатную вилку.	РОЛИ И ГРЕЙДЫ Junior / Middle / Senior / Lead Ступени квалификации — от начинающего специалиста до ведущего и технического лидера команды.	РОЛИ И ГРЕЙДЫ Head of AI / CAIO Chief AI Officer Руководитель AI-направления компании: стратегия, команда, бюджет, ответственность за эффект.
РОЛИ И ГРЕЙДЫ CTO / CPO / CDTO Директора по технологиям (CTO), по продукту (CPO) и по цифровой трансформации (CDTO).	РОЛИ И ГРЕЙДЫ FTE Full-Time Equivalent «Полная штатная единица». 1 FTE — один сотрудник на полной занятости; 0,5 FTE — половина.	РОЛИ И ГРЕЙДЫ Hard / soft skills Твёрдые (технические) и гибкие (коммуникация, командность, мышление) навыки.

Найм и HR-метрики

9 терминов

НАЙМ И HR-МЕТРИКИ Time-to-hire срок найма Срок закрытия вакансии — от старта поиска до выхода кандидата или принятия оффера.	НАЙМ И HR-МЕТРИКИ Quality of hire качество найма Насколько успешны новые сотрудники после выхода (например, доля прошедших испытательный срок).	НАЙМ И HR-МЕТРИКИ Скрининг screening Первичный отбор кандидатов по резюме и короткому интервью.
НАЙМ И HR-МЕТРИКИ Оффер offer Официальное предложение о работе.	НАЙМ И HR-МЕТРИКИ Exit-интервью exit interview Беседа с увольняющимся сотрудником о причинах ухода — источник данных о мотивации.	НАЙМ И HR-МЕТРИКИ Воронка funnel Путь кандидата по этапам отбора с измерением конверсии на каждом шаге.

НАЙМ И HR-МЕТРИКИ Сорсинг / рекрутинг sourcing Активный поиск и привлечение кандидатов.	НАЙМ И HR-МЕТРИКИ Онбординг onboarding Адаптация и ввод нового сотрудника в работу.	НАЙМ И HR-МЕТРИКИ ATS Applicant Tracking System Система управления подбором персонала — хранит вакансии, кандидатов и этапы воронки.
--	--	---

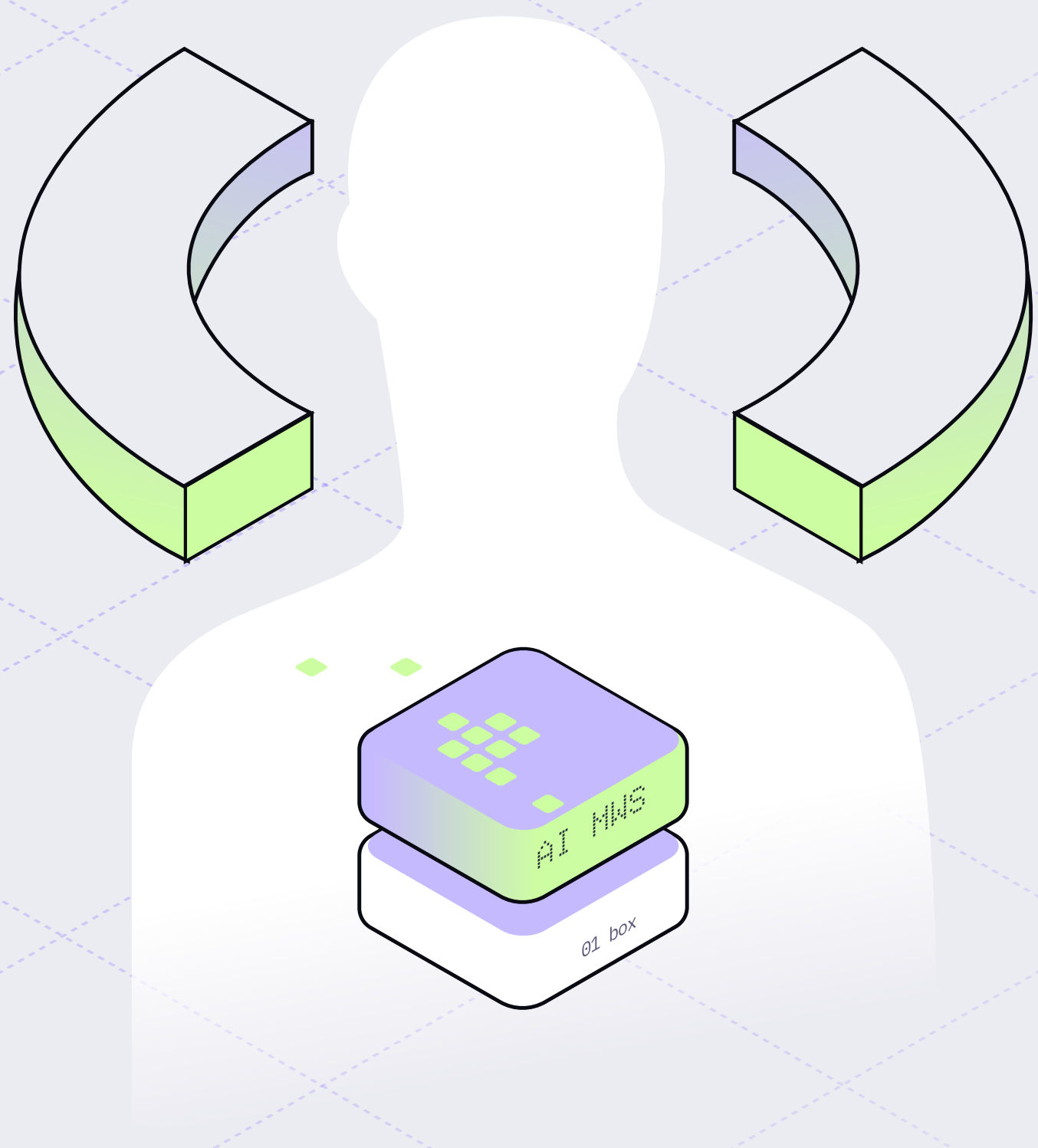
Рынок и стратегия			10 терминов
РЫНОК И СТРАТЕГИЯ Импортозамещение Переход на отечественные решения вместо зарубежных, в том числе в ML-стеке.	РЫНОК И СТРАТЕГИЯ Open-source открытый код ПО с открытым исходным кодом, которое можно свободно использовать и дорабатывать.	РЫНОК И СТРАТЕГИЯ Enterprise Сегмент крупных корпоративных заказчиков с высокими требованиями к надёжности и безопасности.	
РЫНОК И СТРАТЕГИЯ B2B Business-to-Business Продажи и продукты для компаний, а не для частных лиц.	РЫНОК И СТРАТЕГИЯ On-premise Размещение систем на собственных серверах компании, а не в облаке.	РЫНОК И СТРАТЕГИЯ Суверенный AI Sovereign AI AI-инфраструктура и модели под контролем внутри страны или компании, с учётом регуляторных требований (152-ФЗ и др.).	
РЫНОК И СТРАТЕГИЯ Legacy Устаревшие, но всё ещё используемые системы, с которыми приходится считаться.	РЫНОК И СТРАТЕГИЯ Build vs buy Стратегический выбор: разрабатывать решение самим или купить готовое.	РЫНОК И СТРАТЕГИЯ Стек stack Набор технологий и инструментов, на которых строится решение.	
РЫНОК И СТРАТЕГИЯ Фреймворк framework Готовый каркас и набор библиотек для ускорения разработки (LangChain, LlamaIndex и др.).			

Роли и компетенции

10 профилей AI-команды: задачи, требования и уровни мастерства Basic → Expert

В ЭТОМ РАЗДЕЛЕ

- 1 10 профилей AI-команды
- 2 Матрицы компетенций по ролям



04

Роли AI-команды — 10 профилей

Десять ролей, из которых собирается AI-команда: от стратегического лидера до операционных профилей. По каждой — суть роли, три ключевые области (где применимо), задачи, требования и то, что будет плюсом.

1

Head of AI / CAIO

800к – 1,5М ₽/мес

Руководитель, чья компетенция лежит на пересечении трёх областей: стратегического управления AI-направлением, построения и масштабирования высокоэффективных AI-команд, а также трансформации бизнеса через внедрение AI в продукты, процессы и инфраструктуру.

Три ключевые области

AI Strategy & Governance Стратегическое управление AI-направлением на уровне компании	People & Organization Построение и масштабирование высокоэффективных AI-команд	AI-driven Transformation Трансформация бизнеса через внедрение AI в продукты, процессы и инфраструктуру
---	--	---

Ключевые задачи

- Формирование и реализация AI-стратегии: от горизонта исследований до продуктового roadmap и инвестиционных приоритетов
- Построение организационной структуры AI-функции: найм и развитие исследователей, инженеров, архитекторов, аналитиков
- Управление портфелем AI-инициатив: приоритизация, ROI, закрытие неэффективных направлений
- Кросс-функциональное взаимодействие AI-команды с продуктовыми, инженерными, коммерческими подразделениями
- Формирование культуры data-driven и AI-first принятия решений на уровне компании
- Представление AI-повестки на уровне совета директоров, инвесторов и ключевых партнёров
- Выстраивание отношений с российскими регуляторами и госструктурами в части AI-политики
- Управление рисками: этика, безопасность, 152-ФЗ, отраслевые стандарты

Требования

- Подтверждённый опыт руководства AI/ML-направлением с измеримыми бизнес-результатами
- Глубокое понимание AI-стека: LLM/VLM, методы дообучения, RAG, агентные системы, MLOps
- Опыт построения и масштабирования команд 10+ человек с разнородными профилями
- Понимание российского AI-рынка: игроки, регуляторика, enterprise-продажи
- Опыт взаимодействия с C-level: перевод технологий в бизнес-язык
- Английский — достаточный для мониторинга SOTA и участия в международных мероприятиях

Будет плюсом

- Публичная экспертная позиция: AI Journey, AIDAO, SmartData, Skolkovo, FinTech, GovTech
- Академический бэкграунд или взаимодействие с исследовательским сообществом (МФТИ, ВШЭ, ИИП РАН)
- Опыт M&A или инвестиционной деятельности в AI-сегменте
- Участие в формировании национальной AI-повестки или отраслевых стандартов

Технический профиль и экспертиза

Кандидат должен обладать практической экспертизой уровня, достаточного для архитектурного ревью, оценки исследовательских направлений и найма специалистов уровня Staff/Principal — в каждом из доменов:

- **LLM и NLP**
предобучение, SFT, RLHF/DPO, RAG с гибридным поиском и reranking, агентные архитектуры (ReAct, multi-agent, tool use), оценка качества на русскоязычных данных
- **Voice AI**
ASR (CTC, RNN-T, Whisper-подобные архитектуры, streaming/offline), TTS (VITS, XTTS, нейронные вокодеры), KWS, VAD, диаризация, шумоподавление
- **Computer Vision**
детекция (YOLO, DETR), трекинг (ByteTrack), сегментация (SAM), видеоаналитика в реальном времени, интеграция с RTSP/RTMP и VMS
- **Tabular & Structured Data**
градиентный бустинг (CatBoost, LightGBM), временные ряды, скоринг, детекция аномалий, рекомендательные системы
- **MLOps и инфраструктура**
версионирование данных и моделей (MLflow, DVC, W&B), CI/CD для ML, мониторинг качества и data drift, управление GPU-кластерами, оптимизация стоимости инференса на российских и on-premise платформах

2

ML/AI Team Lead

600к – 800к ₽/мес

Специалист на пересечении технического лидерства и развития команды, проектирования и владения ML-системами в production и трансляции исследовательских прототипов в надёжные продуктовые решения.

Три ключевые области

Technical Leadership Техническое лидерство и развитие AI-команды	ML Systems in Production Проектирование и владение ML-системами в production	Research Product Трансляция прототипов в надёжные решения
--	--	---

Ключевые задачи

- Техническое руководство командой 6–12 инженеров: code review, архитектурные сессии, менторинг
- Проектирование и владение архитектурой ML-пайплайнов: от ingestion до деплоя, мониторинга и откатов
- Выбор и обоснование модельного стека: open-source LLM, мультимодальные и специализированные модели
- Проектирование агентных систем и сценарных LLM-ядер: планировщики, RAG, оркестрация
- Управление надёжностью и стоимостью GPU-инфраструктуры
- Построение observability: трассировка генераций, алертинг деградации, cost-per-request
- Формирование технического roadmap с AI Architect и Head of AI

Требования • 5+ лет production-разработки на Python; Go — преимущество • ML и Deep Learning: трансформеры, SFT, LoRA, DPO, диффузионные модели • Деплой на GPU: квантизация (AWQ, GGUF, GPTQ), сервинг через vLLM / TGI / Triton • Микросервисы, Docker, CI/CD, базовый Kubernetes, Prometheus / Grafana • Опыт с российскими облаками и on-premise GPU	Будет плюсом • Production-опыт с LLaMA, Mistral, Qwen под русский язык • RAG корпоративного уровня: гибридный поиск, reranking, eval • Агентные оркестраторы: LangGraph, CrewAI, Autogen • Профильное образование: МФТИ, ВШЭ, МГУ, ИТМО
--	--

3

AI Architect

400к – 800к Р/мес

Специалист на пересечении системного проектирования AI-решений в условиях суверенной инфраструктуры, интеграции моделей в высоконагруженные среды и архитектуры данных и инференс-пайплайнов.

Три ключевые области

Sovereign AI Systems

Системное проектирование на суверенной инфраструктуре

LLMOps / MLOps

Интеграция моделей в высоконагруженные среды

Data & Inference

Архитектура данных и инференс-пайплайнов

Ключевые задачи

- Проектирование сквозной архитектуры AI-продуктов: от источников данных до интерфейсов и SLA
- Выбор и обоснование модельного стека: свои модели, дообучение open-source, RAG, агенты
- Разработка паттернов интеграции LLM в микросервисы с учётом 152-ФЗ и ИБ
- Оптимизация стоимости и производительности инференса на доступном GPU-парке
- Построение evaluation-фреймворков и мониторинга качества в production
- Формирование технических стандартов AI-разработки, ревью архитектур
- Адаптация мировых AI-практик к российским условиям

Требования

- Глубокое понимание LLM/VLM: трансформеры, дообучение, RAG-паттерны, агентные фреймворки
- Опыт вывода AI-систем в production, желательно в высоконагруженной среде
- Python (глубоко); Go или C++ — преимущество. vLLM / Triton, LangChain / LlamaIndex, MLflow, Docker, Kubernetes, Kafka
- Практика с Mistral, LLaMA, Qwen
- Понимание 152-ФЗ и ИБ

Будет плюсом

- Опыт с отечественными AI-платформами, участие в импортозамещении
- Мультиагентные системы корпоративного уровня
- Внутренние Developer Platform
- Вклад в open-source, выступления на конференциях

4

ML Engineer

300к – 600к ₽/мес

Инженер, который берёт сложную бизнес-задачу, быстро собирает рабочий PoC, валидирует на реальных данных и итеративно доводит до production. Ключевая ценность — скорость от идеи до измеримого результата.

Три ключевые области

LLM Application Eng. Разработка и интеграция LLM-систем в продукт	Agentic & RAG Проектирование надёжных AI-пайплайнов и агентов	Eval-driven Dev. Итеративное улучшение качества через метрики
---	---	---

Ключевые задачи

- Проектирование RAG-систем: индексация, гибридный поиск, reranking, query decomposition
- Разработка AI-агентов и copilot-систем: ReAct, function calling, память, оркестрация
- Дообучение моделей под домены: SFT, LoRA/QLoRA, DPO
- Внедрение observability: трейсинг, мониторинг качества, контроль токенов
- Проектирование экспериментов: гипотезы, eval-фреймворки, итерация
- Работа с российскими и open-source моделями при ограниченном доступе к зарубежным API

Требования

- 3+ лет прикладной AI/ML-разработки с фокусом на NLP и LLM
- Сильный Python: типизация, async, ООП, тесты
- Глубокое понимание RAG: векторный и гибридный поиск, чанкинг
- LangChain, LangGraph, LlamaIndex; Milvus, Qdrant, Chroma, FAISS
- LLM observability: LangSmith, LangFuse
- Микросервисы: Kafka, Redis, FastAPI / gRPC

Будет плюсом

- Production-опыт с корпоративным RAG
- Fine-tuning и дистилляция, квантизация (LoRA, AWQ, GGUF), сервинг через vLLM
- LangGraph, CrewAI, Autogen
- Knowledge graph в RAG
- Low-code: LangFlow, n8n

5

ML разработчик-исследователь

200к – 500к Р/мес

Специалист, решающий задачу reduction to practice: преобразование математического описания модели из препринта на arXiv в обучаемый и воспроизводимый тензорный конвейер.

Три ключевые области

Replicable Science

Воспроизводимое научное исследование

HPC

CUDA, распределённое обучение

PyTorch / JAX Eng.

Инженерия динамических вычислительных графов

Ключевые задачи

- Разработка, обучение и валидация LLM/VLM/омни-модальных архитектур
- Проектирование и запуск экспериментов, анализ сходимости, итерация архитектур
- Оптимизация пайплайнов обучения и инференса по latency/throughput/cost
- Distributed training на multi-GPU и multi-node кластерах
- Transfer прототипов в production-ready решения
- Мониторинг SOTA, участие в соревнованиях по ML

Требования

- Глубокое знание методов обучения LLM
- PyTorch, Transformers, TensorFlow, JAX, TRL
- Python (глубоко); C++/CUDA — желательно
- Математика: линейная алгебра, матстатистика
- Hugging Face, PyTorch Lightning, Weights & Biases
- Интерес к научной работе и high-risk исследованиям

Будет плюсом

- Статьи на NeurIPS, ACL, AAAI, EMNLP, ICLR
- Кандидатская степень в ML/NLP
- Участие в open-source

6

Data Scientist

300к – 400к Р/мес

Исследователь и инженер данных в одном лице. Превращает сырые данные в проверяемые гипотезы, рабочие прототипы и production-модели. Находит скрытые закономерности и строит ML-модели под конкретные бизнес-задачи.

Ключевые задачи

- Формулировать гипотезы и ставить ML-задачи: выбор метрик, дизайн эксперимента, оценка эффекта
- Собирать, очищать и исследовать данные: EDA, обработка пропусков, аномалий, дисбаланса
- Разрабатывать и обучать модели: классический ML, ансамбли, deep learning
- Проводить валидацию: кросс-валидация, A/B-тесты, оценка стабильности и интерпретируемости
- Готовить модели к production: сериализация, документирование, передача ML-инженерам
- Мониторить качество после запуска: дрейф, переобучение, деградация признаков
- Визуализировать результаты и доносить инсайты до нетехнических стейкхолдеров

Требования

- Опыт Data Scientist 3+ лет с портфолио ML-задач
- Python: pandas, numpy, scikit-learn; PyTorch или TensorFlow
- Глубокое понимание статистики и ML: метрики, регуляризация, валидация, feature engineering
- SQL, большие объёмы данных
- A/B-тестирование
- Git, Linux, Docker; английский B2+

Будет плюсом

- Временные ряды, NLP, CV, рекомендательные системы или LLM
- MLOps: MLflow, DVC, W&B, Kubeflow, Airflow
- Опыт в регулируемых отраслях
- Публикации, Kaggle, open-source

7

AI Product Manager

300к – 450к ₺/мес

Определяет, запускает и масштабирует AI-продукты, учитывая ограничения данных, стоимость вычислений и этические риски. Не пишет модели — понимает, что нужно рынку, когда AI уместен и как превратить эксперимент в стабильный продукт.

Ключевые задачи

- Формировать продуктовую стратегию и roadmap для AI-функций или AI-native продуктов
- Управлять кроссфункциональной командой: ML-инженеры, разработчики, дизайнеры, аналитики
- Ставить требования к данным: источники, объёмы, качество, разметка, версии датасетов
- Организовывать тестирование, пилоты, A/B-эксперименты, постепенный rollout
- Мониторить поведение моделей после деплоя, планировать ретрейны и откаты
- Выявлять риски AI: безопасность, галлюцинации, интерпретируемость, регуляторика
- Работать с клиентами: выявлять потребности и адаптировать продукт под отрасли

Требования

Будет плюсом

8

Technical Project Manager

250к – 400к ₺/мес

Обеспечивает своевременную, надёжную и экономически эффективную доставку AI-решений от исследовательского прототипа до стабильного production. Выстраивает прозрачные процессы и поддерживает инженерные стандарты.

Ключевые задачи

- Планирование и контроль сроков, бюджета, ресурсов; управление рисками; работа с заказчиками
- Разработка и согласование проектной документации (план, устав, ФТ, ТЗ)
- Управление распределёнными командами
- Постановка задач (ведение backlog), контроль сроков и качества, приёмка работ
- Подготовка презентационных и отчётных материалов
- Выстраивание кросс-командного взаимодействия

Требования

Будет плюсом

9

AI-тренер / Тренер ИИ

100к – 200к ₺/мес

Выполняет аннотацию, ранжирование и генерацию ground-truth данных для этапов SFT и RLHF. Ключевое отличие от промпт-инженера: AI-тренер создаёт датасеты, которые меняют веса модели.

Ключевые задачи

- Изучение в короткие сроки сложных тем (от математики до литературы) для знакомства модели с ними
- Создание и редактирование текстов, обобщающих опыт людей в решении задач
- Проверка, оценка и исправление текстов, включая сгенерированные нейросетями
- Создание эталонов ответов и ранжирование ответов
- Этический контроль ответов AI
- Анализ информации на достоверность и выявление сомнительных источников (фактчекинг)

Требования

- Опыт работы с языковыми моделями (OpenAI, ChatGPT и аналоги)
- Идеальная грамотность и знание языка
- Навык фактчекинга: поиск первоисточников, быстрая проверка
- Системное и критическое мышление: видеть логические ошибки в красивом, но неверном тексте
- Английский — минимум C1–C2

Будет плюсом

- Высшее образование в области лингвистики, филологии, журналистики
- Опыт в роли учителя, копирайтера, переводчика, технического писателя, редактора или контент-менеджера

10

Prompt Engineer

150к – 250к ₺/мес

Новая профессия на пересечении гуманитарных и технических дисциплин, ориентированная на работу с большими языковыми моделями. Для тех, кто хочет участвовать в развитии AI-систем и глубоко понимать их работу.

Ключевые задачи

- Создание, тестирование и оптимизация промптов с использованием разных техник и фреймворков
- Тестирование и итерации (промпт-хитинг)
- Анализ и интерпретация результатов работы моделей
- Устранение галлюцинаций AI
- Разработка цепочек промптов (Chain-of-Thought)
- Создание шаблонов (промпт-библиотек) для бизнеса
- Интеграция с внешними инструментами через функции (function calling)
- Взаимодействие с исследовательской и продуктовой командами

Требования

- Опыт работы с языковыми моделями, решение задач с использованием LLM
- Понимание принципов prompt engineering и ограничений LLM
- Базовые навыки Python: pandas, re, numpy
- Опыт работы с LLM через API
- Уверенное владение JSON и CSV
- Английский B2–C1

Будет плюсом

- Образование: лингвистика, филология, журналистика или техническое
- Понимание трансформеров, токенизации, температуры, top-k/top-p
- Оптимизация промптов под архитектуру
- LangFlow, LlamaIndex
- Embedding-решения (FAISS, Pinecone)
- FastAPI, Streamlit, Gradio, Flask

05

Матрицы компетенций

Инструмент для оценки кандидатов на интервью и грейдирования. По каждой из 10 ролей — ключевые домены компетенций и описание четырёх уровней владения: Basic → Good → Strong → Expert.

Head of AI / CAIO

Стратегическое управление AI-направлением, построение и масштабирование высокоэффективных AI-команд, трансформация бизнеса через внедрение AI в продукты, процессы и инфраструктуру.

Стратегия и управление

Компетенция	Basic	Good	Strong	Expert
AI-стратегия	Понимает основы AI-ландшафта, участвует в стратегических дискуссиях	Формулирует AI-стратегию для отдельного направления, определяет приоритеты	Разрабатывает и защищает AI-стратегию компании на уровне C-level	Формирует долгосрочное AI-видение, адаптирует стратегию под глобальные технологические сдвиги
Управление портфелем инициатив	Ведёт список AI-инициатив, собирает базовые метрики	Приоритизирует инициативы по ROI, управляет ресурсами небольшого портфеля	Управляет портфелем 10+ инициатив: приоритизация, ROI, закрытие неэффективных	Формирует инвестиционную политику AI, принимает решения о M&A в AI-сегменте
AI Governance & Compliance	Знает основные требования 152-ФЗ, базовые риски AI	Выстраивает политики обработки данных, управляет этическими рисками	Формирует полноценную систему AI Governance: этика, безопасность, регуляторика	Участвует в формировании национальной AI-повестки, отраслевых стандартов
Взаимодействие с C-level и Board	Готовит материалы для руководства, участвует в презентациях	Самостоятельно представляет AI-повестку на уровне топ-менеджмента	Защищает долгосрочные AI-инвестиции, переводит технологии в бизнес-язык	Является полноправным членом стратегического руководства, формирует AI-видение компании

Люди, организация и лидерство

Компетенция	Basic	Good	Strong	Expert
Построение AI-команды	Нанимает отдельных специалистов, базовое понимание AI-профилей	Выстраивает функцию 5–10 человек, формирует инженерную культуру	Масштабирует AI-функцию до 30+ чел., строит разнородные команды (Research / Engineering / Product)	Формирует организационную структуру AI на уровне компании, управляет несколькими подразделениями
Развитие и удержание талантов	Проводит регулярные 1-1: знает личные цели каждого (рост / стабильность / баланс), даёт обратную связь, замечает признаки выгорания и приходит с разговором, а не ждёт.	Формирует планы развития и систему грейдирования; ежеквартально ревьюит нагрузку и зоны роста, подбирает задачи под цели, убирает блокеры (доступы, бюджет на обучение), делегирует задачи и полномочия.	Создаёт карьерные треки для AI-специалистов и удерживает ключевых лидеров; формирует культуру горизонтального признания и право ошибаться, балансирует траектории людей и интересы системы на горизонте 1–2 лет. Сигналы: не уходят при первом оффере, хотят попасть из других команд.	Формирует EVP для AI-функции и масштабирует развивающую среду на всю организацию: прозрачные карьерные треки и грейды, внутреннее AI-сообщество и академия, измеримое удержание талантов (retention, eNPS).
Саморегуляция	Замечает свои сильные реакции (раздражение, тревогу) до того, как они проявятся в общении с командой; берёт паузу перед сложным разговором в реактивном состоянии.	Знает свои паттерны под давлением (микроменеджмент, избегание конфликта) и осознанно с ними работает; быстро восстанавливается после сложных разговоров, не переносит «хвост» дальше.	Сохраняет ясность в кризисах и многосторонних конфликтах; своим состоянием задаёт норму — спокойствие под нагрузкой, признание ошибок без защиты.	Делает саморегуляцию общекомандной практикой и служит ролевой моделью для других руководителей AI-функции; помогает лидерам возвращаться в устойчивое состояние.

Люди, организация и лидерство (продолжение)

Компетенция	Basic	Good	Strong	Expert
Управление поляризацией	Замечает нарастающее напряжение между людьми до открытого конфликта; удерживает разговор в конструктивном русле.	Самостоятельно фасилитирует сложные переговоры; помогает сторонам выработать договорённости и правила, предотвращающие повтор конфликта.	Выявляет системные паттерны конфликтов и устраняет их на уровне культуры и процессов; формирует среду, где разногласия — ресурс, а не угроза.	Выстраивает в организации механизмы разрешения конфликтов между командами и обучает руководителей фасилитации.
Видение перспектив развития	Переводит стратегию компании в конкретные ориентиры для команды («зачем»); смотрит на горизонт 6–12 мес. и заранее готовится к изменениям.	Вовлекает команду в формирование общего видения; создаёт условия, где команда сама генерирует инициативы; добывает ресурсы под рост.	Формирует долгосрочную стратегию развития команды, согласованную с бизнесом; влияет на оргрешения исходя из перспектив системы.	Задаёт видение всей AI-функции на горизонте 3–5 лет и синхронизирует его с продуктовой и технологической стратегией компании.
Представление команды вовне	Конструктивно отстаивает интересы команды (аргументы, решения, без конфронтации); учитывает интересы организации в целом.	Защищает интересы своей системы, удерживая баланс с интересами организации.	Формирует устойчивые кросс-функциональные партнёрства на общий результат; согласовывает позиции нескольких систем, находя точки взаимной выгоды.	Представляет AI-повестку на уровне совета директоров и внешних стейкхолдеров; выстраивает альянсы, влияющие на стратегию компании.
Принятие обратной связи	Отделяет субъективные оценки от фактов, выделяет повторяющиеся паттерны; инициирует конкретные изменения по итогам фидбэка.	Выстраивает регулярный сбор обратной связи от ключевых стейкхолдеров; встраивает уроки в процессы — изменения устойчивы.	Формирует культуру открытого фидбэка (люди сами запрашивают ОС); работает с противоречивой и болезненной обратной связью, сохраняя объективность.	Внедряет 360°-обратную связь на уровне функции и использует фидбэк-данные для системных управленческих решений.

Технологическая экспертиза и рынок

Ориентир уровня Expert — опыт руководства ML-командами (100+ специалистов) и одновременного вывода в production нескольких AI-продуктов разных доменов с измеримыми бизнес-результатами и SLA. Уровень Strong+ подразумевает способность к архитектурному ревью, оценке исследовательских направлений и найму специалистов уровня Staff/Principal.

Компетенция	Basic	Good	Strong	Expert
LLM и NLP	Понимает, что такое LLM, RAG и агенты; ориентируется в базовых терминах.	Понимает методы дообучения (SFT, RLHF/DPO) и устройство RAG; корректно ставит задачи команде.	Проводит архитектурное ревью LLM-решений, оценивает компромиссы «размер / качество / стоимость инференса», нанимает сильных NLP-инженеров.	Владеет полным циклом LLM-продукта: предобучение и continual pretraining, instruction tuning (SFT), выравнивание (RLHF/DPO/PPO), RAG с гибридным поиском и reranking, агентные архитектуры (ReAct, multi-agent, tool use); задаёт стандарты оценки на русскоязычных данных.
Voice AI	Понимает, что такое ASR и TTS, и общую логику голосовых систем.	Ориентируется в стеке ASR / TTS / KWS / VAD, понимает требования к latency и качеству.	Проводит ревью архитектур речевых решений (streaming vs offline, шумоподавление, диаризация), ставит ТЗ и оценивает команду/подрядчиков.	Владеет речевым стеком на уровне архитектурных решений: ASR (CTC, RNN-T, Whisper-подобные), TTS (VITS, XTTS, нейронные вокодеры), KWS, VAD, адаптация под акцент и домен.
Computer Vision	Понимает базовые задачи CV (детекция, распознавание) и их применимость.	Ориентируется в семействах моделей (YOLO, DETR), задачах трекинга и сегментации; ставит задачи команде.	Проводит ревью CV-архитектур, оценивает специфику production-видеоаналитики (реальное время, edge/GPU, RTSP/RTMP, VMS).	Архитектурная экспертиза в детекции (YOLO, DETR), трекинге (ByteTrack, StrongSORT), сегментации (SAM, Mask R-CNN), распознавании (face / OCR / action recognition).

Технологическая экспертиза и рынок (продолжение)

Компетенция	Basic	Good	Strong	Expert
Tabular & Structured Data	Понимает, что такое классический ML и для каких задач он применим.	Ориентируется в градиентном бустинге и базовых метриках; ставит задачи DS-команде.	Проводит ревью решений на структурированных данных: бустинг, ансамбли, feature engineering, временные ряды, дисбаланс классов.	Экспертно оценивает рекомендательные, скоринговые и anomaly-detection системы в финансовом и операционном контексте (CatBoost, LightGBM, XGBoost).
MLOps и инфраструктура	Понимает, зачем нужны версионирование и мониторинг моделей.	Ориентируется в инструментах (MLflow, DVC, W&B), CI/CD для ML и базовом мониторинге.	Проводит ревью production-цикла ML: мониторинг качества и data drift, управление GPU-кластерами, контроль стоимости вычислений.	Задаёт MLOps-стандарты функции; оптимизирует стоимость инференса на российских облачных и on-premise платформах в условиях ограниченного доступа к зарубежным GPU.
Мониторинг AI-рынка	Следит за ключевыми новостями AI-рынка через общедоступные источники	Анализирует конкурентную среду, идентифицирует ключевых игроков и тренды	Формирует технологический радар, своевременно адаптирует стратегию под изменения	Задаёт направления развития AI-рынка, является публичным экспертом отрасли

AI-компетенции: применение AI

Компетенция	Basic	Good	Strong	Expert
Формирование бизнес-процесса	Процессы не описаны, работа на устных договорённостях; нет понимания границ процесса.	Описывает свой процесс текстом или схемой, понимает входы, выходы и ответственных; использует AI для структурирования описания.	Проектирует процессы для команды: SOP, регламенты, схемы; выявляет узкие места и предлагает улучшения.	Формализует процессы как алгоритмы для автоматизации (ветвления, триггеры, исключения) и проектирует сквозные процессы, где AI-агенты — полноправные исполнители; управляет через метрики и алерты.
AI-грамотность (AI-adoption)	AI воспринимается как поисковик; использование эпизодическое, без системы.	Регулярно использует AI для текстовых задач; различает «попросить» и «спроектировать запрос».	Думает в логике «что делает AI, а что — человек», встраивает AI в командные процессы.	Проектирует решения «AI-first» с агентами и интеграциями; оценивает зрелость AI-adoption организации, выстраивает стратегию внедрения, строит AI-продукты и системы.
Работа с данными и достоверность AI	Принимает ответ AI как готовый, не проверяет факты; загружает данные без подготовки.	Понимает, что AI ошибается: сверяет ключевые цифры, даты, имена; даёт модели контекст (документ, роль).	Системно готовит вход (чистит, структурирует, подаёт источники), проверяет ответы по первоисточникам, знает, где AI надёжен, а где нужна верификация; уверенно работает с длинным контекстом.	Проектирует процесс подготовки данных и регулярные тесты на достоверность; строит многоуровневую верификацию (кросс-проверка моделей, human-in-the-loop) и задаёт стандарты качества данных для организации.

ML/AI Team Lead

Техническое лидерство и развитие AI-команды 6–12 человек. Проектирование и владение ML-системами в production. Трансляция исследовательских прототипов в надёжные продуктовые решения.

Техническое лидерство

Компетенция	Basic	Good	Strong	Expert
Code Review & архитектурные сессии	Проводит code review по чеклисту, комментирует очевидные проблемы	Самостоятельно проводит архитектурные сессии, формирует инженерную культуру	Выстраивает системный процесс review и менторинга для команды 6–10 чел.	Формирует технические стандарты на уровне нескольких команд, обучает лидов
Менторинг и развитие команды	Проводит 1-on-1, даёт техническую обратную связь по запросу	Ведёт планы развития для каждого инженера, помогает с грейдовым ростом	Формирует карьерные треки для ML-инженеров, удерживает сеньоров	Создаёт систему технического онбординга и обучения для AI-функции

ML-архитектура и Production

Компетенция	Basic	Good	Strong	Expert
ML-пайплайны в production	Деплоит модели по готовым паттернам, настраивает базовый мониторинг	Проектирует надёжные пайплайны: ingestion, препроцессинг, деплой, откаты	Управляет полным lifecycle ML-систем: от POC до production с SLA	Определяет архитектурную стратегию ML-платформы, стандарты для всей организации
Выбор модельного стека	Использует знакомые open-source модели, аргументирует базовые выборы	Обоснованно выбирает между open-source LLM с учётом качества, стоимости и доступности	Формирует технические решения по всему модельному стеку (LLM, VLM, специализированные)	Определяет технологический радар по моделям, следит за SOTA и адаптирует под продукт
GPU-инфраструктура	Запускает задачи на GPU, понимает базовые метрики утилизации	Управляет GPU-кластером, оптимизирует шедулинг и cost-per-request	Проектирует GPU-инфраструктуру для высоконагруженных сервисов, FinOps-подходы	Определяет стратегию GPU-парка, мультирегиональность, autoscaling, бюджетирование
Observability ML-систем	Настраивает базовое логирование, читает метрики в Grafana	Строит систему трейсинга генераций до промпта, алертинг деградации качества	Проектирует полноценный observability стек: метрики, трейсы, cost-per-request	Формирует стандарты observability для всех ML-компонентов организации

Агентные системы и LLM

Компетенция	Basic	Good	Strong	Expert
Проектирование агентных систем	Знает ReAct-паттерн, реализует простых агентов с function calling	Проектирует сценарные LLM-ядра: планировщики, RAG, оркестрация	Строит надёжные мультиагентные системы с управлением памятью и state recovery	Определяет корпоративный фреймворк агентных систем, SLA, аудит, безопасность
Python + Go	Уверенный Python: типизация, async/await, ООП, unit-тесты	Проектирует высококачественный Python-код для ML-сервисов, базовый Go	Разрабатывает на Python и Go продакшн-сервисы с учётом производительности	Определяет стандарты разработки AI-сервисов, code review, tech debt management

AI Architect

Системное проектирование AI-решений на суверенной инфраструктуре. Интеграция моделей в высоконагруженные среды. Архитектура данных и инференс-пайплайнов.

Архитектура AI-систем

Компетенция	Basic	Good	Strong	Expert
LLM/VLM архитектуры	Понимает архитектуры трансформеров, методы дообучения, RAG-паттерны	Проектирует сквозную архитектуру: источники данных → пайплайны → UI с учётом SLA	Обоснованно выбирает модельный стек под бизнес-задачи, защищает перед СТО	Проектирует новые архитектурные паттерны, публикует решения, является экспертом рынка
Интеграция LLM в микросервисы	Оборачивает модель в REST API, деплоит в Docker	Разрабатывает паттерны интеграции LLM в микросервисную архитектуру	Проектирует корпоративные интеграции с учётом 152-ФЗ, ИБ-требований	Формирует архитектурные стандарты для всей AI-платформы компании
RAG и агентные фреймворки	Реализует базовый RAG (чанкинг → индекс → поиск → генерация)	Строит гибридный поиск (dense+sparse), reranking, eval-фреймворки	Проектирует корпоративные RAG-системы с incremental sync, multi-source, дедупликацией	Определяет архитектуру enterprise RAG-платформы, руководит кросс-командной разработкой

LLMOps / Оптимизация

Компетенция	Basic	Good	Strong	Expert
Оптимизация инференса	Понимает квантизацию (FP16, INT8), экспортирует в ONNX	Применяет AWQ, GGUF, GPTQ, настраивает vLLM/TGI/Triton для production	Оптимизирует inference-pipeline: KV-cache, speculative decoding, batching	Проектирует inference-стратегию компании: аппаратная архитектура, FinOps, SLA
Evaluation-фреймворки	Знает базовые метрики (BLEU, ROUGE, F1), запускает готовые скрипты	Строит eval-фреймворки с LLM-as-judge, RAGAS, автоматические тесты	Проектирует системы мониторинга качества в production: дрейф, деградация, алерты	Формирует корпоративный стандарт eval всех AI-продуктов, культуру data-driven решений
Суверенная инфраструктура	Работает с VK Cloud, Yandex Cloud на базовом уровне	Уверенно деплоит ML-нагрузки на российских облаках и on-premise GPU-кластерах	Проектирует AI-системы с учётом требований импортозамещения и регуляtorики	Является экспертом по суверенному AI-стеку, формирует стандарты для отрасли

ML Engineer

Разработка и интеграция LLM-систем. Проектирование RAG и агентных архитектур. Итеративное улучшение качества через метрики и эксперименты. Быстро собирает рабочий PoC и доводит до production.

LLM Engineering

Компетенция	Basic	Good	Strong	Expert
LLM-инженерия	Использует LLM через API, составляет простые промпты, понимает провайдеров	Строит цепочки промптов (few-shot, CoT), версионировует промпты, оптимизирует стоимость	Разрабатывает RAG-архитектуры, LangGraph, Tool Calling, мониторинг LLM в prodaction	Создаёт корпоративные multi-agent системы, LLM-шлюзы, маршрутизацию, guardrails
Разработка AI-агентов	Знает ReAct-паттерн, создаёт простых агентов с function calling	Строит агентов с планированием, историей диалога, отладкой через LangSmith/LangFuse	Проектирует мультиагентные системы с иерархией, разделением ответственности, fault tolerance	Определяет корпоративный фреймворк AI-агентов, SLA, безопасность, ROI-трекинг
RAG Development	Реализует базовый RAG: чанкинг → индексация → поиск → генерация	Гибридный поиск (BM25 + семантика), re-ranking, оценка через RAGAS, LLM-as-judge	Решает сложные RAG-кейсы: multi-hop, conflicting knowledge, мониторинг в prodaction	Определяет архитектуру корпоративной RAG-платформы, руководит cross-team разработкой

ML / NLP

Компетенция	Basic	Good	Strong	Expert
ML / NLP	Использует готовые модели из HuggingFace, применяет sentence-transformers	Дообучает BERT-подобные модели (fine-tuning, PEFT, LoRA/QLoRA) для downstream-задач	Применяет дистилляцию, квантование, мультиязычные модели, воспроизводит SOTA-статьи	Разрабатывает кастомные архитектуры, определяет стратегию AI-исследований, публикует результаты
AI Evaluation	Знает базовые метрики (точность, F1, BLEU, ROUGE), проверяет вручную	Создаёт датасеты и скрипты автоматической оценки, LLM-as-judge, A/B-тесты	Внедряет непрерывную оценку в CI/CD, дашборды метрик, измерение бизнес-эффекта	Определяет корпоративный стандарт оценки AI-продуктов, культуру data-driven решений

Разработка и MLOps

Компетенция	Basic	Good	Strong	Expert
Python Backend	Создаёт простые FastAPI-приложения с Docker, базовый CRUD	Разрабатывает production-ready сервисы: asynс, Redis, тесты, OpenAPI	Проектирует масштабируемые микросервисы: Kubernetes, Kafka, батчинг инференса	Определяет архитектуру backend-платформы: DDD, event sourcing, стандарты команд
MLOps	Оборачивает модель в Docker, запускает локально через FastAPI	Настраивает vLLM/TGI/BentoML, ведёт реестр моделей (MLflow, W&B), базовый мониторинг	Деплоит в Kubernetes с autoscaling, CI/CD для моделей, мониторинг data drift	Строит production-платформу: GPU-оркестрация, streaming, изоляция клиентов, SLA 99.9%+
System Design	Понимает API → сервис → БД, базовые требования к производительности	Проектирует микросервис с учётом ML-нагрузок: батчинг, кэширование, пики инференса	Проектирует end-to-end пайплайны с observability, балансирует trade-offs	Формирует архитектурную стратегию AI-направления: стандарты, шаблоны RAG/агентов, безопасность

ML Researcher

Воспроизводимые научные исследования. Высокопроизводительные вычисления (CUDA, distributed training).
Преобразование математического описания из препринтов arXiv в обучаемые тензорные пайплайны.

Машинное обучение и нейросети

Компетенция	Basic	Good	Strong	Expert
Машинное обучение	Знает матанализ, линейную алгебру, применяет классические методы (деревья, ансамбли, бустинг)	Применяет регуляризацию, ансамблирование, кросс-валидацию, интерпретирует результаты	Знает LIME/SHAP, временные ряды, ранжирование, anomaly detection, выбирает метод под нестандартную задачу	Разрабатывает новые ML-методы, формирует стандарты команды, обучает продвинутым подходам
Нейронные сети	Знает математику нейрона, backprop, реализует FFNN на PyTorch по шаблону	Применяет CNN, LSTM, использует HuggingFace Transformers, строит кастомный training loop	Знает трансформеры, attention, GAN, LoRA/QLoRA, обучает диффузионные модели, VAE	Проектирует нестандартные архитектуры, знает Mamba/RWKV, формирует технологический радар
Обработка текстов (NLP)	Применяет токенизацию, TF-IDF, word2vec, решает базовые NLP-задачи готовыми моделями	Fine-tuning BERT/GPT/T5, промпт-инжиниринг, zero-shot и few-shot learning	SFT, RLHF, DPO, Reward model, PEFT, строит RAG-системы, мультиагентные системы	Определяет NLP-стратегию продукта, формирует исследовательскую повестку, публикует результаты

Инфраструктура обучения

Компетенция	Basic	Good	Strong	Expert
Фреймворки (PyTorch / HuggingFace)	Базовый PyTorch: тензоры, training loop, загрузка HuggingFace моделей	Кастомные архитектуры, LoRA/QLoRA/PEFT, PyTorch Lightning, structured code	Mixed Precision, Gradient Accumulation, DistributedDataParallel, профилирование	Масштабирование на Multi-node, проектирует внутренние библиотеки-обёртки для команды
MLOps-инфраструктура (ClearML / W&B)	Настраивает Python-окружение, Docker, базовое логирование в TensorBoard	Настраивает ClearML/W&B, Hydra, запускает DDP, ранняя остановка, gradient clipping	Продвинутые возможности ClearML: HPO, Expert Parallel, оптимизация GPU	Проектирует MLOps-инфраструктуру для команды, создаёт стандартизированные библиотеки
Рантаймы инференса	Знает ONNX, экспортирует модели, замеряет latency, использует ollama	Квантизация (FP16, INT8), vLLM, ONNX Runtime для production	TensorRT, Triton Inference Server, PagedAttention, настраивает параметры serving	KV-cache оптимизация, speculative decoding, проектирует inference-архитектуру для high-load

R&D и данные

Компетенция	Basic	Good	Strong	Expert
Чтение публикаций и реализация	Читает обзорные статьи, запускает код из репозиториев по туториалам	Самостоятельно воспроизводит SOTA, сравнивает подходы экспериментально	Реализует метод без официального кода, критически оценивает применимость к бизнес-задачам	Комбинирует идеи из публикаций, формирует технологический радар, публикует на топ-конференциях
Разметка данных	Работает с Label Studio/Доссано как исполнитель, знает форматы аннотаций	Составляет инструкции для ассессоров, готовит данные, базовые метрики согласия	Active Learning, Dataset Cartography, проектирует многоэтапные пайплайны разметки	Разрабатывает методологию для новых задач, управляет экономикой разметки, формирует стратегию датасетов
Верификация качества моделей	Знает accuracy, Precision, Recall, F1, корректно разделяет train/val/test	ROC-AUC, BLEU, ROUGE, BERTscore, проводит Error Analysis, анализ соответствия задаче	Методы валидации RAG и агентных систем, ablation study, статистическая значимость	Формирует стандарты верификации для всех NLP-компонентов, создаёт фреймворки автотестирования

Data Scientist

Исследователь и инженер данных в одном лице. Превращает сырые данные в проверяемые гипотезы, рабочие прототипы и production-модели. Ключевая задача — найти в данных скрытые закономерности и решить конкретные бизнес-задачи.

Анализ данных и статистика

Компетенция	Basic	Good	Strong	Expert
Разведочный анализ данных (EDA)	Строит базовые визуализации, вычисляет описательные статистики	Самостоятельно проводит EDA: пропуски, аномалии, распределения, корреляции	Выявляет неочевидные паттерны, формирует гипотезы, документирует инсайты	Формирует стандарты EDA в команде, выстраивает процессы мониторинга качества данных
Feature Engineering	Применяет базовые преобразования признаков (нормализация, кодирование)	Создаёт информативные признаки, устраняет data leakage, работает с временными рядами	Применяет автоматический feature selection (SHAP, RFE), создаёт нелинейные признаки	Разрабатывает feature store для команды, формирует стандарты работы с признаками
A/B-тесты и эксперименты	Понимает концепцию A/B-тестов, базовые статистические тесты	Самостоятельно проектирует A/B-тесты, интерпретирует результаты, p-value	Проводит сложные эксперименты (CUPED, стратификация), оценивает бизнес-эффект	Формирует культуру экспериментирования, разрабатывает фреймворк A/B-тестов для продуктов

Моделирование

Компетенция	Basic	Good	Strong	Expert
Классический ML	Применяет базовые алгоритмы (логрег, деревья) с помощью sklearn	Строит ансамбли, оптимизирует гиперпараметры, работает с дисбалансом классов	Выбирает метод под нестандартную задачу, анализирует ошибки, работает с нелинейными задачами	Разрабатывает новые подходы, формирует стандарты моделирования для команды
Deep Learning для DS-задач	Использует предобученные модели для базовых задач (классификация, NER)	Fine-tuning BERT-подобных моделей, построение эмбеддингов, семантический поиск	Применяет LLM для DS-задач: табличные данные, генерация признаков, аугментация	Определяет стратегию применения DL в DS-продуктах, обучает команду
Мониторинг моделей (Prod)	Отслеживает базовые метрики модели в production	Настраивает алерты на деградацию, отслеживает data drift с помощью инструментов	Проектирует систему мониторинга: concept drift, model drift, автоматические retraining триггеры	Определяет стандарты операционного ML для компании, SLA по качеству моделей

Инженерия и инструменты

Компетенция	Basic	Good	Strong	Expert
Python и ML-библиотеки	pandas, numpy, scikit-learn на уровне базового использования, простые скрипты	Уверенный Python: ООП, типизация, тесты, PyTorch/ TensorFlow для практических задач	Проектирует переиспользуемые ML-компоненты, multi-processing, профилирование	Разрабатывает внутренние DS-фреймворки, управляет техдолгом, определяет стандарты
SQL и работа с данными	Пишет простые SELECT-запросы, работает с таблицами по инструкции	JOIN, агрегации, оконные функции, подготовка датасетов из БД	Оптимизирует сложные запросы, работает с большими объёмами, проектирует датасеты	Определяет стандарты работы с данными, проектирует data pipelines для DS-команды
MLOps-инструменты	Знает концепции MLflow, DVC, Weights & Biases на базовом уровне	Трекает эксперименты, версионизирует модели и данные, настраивает базовый CI/CD	Строит воспроизводимые ML-пайплайны с трекингом, автоматизирует retraining	Проектирует ML-платформу для DS-команды, определяет MLOps-стандарты организации

AI Product Manager

Определяет, запускает и масштабирует AI-продукты. Управляет кросс-функциональной командой. Понимает ограничения AI, стоимость вычислений и этические риски.

Продуктовое исследование и стратегия

Компетенция	Basic	Good	Strong	Expert
Product Discovery & Research	Участвует в исследованиях под руководством ментора, знает JTBD, Lean Canvas	Самостоятельно проводит исследования рынка, валидирует гипотезы, формирует ценностное предложение	Выстраивает процесс непрерывного discovery, синтезирует инсайты, влияет на стратегию	Формирует культуру product discovery на уровне организации, наставляет PM'ов
Customer Development	Участвует в интервью как наблюдатель, ведёт заметки и систематизирует обратную связь	Самостоятельно проводит проблемные и решенческие интервью, работает с B2B-клиентами	Проводит сложные B2B-интервью с C-level, выстраивает систему регулярного CustDev	Определяет методологию CustDev, ведёт стратегические интервью, формирует партнёрства
Roadmap & Prioritization	Понимает структуру roadmap, помогает в приоритизации по RICE/ICE	Самостоятельно формирует roadmap, балансирует быстрые результаты и долгосрочное развитие	Формирует стратегический roadmap на 6–12 мес., управляет зависимостями, презентует C-level	Определяет долгосрочное видение продуктовой линейки, координирует несколько roadmap'ов

Метрики, аналитика и GTM

Компетенция	Basic	Good	Strong	Expert
Метрики и аналитика данных	Знает DAU, MAU, retention, conversion, строит простые отчёты	Определяет дерево метрик, настраивает отслеживание, работает с data-аналитиками	Выстраивает систему аналитики, формулирует B2B-метрики, проводит A/B-тестирование	Определяет стандарты аналитики, является экспертом по метрикам AI/ML-продуктов
Go-to-Market стратегия	Понимает базовые принципы GTM, участвует в подготовке материалов	Готовит план запуска, формулирует позиционирование, участвует в ценовой политике	Разрабатывает GTM-стратегию для внешнего рынка РФ: сегменты, каналы, ценообразование	Формирует GTM на уровне портфеля, определяет стратегические партнёрства
Pre-sales и работа с клиентами	Участвует в демо как наблюдатель, фиксирует требования клиентов	Самостоятельно проводит демо, собирает и приоритизирует требования заказчиков	Ведёт pre-sales для крупных B2B, управляет пилотами, формирует отраслевые решения	Определяет стратегию pre-sales, ведёт переговоры с C-level, является лицом продукта

Техническая грамотность и инструменты

Компетенция	Basic	Good	Strong	Expert
Техническая грамотность (AI/ML)	Базовое понимание AI/ML: модели, датасеты, обучение, продукты компании	Понимает LLM, NLP-пайплайны, ASR/TTS, общается с ML-инженерами, знает метрики качества	Глубокое понимание AI-архитектур, формулирует технические требования, следит за трендами	Экспертное понимание возможностей AI, определяет технологическую стратегию продуктов
Работа с LLM и AI-инструментами	Активно использует LLM для ежедневных задач, грамотно формулирует запросы	Применяет LLM для product research, промпт-инжиниринг, автоматизация рутины PM	Выстраивает рабочие процессы с AI, создаёт промпты для типовых задач, интегрирует в delivery	Формирует стратегию AI-инструментов, создаёт best practices, обучает команды
PRD и управление delivery	Пишет user stories по шаблону, ведёт задачи в Jira, знает основы Scrum	Самостоятельно пишет PRD, управляет backlog и спринтами, координирует команду	Пишет комплексные PRD для AI-продуктов, оптимизирует delivery, управляет зависимостями	Разрабатывает стандарты документации и delivery для продуктового направления

Technical PM

Обеспечивает надёжную и экономически эффективную доставку AI-решений от прототипа до production.
Выстраивает прозрачные процессы, контролирует технические риски, поддерживает инженерные стандарты.

Управление проектами

Компетенция	Basic	Good	Strong	Expert
Планирование и контроль	Ведёт задачи в Jira, отслеживает статусы, заполняет шаблоны документации	Самостоятельно планирует спринты, управляет рисками, готовит отчётность для стейкхолдеров	Управляет программой проектов: сроки, бюджет, ресурсы, зависимости между командами	Формирует РМО-стандарты для AI-направления, наставляет менеджеров
Управление требованиями	Фиксирует требования по шаблону, разбивает задачи на атомарные истории	Составляет ТЗ и ФТ для ML-компонентов, согласовывает с DS/ML командой и бизнесом	Полный цикл: от бизнес-требований до технических спецификаций, управление изменениями	Разрабатывает стандарты документирования для AI-систем на уровне организации
Управление рисками	Ведёт реестр рисков по шаблону, эскалирует проблемы по цепочке	Самостоятельно выявляет и митигирует риски ML-проектов (данные, качество, сроки)	Управляет рисками программы AI-трансформации: технические, организационные, регуляторные	Формирует корпоративный подход к управлению рисками AI-проектов

Технические знания для PM

Компетенция	Basic	Good	Strong	Expert
Понимание AI/ML workflow	Знает этапы ML-pipeline: данные → обучение → деплой → мониторинг	Понимает специфику AI-проектов: неопределённость качества, итеративность, зависимость от данных	Ставит реалистичные оценки для ML-задач, управляет ожиданиями стейкхолдеров	Является экспертом по оценке AI-проектов, определяет методологию для организации
Технический кругозор (Python / SQL / API)	Понимает принципы программирования, читает код, базовый Python/SQL	Уверенный SQL, базовый Python (pandas, sk!), понимает REST API и микросервисы	Понимает LangChain/LlamaIndex, MLOps-инструменты, архитектуру AI-систем на уровне PM	Может самостоятельно прототипировать решения (vibe-coding), определяет техдолг

Процессы и коммуникации

Компетенция	Basic	Good	Strong	Expert
Agile / Scrum / Kanban	Участвует в церемониях, ведёт backlog, знает основы Scrum	Самостоятельно проводит планирования, ретро, управляет velocity команды	Адаптирует методологию под специфику AI-проектов, внедряет лучшие практики	Формирует стандарты delivery для AI-направления, масштабирует на несколько команд
Кросс-командное взаимодействие	Коммуницирует в рамках своей команды, эскалирует блокеры	Самостоятельно выстраивает взаимодействие между ML, backend, продуктом	Управляет зависимостями AI-программы: Research, Engineering, Data, Product, Infra	Является центром коммуникаций для крупной AI-трансформации, работает с C-level
Документация (ТЗ / Устав / ФТ)	Заполняет шаблоны документации под руководством, ведёт Confluence	Самостоятельно разрабатывает ТЗ, ФТ, устав проекта для AI-систем	Пишет комплексную документацию AI-систем: архитектура, данные, безопасность, compliance	Разрабатывает стандарты документирования AI-проектов для организации

AI Тренер

Создаёт ground-truth данные для этапов SFT и RLHF. В отличие от промпт-инженера работает с замороженной моделью — AI-тренер создаёт датасеты, которые меняют веса модели.

Создание обучающих данных

Компетенция	Basic	Good	Strong	Expert
Аннотация и ранжирование ответов	Работает с инструментами разметки по инструкции, оценивает ответы по готовым критериям	Самостоятельно ранжирует ответы модели, выявляет нюансы качества, обосновывает выборы	Создаёт сложные датасеты предпочтений (preference data) для RLHF, обучает ассессоров	Разрабатывает методологию разметки для принципиально новых задач, формирует стандарты
Создание эталонных ответов (SFT)	Пишет простые эталонные ответы по шаблону в заданных доменах	Создаёт высококачественные эталоны для разных типов задач: инструкции, диалоги, рассуждения	Разрабатывает эталоны для сложных задач: математика, код, многошаговые рассуждения	Определяет стратегию сбора SFT-данных, управляет командой тренеров, контролирует качество
Генерация и редактирование текстов	Пишет чистые, грамотные тексты по шаблонам, следует стилевым указаниям	Создаёт разнообразные тексты в разных стилях, доменах, форматах без шаблонов	Пишет тексты экспертного уровня в сложных доменах (медицина, право, технологии)	Является редактором и методологом для команды тренеров, формирует гайды качества

Оценка и контроль качества

Компетенция	Basic	Good	Strong	Expert
Фактчекинг	Проверяет очевидные фактические ошибки, использует популярные поисковики	Находит первоисточники, проверяет цифры, даты, цитаты по академическим источникам	Верифицирует сложные факты в специализированных доменах, выявляет тонкие фактические ошибки	Разрабатывает стандарты фактчекинга для команды, строит базы доверенных источников
Этический контроль ответов	Идентифицирует очевидно вредный контент по заданным критериям	Оценивает контент по многомерным этическим критериям, выявляет тонкие нарушения	Разрабатывает политики этической оценки для специфических доменов и аудиторий	Формирует этические стандарты для всей AI-разработки, консультирует по сложным кейсам
Критическое мышление и логика	Выявляет очевидные логические ошибки в тексте	Анализирует аргументацию, выявляет fallacies, проверяет согласованность утверждений	Оценивает сложные рассуждения в нестандартных доменах, работает с противоречивыми источниками	Обучает команду критическому анализу, разрабатывает методы оценки качества рассуждений

Предметные знания и языки

Компетенция	Basic	Good	Strong	Expert
Работа с LLM-инструментами	Использует ChatGPT/Claude для ежедневных задач, понимает возможности и ограничения	Анализирует паттерны ошибок модели, понимает связь промптов и качества вывода	Понимает связь обучающих данных с поведением модели, может предсказывать эффект данных	Является экспертом по pipeline SFT/RLHF, консультирует исследователей по дизайну датасетов
Грамотность (русский / целевой язык)	Пишет без орфографических и пунктуационных ошибок, следует нормам языка	Владеет стилистикой, адаптирует регистр под задачу, исправляет чужие тексты	Экспертный уровень: знает тонкие нормы языка, стилистику разных жанров, редкие правила	Является языковым экспертом команды, обучает тренеров грамотности и стилистике
Быстрое освоение новых доменов	Осваивает базовые понятия нового домена за 1–2 дня по инструкции	Самостоятельно изучает сложные домены (медицина, право, технологии) для задач разметки	Экспертно ориентируется в 2–3 специализированных доменах, понимает специфику каждого	Формирует учебные материалы по доменам для команды, определяет источники знаний

Prompt Engineer

Профессия на пересечении гуманитарных и технических дисциплин. Создаёт, тестирует и оптимизирует промпты для LLM. Работает с замороженной моделью через интерфейс — в отличие от AI-тренера, который меняет веса.

Промпт-инжиниринг

Компетенция	Basic	Good	Strong	Expert
Техники промпт-инжиниринга	Пишет базовые промпты, понимает роли (system/user/assistant), использует примеры	Применяет few-shot, CoT, zero-shot, структурированный вывод (JSON, Pydantic)	Применяет Tree of Thought, self-consistency, ReAct, оптимизирует по качеству и стоимости	Разрабатывает корпоративные методологии промптинга, проводит обучение команд
Устранение галлюцинаций	Выявляет очевидные галлюцинации, базовые методы верификации ответов	Применяет техники снижения галлюцинаций: grounding, ограничения в промпте, RAG	Системно анализирует паттерны галлюцинаций, разрабатывает eval-фреймворки для мониторинга	Определяет стратегию борьбы с галлюцинациями для продукта, формирует best practices
Разработка цепочек промптов (CoT)	Составляет линейные цепочки промптов по шаблону	Проектирует многошаговые цепочки с ветвлением, управлением контекстом и состоянием	Создаёт сложные оркестрированные пайплайны с несколькими LLM и инструментами	Является архитектором промпт-систем, формирует корпоративные стандарты цепочек

Тестирование и оптимизация

Компетенция	Basic	Good	Strong	Expert
Итерация (промпт-хитинг)	Тестирует промпты вручную, фиксирует результаты в таблице	Систематически итерирует промпты, ведёт эксперименты с метриками, версиониров	Строит автоматизированные eval-пайплайны, сравнивает версии статистически	Внедряет continuous prompt evaluation в CI/CD, формирует культуру экспериментирования
Анализ результатов моделей	Оценивает качество ответов по заданным критериям, фиксирует паттерны ошибок	Декомпозирует ошибки по типам, анализирует влияние параметров (temp, top-k, top-p)	Строит системный анализ качества с метриками, LLM-as-judge, понимает внутреннее устройство	Является экспертом по оценке LLM-качества, консультирует исследователей и PM'ов

Техническая база и интеграции

Компетенция	Basic	Good	Strong	Expert
Python и работа с данными	Базовый Python: скрипты, pandas, re, numpy, чтение JSON/CSV	Уверенный Python: парсинг данных, автоматизация, вызов LLM через API	Проектирует масштабируемые скрипты для промпт-тестирования, работает с LangChain/LlamaIndex	Разрабатывает внутренние инструменты оценки промптов, определяет tooling для команды
Работа с LLM через API	Вызывает LLM API через готовые примеры, понимает параметры запроса	Самостоятельно работает с OpenAI/Anthropic/open-source API, управляет контекстом	Интегрирует несколько LLM-провайдеров, настраивает streaming, управляет токенами	Проектирует LLM-шлюзы и middleware для корпоративных систем
Промпт-библиотеки для бизнеса	Создаёт простые шаблоны для конкретных задач по инструкции	Разрабатывает переиспользуемые промпт-шаблоны для бизнес-кейсов, документирует	Строит корпоративную промпт-библиотеку с версионированием, eval и governance	Определяет стратегию управления промптами для организации, формирует стандарты
Понимание LLM (токенизация, контекст)	Знает что такое токены, контекстное окно, температура	Понимает токенизацию, top-k/top-p, влияние system prompt на поведение модели	Глубоко понимает механику LLM: attention, positional encoding, inference-time compute	Является экспертом по архитектуре LLM, консультирует команды по применению

Выводы и применение

1. Насколько эти матрицы применимы — и почему?

Применимы как базовая карта ориентиров

Матрицы покрывают компетенции, которые востребованы в большинстве AI-команд на российском рынке: от базовой разработки до архитектуры production-систем и research-экспертизы. Уровни Basic → Expert отражают реальный профессиональный путь в каждой роли и могут быть использованы как есть на старте — без длительной кастомизации.

Применимы для структурирования найма

Без матрицы оценка кандидата субъективна: каждый интервьюер проверяет то, что считает важным лично. Матрица создаёт общий язык между рекрутером, нанимающим менеджером и техническим интервьюером — и делает оценку воспроизводимой. Это особенно важно для AI-ролей, где разрыв в понимании профиля между HR и техническим руководителем традиционно высок.

Применимы для грейдинга и ревью

Матрица даёт объективную основу для performance review и карьерных разговоров: сотрудник видит, где он находится сейчас, что конкретно нужно сделать для следующего уровня, и какие компетенции у него избыточны или недостаточны для текущей роли. Это переводит разговор из «оцениваю тебя субъективно» в «вот конкретные критерии».

Ограничение: матрица — карта, а не территория. Матрица описывает компетенции, но не заменяет суждение. Человек с уровнем Strong по всем компетенциям не автоматически сильнее того, у кого Expert в двух ключевых и Basic в периферийных. Контекст задачи, soft skills, культурное соответствие — вне матрицы. Используйте её как структуру разговора, а не как формулу найма.

2. Универсальные ли они — или нужна адаптация под компанию?

Короткий ответ: базовая структура универсальна, конкретное наполнение — нет. Матрицы отражают рыночный стандарт AI-компетенций в 2026 году, но не учитывают специфику вашего стека, домена и стадии развития команды. Адаптация обязательна.

Что оставить как есть

- Четырёхуровневая структура Basic → Expert
- Доменное деление компетенций
- Базовые технические требования (Python, ML-фреймворки, основы LLM)
- Управленческие компетенции для Lead и Head

Что адаптировать под компанию

- Конкретный технологический стек (ваши облака, БД, оркестраторы)
- Доменные знания (промышленность, финтех, медицина, ритейл)
- Весовые коэффициенты: что важнее в вашем контексте
- Пороговые уровни для грейдов и найма

Как адаптировать — пошагово

- Проведите аудит текущего стека.** Какие инструменты, фреймворки и платформы реально используются? Добавьте их как конкретные критерии в компетенции — вместо общих «Docker, Kubernetes» пропишите версии и конфигурации, с которыми работает команда.
- Определите доменные знания.** Финтех требует понимания AML и скоринга. Промышленность — предиктивного обслуживания и компьютерного зрения на производственных данных. Добавьте домен как отдельный блок компетенций.
- Откалибруйте уровни на реальных людях.** Возьмите 2–3 сотрудника, которых считаете сильными в роли, и пройдите с ними по матрице. Где их реальный уровень не совпадает с описанием — описание нужно пересмотреть.
- Установите пороги для грейдов.** Не все компетенции одинаково важны. Определите 3–5 ключевых компетенций для каждой роли, которые являются hard requirement, и остальные — как желательные. Минимальный порог для грейда = Strong по ключевым.
- Пересматривайте раз в полгода.** AI-стек меняется быстрее большинства других технологических областей. Компетенция «знает GPT-3» через год уже не является дифференцирующей. Матрица без обновления устаревает быстрее, чем кажется.

3. Как развивать уровни компетенций в команде?

Basic	Good	Strong	Expert
Онбординг-программа, базовые курсы, работа по шаблонам под менторством. Горизонт перехода: 2–4 мес.	Самостоятельные задачи с ревью, разбор кейсов, участие в code review как ревьюируемый. Горизонт: 4–8 мес	Ведение архитектурных решений, менторство джунов, участие в найме, конференции и чтение статей. Горизонт: 6–18 мес.	Публичная экспертиза, open-source, публикации, формирование стандартов. Не растят в большинстве компаний — нанимают или удерживают.

Регулярный 1-on-1 по матрице раз в квартал. Руководитель и сотрудник совместно оценивают текущий уровень по каждой компетенции, фиксируют разрыв и договариваются о конкретных действиях на следующий квартал. Не «развивай коммуникацию» — а «проведи 3 архитектурные сессии и получи обратную связь».

Развивать в зоне ближайшего развития, не в зоне разрыва. Человек на уровне Basic по компетенции не прыгнет сразу на Expert. Цель — следующий уровень, не финальный. Попытка перепрыгнуть через ступень ведёт к демотивации и ощущению недостижимости.

Не выравнивать всё до Strong — это нормально. Реалистичная цель для большинства специалистов — Strong в 2–3 ключевых компетенциях роли и Good в остальных. Expert — это штучная история. Попытка вырастить Expert по всем компетенциям у всех — трата ресурсов и разочарование.

Инструменты роста для AI-специалистов: что работает. Для Basic→Good: курсы + задачи с ревью. Для Good→Strong: участие в реальных production-инцидентах, самостоятельные архитектурные решения, публичное объяснение своих решений команде. Для Strong→Expert: публикации, open-source, конференции, менторство других команд.

4. Как применять матрицу в найме и оценке кандидата?

ДО ИНТЕРВЬЮ

- 1. Определите целевой профиль: какой уровень по каким компетенциям нужен для данной вакансии прямо сейчас. Не «Strong везде», а конкретные приоритеты.
- 2. Составьте вопросы под каждую проверяемую компетенцию. Один вопрос — одна компетенция. Вопросы поведенческие («расскажите о случае, когда вы...») для soft и технические — для hard.
- 3. Договоритесь с техническим интервьюером, какие компетенции проверяет он, а какие — рекрутер или нанимающий менеджер. Без разделения — дублирование и пробелы одновременно.

ВО ВРЕМЯ ИНТЕРВЬЮ

- 4. Оценивайте ответ по шкале матрицы в реальном времени. Фиксируйте конкретные формулировки кандидата — они понадобятся при дебрифе с командой.
- 5. Если ответ неоднозначный — задайте уточняющий вопрос на следующий уровень. «Вы сказали, что строили RAG. Как вы решали проблему conflicting context при multi-hop вопросах?»
- 6. Не пытайтесь проверить всё. 3–5 ключевых компетенций на одно интервью — достаточно. Лучше глубоко проверить важное, чем поверхностно — всё.

ПОСЛЕ ИНТЕРВЬЮ — ДЕБРИФ ПО МАТРИЦЕ

Заполните scorecard

Каждый интервьюер отдельно проставляет уровни по своим компетенциям до совместного обсуждения. Это исключает эффект anchoring — когда первый голос определяет всех остальных.

Обсудите расхождения

Если оценки разошлись больше чем на один уровень — это точка для разговора, а не усреднения. Почему один видит Strong, а другой Basic? Возможно, проверялись разные аспекты.

Зафиксируйте разрывы

Где кандидат ниже целевого уровня — это не автоматический отказ. Это план развития: можно ли закрыть разрыв за 3–6 месяцев онбординга, или это hard requirement без компромисса?

Типичные ошибки при применении матрицы в найме:

- 1. Искать Expert по всем компетенциям (такого человека нет или он стоит как Head of AI).
- 2. Использовать матрицу как чеклист вместо карты — оценивать наличие слова, а не реального уровня.
- 3. Не договариваться заранее, кто что проверяет (в итоге Python проверяют трижды, а system design — никто).
- 4. Отказывать из-за низкого уровня в периферийных компетенциях при высоком в ключевых.

Итоговый вывод

Матрица компетенций — это не документ, а инструмент.

Её ценность определяется тем, как она используется.

Большинство компаний создают матрицы компетенций, кладут их в Confluence и забывают. Через год выясняется, что найм ведётся по интуиции, грейды выдаются по принципу «заслужил лояльностью», а оценка кандидата зависит от настроения интервьюера. Матрица работает только в одном случае — если она живёт в процессах: найма, ревью, планирования развития, разговора об офере. Без этого она остаётся красивым артефактом.

О применимости

Представленные матрицы отражают рыночный стандарт AI-компетенций на 2026 год и применимы как отправная точка для любой компании, которая строит или масштабирует AI-функцию. Они охватывают весь спектр ролей — от операционных (AI Trainer, Prompt Engineer) до стратегических (Head of AI) — и дают единый язык для оценки людей разных специализаций. Это особенно ценно в момент, когда рынок ещё не выработал общепринятых стандартов: что именно должен уметь «ML-инженер уровня Strong» — в разных компаниях понимают по-разному. Матрица снимает эту двусмысленность.

Об адаптации

Адаптация под компанию — не опция, а необходимость. Матрица, созданная для продуктовой tech-компании, не учитывает специфику промышленного AI, финтеха или государственных систем. Там свои данные, свои ограничения, свои регуляторные требования — и свои приоритеты в компетенциях. Компания, которая использует матрицу без адаптации, рискует нанимать правильных людей для чужой задачи. Правило простое: структура универсальна, наполнение — нет. Чем точнее матрица отражает реальные задачи и стек вашей команды, тем точнее она работает как инструмент найма и оценки.

О развитии уровней

Самая распространённая ошибка — использовать матрицу как аттестационный список, а не как карту роста. Реалистичная цель для большинства специалистов — Strong в ключевых компетенциях роли, Good в остальных. Expert — это штучная история: таких людей либо нанимают, либо удерживают особыми условиями, но не «выращивают» стандартными программами обучения. Развитие работает, когда есть три вещи одновременно: понятный следующий уровень (не финальная цель, а ближайший шаг), конкретные действия для его достижения, и регулярная обратная связь от руководителя о том, движется ли человек в нужном направлении. Без любого из трёх элементов матрица превращается в декорацию.

О применении в найме

Матрица компетенций меняет найм принципиально: вместо «нравится / не нравится» появляется структурированная оценка по конкретным критериям. Это снижает влияние субъективности, помогает сравнивать кандидатов на одинаковых основаниях и даёт чёткий аргумент при несогласии в команде интервьюеров. Но инструмент работает только при правильном использовании. Три условия, без которых матрица в найме не работает: во-первых, заранее договориться, кто какие компетенции проверяет — иначе одни блоки проверяются трижды, другие не проверяются вообще. Во-вторых, зафиксировать целевой профиль до начала поиска — без него каждый интервьюер оценивает «идеального» кандидата по своим критериям, и финальное решение становится результатом борьбы мнений, а не оценки. В-третьих, отличать hard requirement от nice-to-have — кандидат с Expert в ключевых компетенциях и Basic в периферийных сильнее кандидата с равномерным Good по всем. Матрица помогает это увидеть — если её правильно читать.

Главный вывод: матрица компетенций — это инвестиция, которая окупается не в момент создания, а в момент первого найма, первого карьерного разговора и первого performance review, проведённого по единым стандартам. Компании, выстроившие эту систему до того, как AI-команда выросла до 10+ человек, избегают хаоса несогласованных ожиданий, субъективных оценок и дорогостоящих ошибок найма, с которыми сталкиваются все, кто откладывал структуризацию на потом.

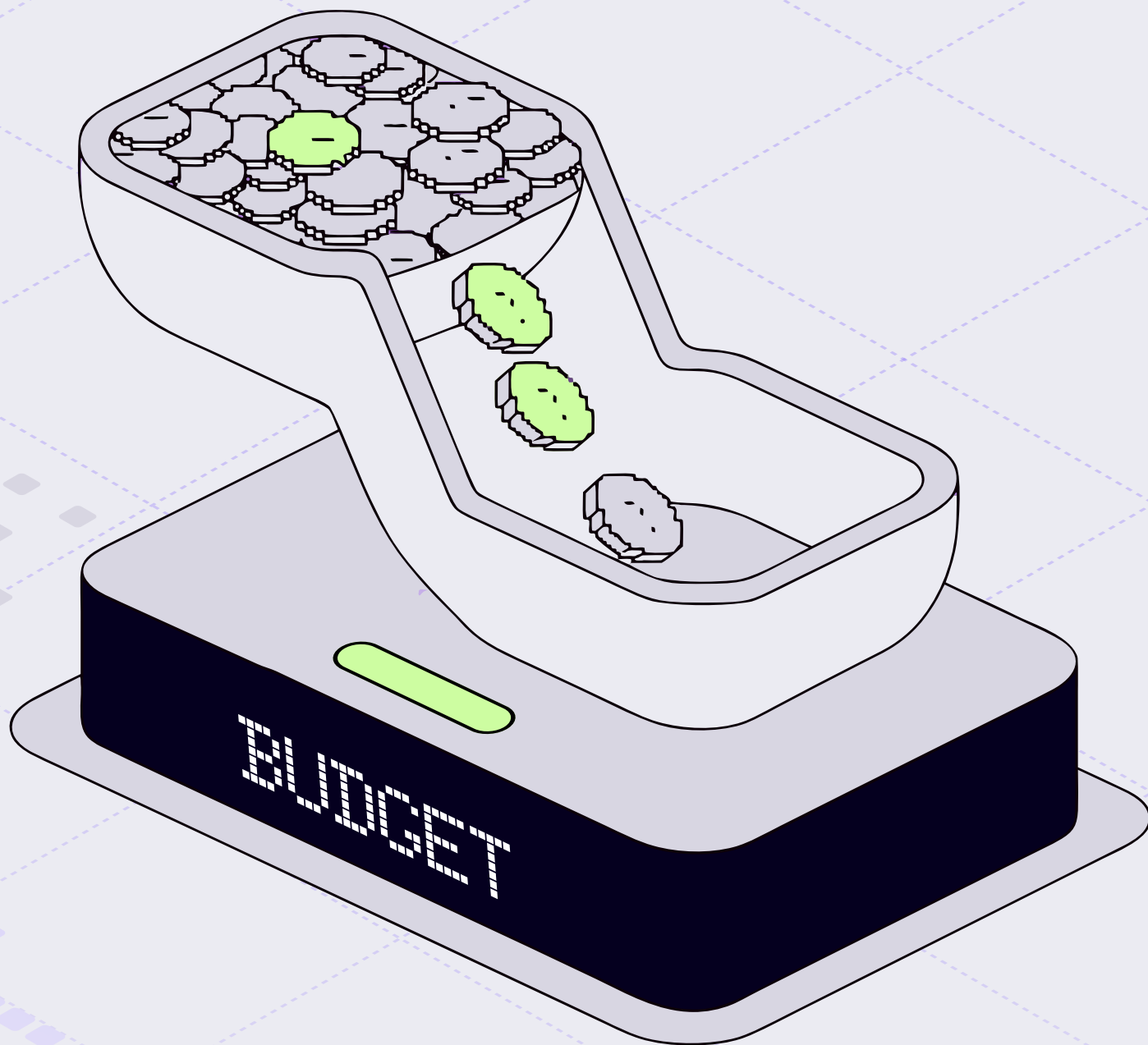


Деньги

Вилки зарплатных ожиданий рынка, оргструктуры команд, ФОТ-бюджет и выводы по экономике AI-направления.

В ЭТОМ РАЗДЕЛЕ

- 1 Вилки зарплатных ожиданий — рынок 2026
- 2 Оргструктуры: PoC и Production
- 3 ФОТ — бюджет команды
- 4 Выводы по ролям и ФОТу



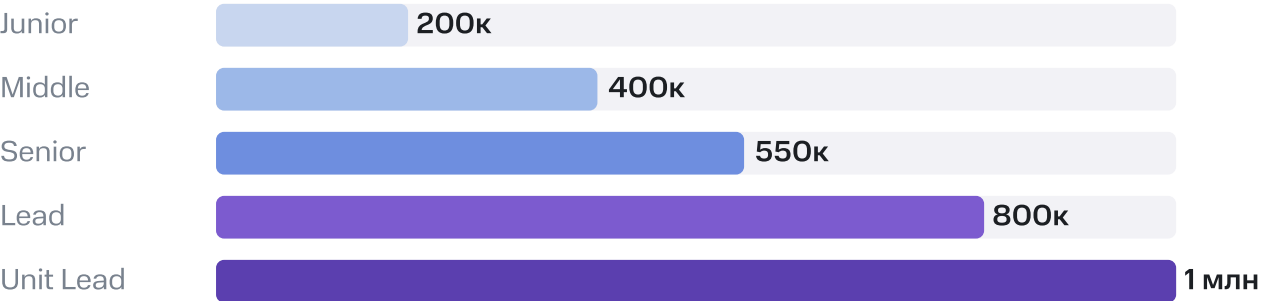
06

Зарплатные ожидания AI-специалистов по грейдам, март 2026

Все роли

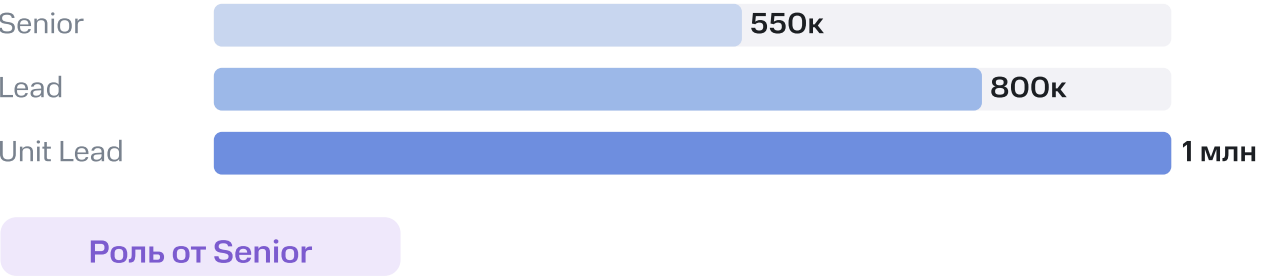
ML-инженерия

ML Engineer



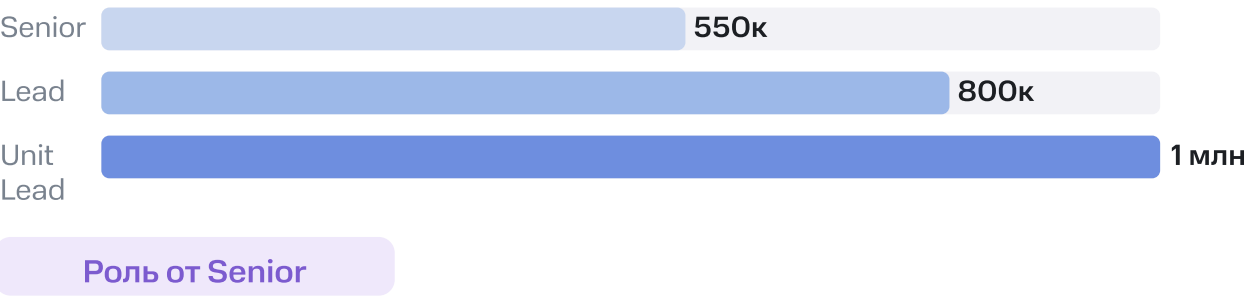
ML-инженерия

AI Архитектор



ML-инженерия

ML/AI Team Lead



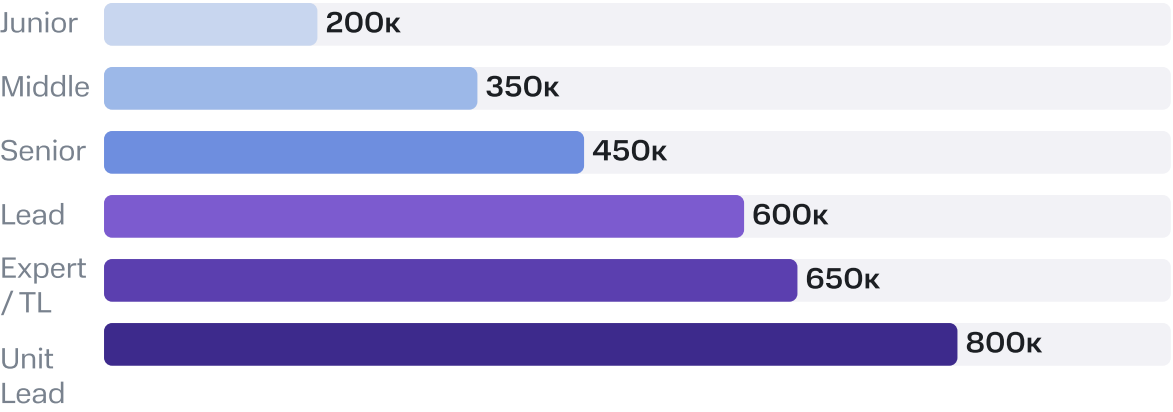
ML-инженерия

Head of AI / CAIO



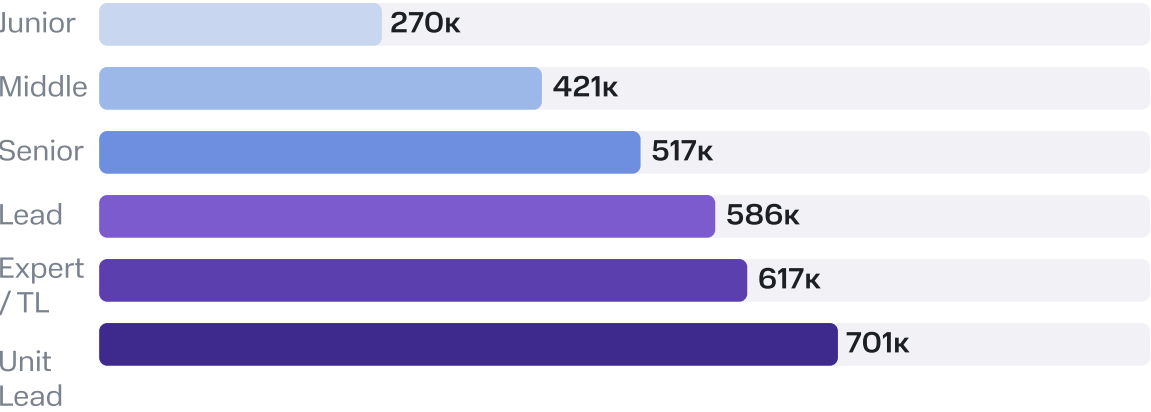
Исследователи

ML Разработчик-исследователь



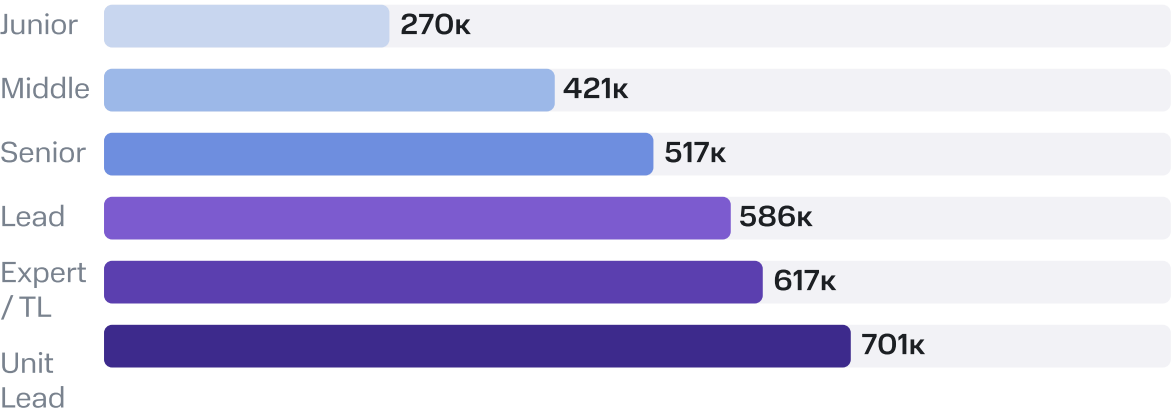
Продуктовый офис

AI Product Manager



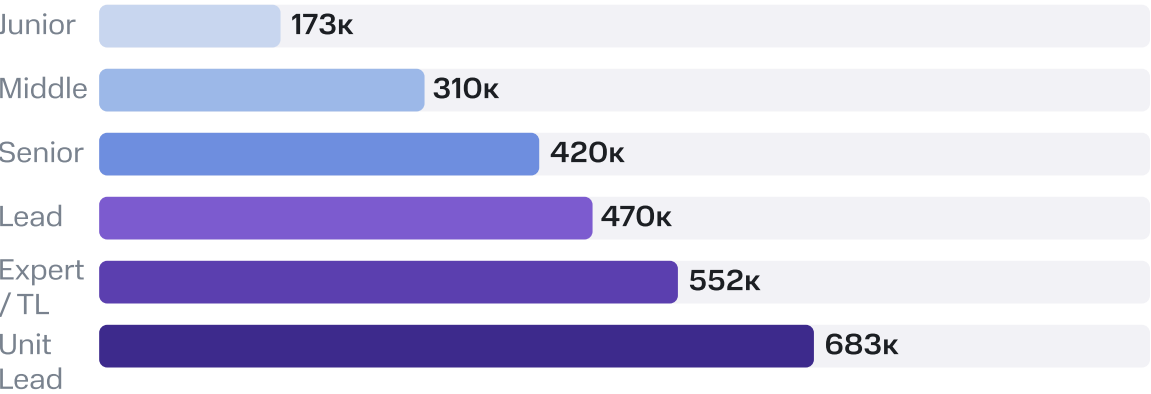
Продуктовый офис

Data Scientist



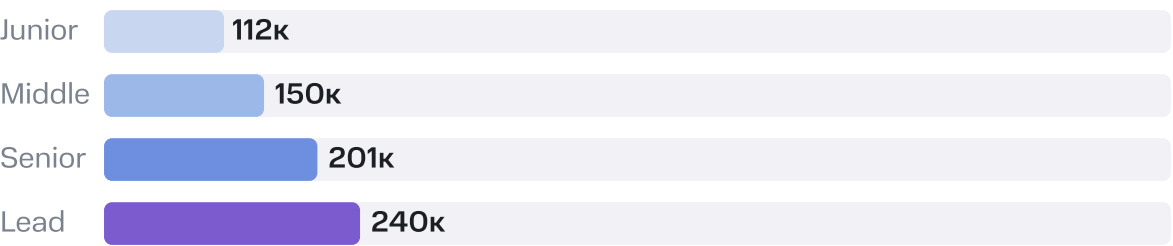
Проектный офис

Project Manager



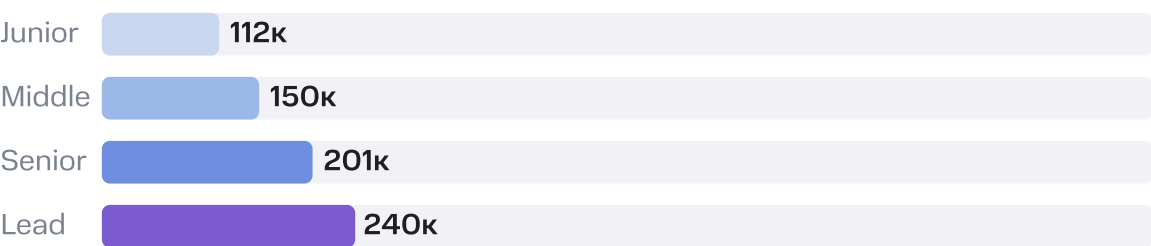
AI-тренеры / Промпт

Prompt Engineer



AI-тренеры / Промпт

AI Тренер

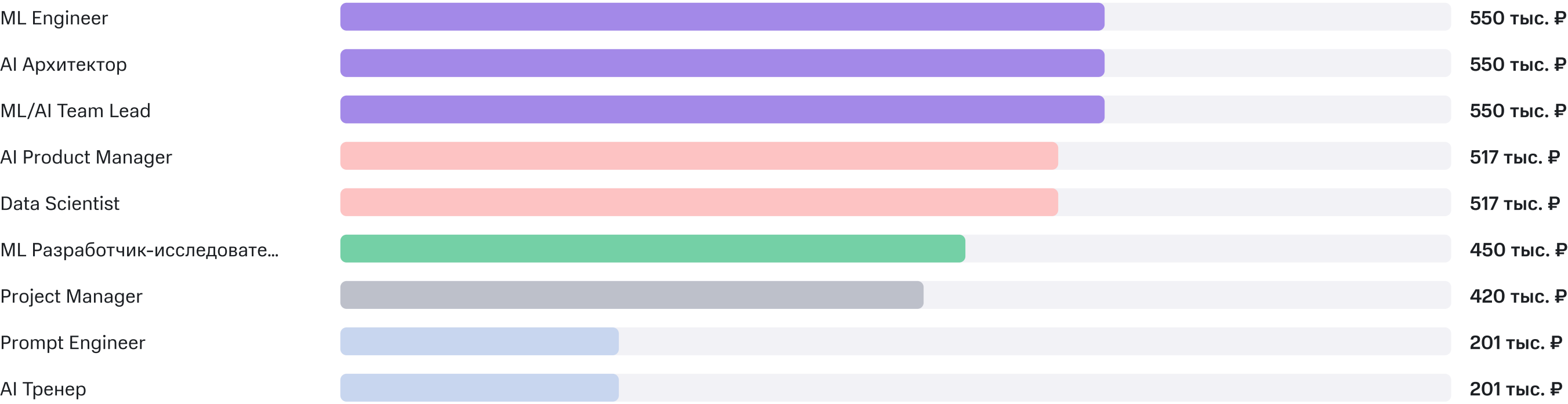


Разрыв Junior → Unit Lead достигает 5х. Самые высокие вилки — у Head of AI/CAIO и ML/AI Team Lead; самые скромные — у Prompt Engineer и AI Тренера.

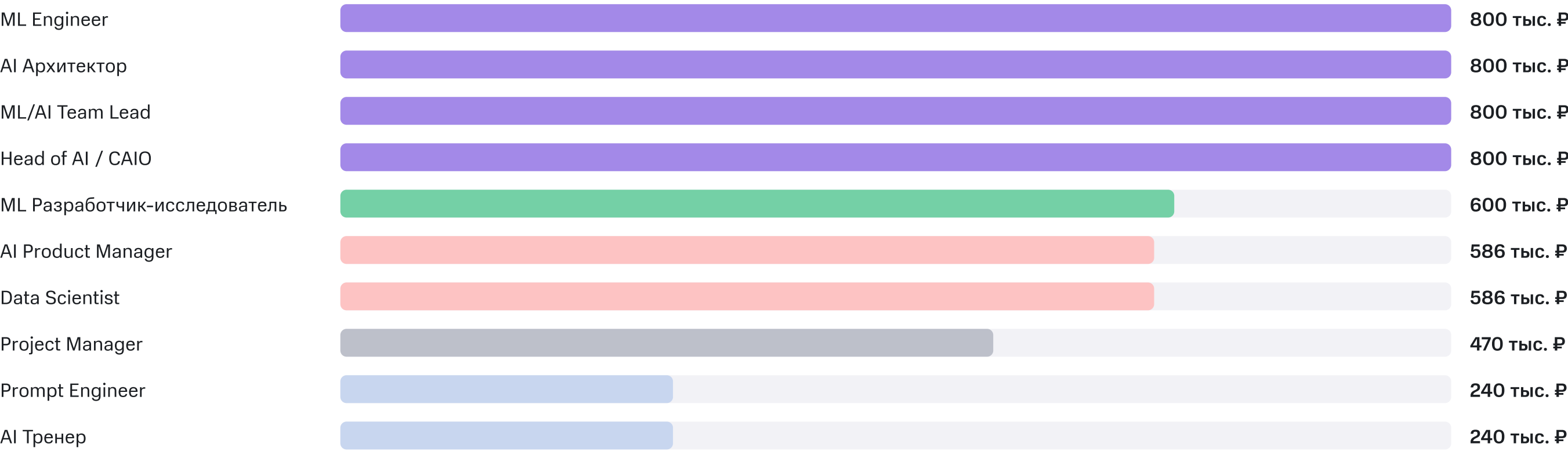
Сравнение зарплатных ожиданий Senior

ML-инженерия Исследователи Продуктовый офис Проектный офис AI-тренеры / Промпт

Медианные ожидания по зарплате — грейд Senior, ₽/мес.



Медианные ожидания по зарплате — грейд Lead, ₽/мес.

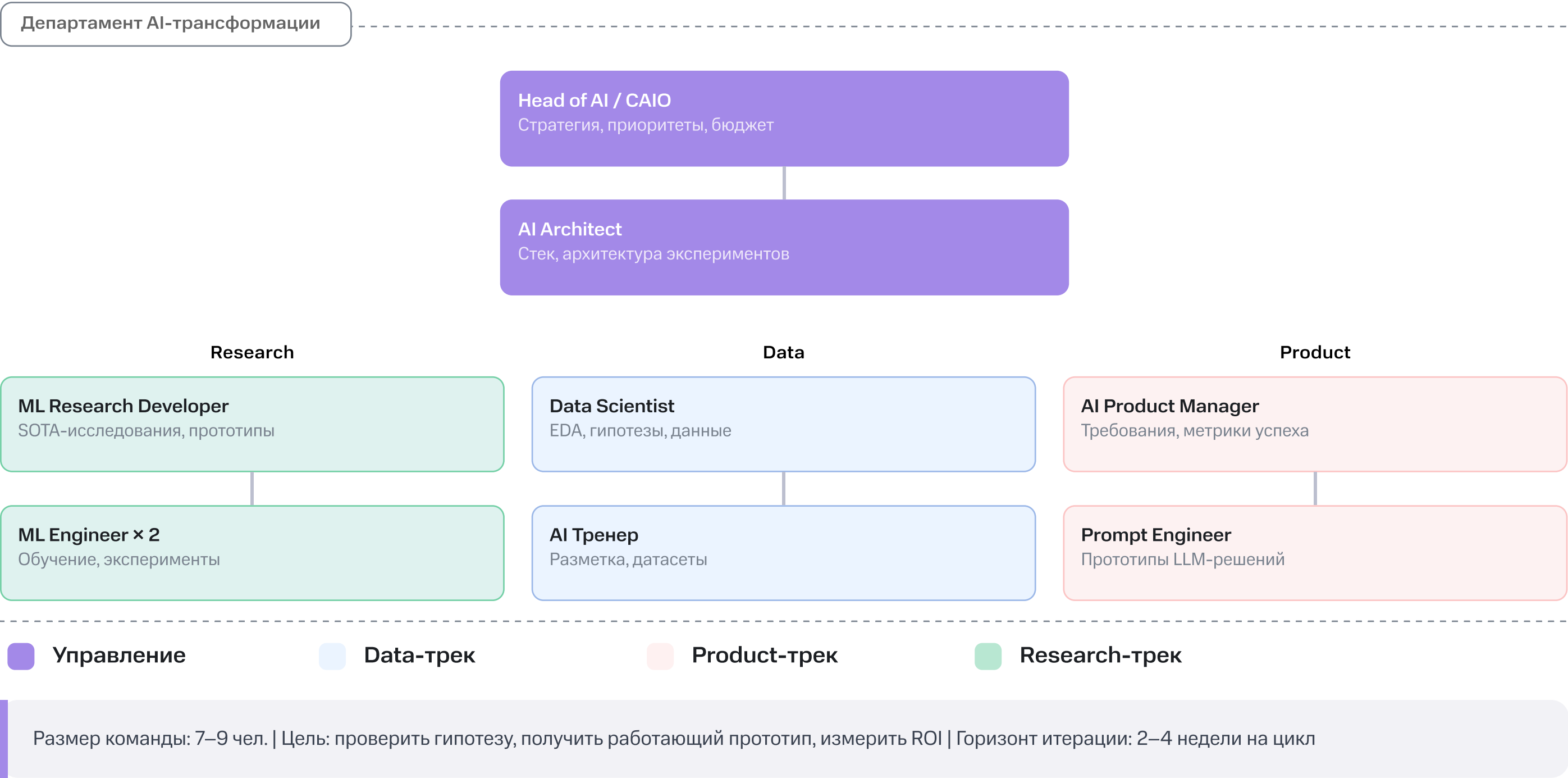


07

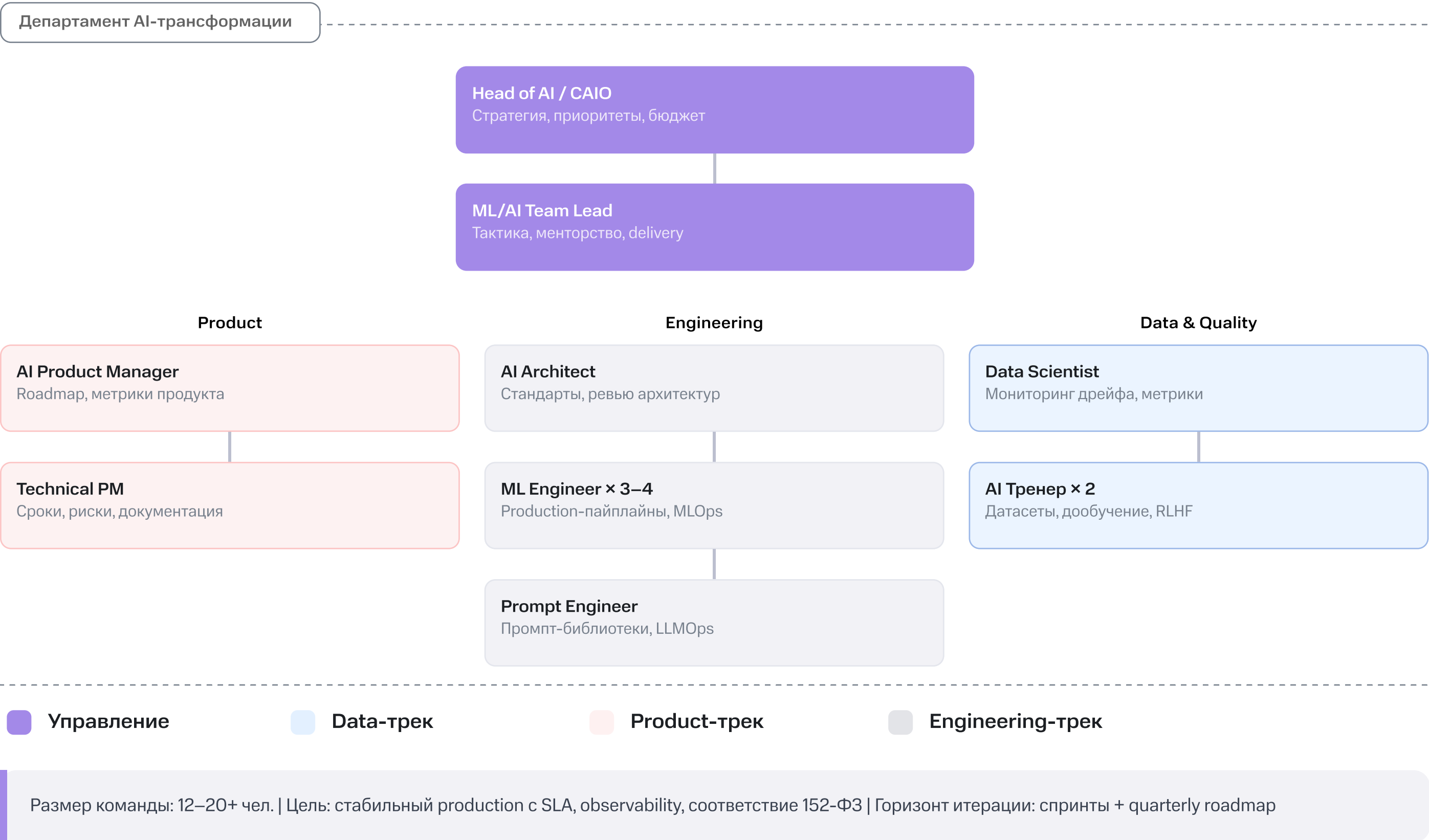
Оргструктуры: PoC и Production

Структура Департамента AI-трансформации в двух вариантах — для стадии проверки гипотез (PoC) и для промышленного внедрения (Production), с триггерами перехода между ними.

Вариант 1 — PoC-команда



Вариант 2 — Prod-команда



Сравнительный анализ структур

Параметр	Вариант 1 — PoC	Вариант 2 — Prod
Размер команды	7–9 чел.	12–20+ чел.
Главная ценность	Скорость от гипотезы до сигнала	Надёжность и управляемость delivery
Технический лидер L2	AI Architect — проектирует стек экспериментов, без управленческой нагрузки	ML/AI Team Lead — ведёт команду 6–12 чел., отвечает за инженерную культуру
Позиция AI Architect	L2 Выбирает стек и архитектуру с нуля	L3 Ревьюит решения в рамках устоявшейся системы
Technical PM	Нет AI PM самостоятельно ведёт backlog и коммуникации с бизнесом	Есть Управляет рисками, документацией и кросс-командным взаимодействием
ML Research Developer	Есть Ядро команды, reduction to practice	Нет Уходит в R&D-трек или роль не нужна
Data Scientist	Ищет закономерности, формулирует гипотезы для экспериментов	Мониторит дрейф модели, анализирует деградацию в production
Иерархия	Плоская — быстрый pivot	Многоуровневая — управляемость масштаба
Горизонт итерации	2–4 недели на цикл гипотеза → прототип	Спринты + квартальный roadmap
Observability	Базовая: LangSmith / LangFuse	Полная: Prometheus, Grafana, алерты на деградацию
Соответствие требованиям	Минимальное	152-ФЗ, стандарты ИБ, аудиты AI-систем
Ключевой риск	Слишком долго не выходим из PoC	Накопление техдолга при масштабировании
Триггер перехода	PoC прошёл валидацию на реальных данных + бизнес принял решение о масштабировании. Антипаттерн: создавать Prod-структуру под непроверенную гипотезу	

Что стоит за этими различиями

- Размер команды.** Разница не в сложности, а в параллельности: в PoC вся команда работает на одну гипотезу, в Prod одновременно идут поддержка, улучшение качества, инфраструктура и новая разработка.
- Главная ценность** — это противоречие, которое нельзя снять. Оптимизация под скорость и под надёжность — это разные архитектурные решения. Команда на быстрый pivot будет медленно реагировать на инциденты. Prod-команда с тяжёлыми процессами не сможет быстро проверять гипотезы. Поэтому это два разных режима, а не одна команда на двух скоростях.
- AI Architect L2 → L3.** Не понижение значимости, а смена горизонта ответственности. В PoC Architect делает стратегический выбор стека, который определяет всё остальное. В Prod этот выбор уже сделан — он следит, чтобы масштабирование не создавало архитектурного долга.
- ML Research Developer: есть vs нет** — самый показательный индикатор фазы команды. Исследователь нужен, когда неизвестно, что будет работать. После валидации он либо уходит в отдельный R&D-трек, либо его экспертиза просто не применяется.
- Observability: базовая vs полная.** В PoC достаточно понять, что модель даёт нужное качество. В Prod это нужно знать в реальном времени — с алертами, историей деградации и cost-per-request. Это не «больше инструментов», а другой уровень операционной зрелости.
- Ключевые риски.** Риск PoC — «вечный прототип»: бесконечное улучшение демо без перехода в продукт. Лечится явным критерием подтверждения гипотезы. Риск Prod — техдолг: архитектурные компромиссы ради скорости становятся проблемой через 6–12 месяцев. Лечится регулярными архитектурными ревью и контролем техдолга в backlog.
- Триггер перехода** — ключевой пункт. Переход — это не найм людей, а смена операционной модели целиком. Слишком рано — тратите ресурсы Prod-структуры на невалидированные гипотезы. Слишком поздно — работаете в Prod на PoC-инфраструктуре, что неизбежно приведёт к инцидентам.

08

ФОТ на команды — расчёт бюджета

Сравнение фонда оплаты труда для двух стадий: PoC (7–9 человек) и Production (12–18 человек).
С учетом рыночных зарплатных ожиданий кандидатов март 2026, Москва / гибрид.

PoC-команда — расчёт годового ФОТ

AI-трансформация · 7–9 чел. · С учетом рыночных зарплатных ожиданий кандидатов · март 2026 · Москва / гибрид

Роль / должность	Трек	Кол-во FTE	Грейд	Оклад в мес., ₽	Годовой оклад, ₽	Бонус (%)	Годовой ФОТ на позицию, ₽	Примечание
Head of AI	Управление	1	Lead	900 000	10 800 000	50%	16 200 000	
ML/AI Team Lead	Research	1	Lead	800 000	9 600 000	30%	12 480 000	
ML Researcher	Research	1	Senior	450 000	5 400 000	25%	6 750 000	
ML Engineer	Research	2	Middle	400 000	9 600 000	25%	12 000 000	2 штатные единицы
Data Scientist	Data	1	Middle	400 000	4 800 000	25%	6 000 000	
Project Manager	Operations	1	Senior	420 000	5 040 000	20%	6 048 000	
AI Trainer	Data	1	Middle	150 000	1 800 000	20%	2 160 000	
Prompt Engineer	Product	1	Middle	150 000	1 800 000	20%	2 160 000	
AI Product Manager	Product	0	Lead	517 000	3 102 000	20%	3 722 400	0,5 FTE — shared из продуктового департамента
ИТОГО — штатные позиции (FTE)					48 840 000		63 798 000	
+ Shared / привлечённые роли					3 102 000		3 722 400	
ИТОГО ФОТ (gross + бонус)					51 942 000		67 520 400	
Средний годовой ФОТ на 1 FTE (штат)							413 600	

Production-команда — расчёт годового ФОТ

AI-трансформация · 12–18 чел. · С учетом рыночных зарплатных ожиданий кандидатов · март 2026 · Москва / гибрид

Роль / должность	Трек	Кол-во FTE	Грейд	Оклад в мес., ₽	Годовой оклад, ₽	Бонус (%)	Годовой ФОТ на позицию, ₽	Примечание
Head of AI / CAIO	Управление	1	Unit Lead	1 050 000	12 600 000	50%	18 900 000	
AI Architect	Engineering	1	Lead	700 000	8 400 000	30%	10 920 000	
ML/AI Team Lead	Research	1	Lead	800 000	9 600 000	30%	12 480 000	
AI Product Manager	Product	1	Lead	586 000	7 032 000	20%	8 438 400	
Project Manager	Operations	1	Lead	470 000	5 640 000	20%	6 768 000	
ML Engineer (Senior)	Research	2	Senior	550 000	13 200 000	25%	16 500 000	
ML Engineer (Middle)	Research	2	Middle	400 000	9 600 000	25%	12 000 000	
Data Scientist	Data	2	Middle	400 000	9 600 000	25%	12 000 000	
AI Trainer	Data	2	Middle	150 000	3 600 000	20%	4 320 000	
Prompt Engineer	Product	1	Middle	150 000	1 800 000	20%	2 160 000	
MLOps / DevOps	Engineering	0	Senior	500 000	3 000 000	25%	3 750 000	0.5 FTE — shared из ИТ-инфраструктуры
ИТОГО — штатные позиции (FTE)					81 072 000		104 486 400	
+ Shared / привлечённые роли					3 000 000		3 750 000	
ИТОГО ФОТ (gross + бонус)					84 072 000		108 236 400	
Средний годовой ФОТ на 1 FTE (штат)							267 857	

09

Выводы по ролям и ФОТу

Практические рекомендации по каждой роли, последовательности найма и реальной стоимости владения AI-командой.

Выводы по ролям

Head of AI / CAIO (900k–1.05M/мес.) — самая дорогая и дефицитная позиция. Срок закрытия на рынке: 3–6 месяцев даже при сильном HR-бренде. Ошибка найма обходится в 3–6 месячных окладов + потеря темпа команды.

Рекомендация для HRD: запускать поиск параллельно с формированием команды, не ждать утверждения стратегии.

Рекомендация для CDTO: если роль закрыть быстро не получается — первые 3–6 месяцев её может совмещать CTO или VP Engineering при наличии глубокой AI-экспертизы.

AI Architect (700k/мес.) — в PoC отсутствует намеренно: на этапе эксперимента достаточно ML TL. В Production — критическая роль, без которой масштабирование превращается в хаос. На рынке почти нет готовых кандидатов: роль вырастает из Senior ML Engineer за 1.5–2 года.

Рекомендация: выращивать изнутри с момента старта PoC, заложить трек развития сразу.

ML/AI Team Lead (800k/мес.) — ключевой найм всей команды, определяет инженерную культуру. Конкуренция максимальная: Яндекс, Сбер, МТС, VK платят сопоставимо. Без публичного технического бренда компании (open-source, конференции, сильный стек) реальная ставка найма — 850–950k.

Рекомендация: при офере добавлять нефинансовые рычаги: автономия в выборе стека, возможность публиковаться, GPU-бюджет.

ML Engineer Middle (400k/мес.) — самый перегретый сегмент рынка. Воронка найма в 3–5 раз длиннее, чем для аналогичного backend-разработчика. Кандидаты с реальным production-опытом уходят в оффер за 3–5 дней.

Рекомендация HRD: держать pipeline всегда открытым, не ждать появления вакансии.

ML Research Developer (450k/мес., Senior) — нишевой профиль. Реалистичная альтернатива найму — стажировки из МФТИ/ВШЭ/ИТМО с конвертацией в штат. 80% сильных кандидатов имеют академический бэкграунд и сравнивают оффер не только по деньгам, но и по интересным задачам и возможностями публиковаться в том числе в международных изданиях.

Data Scientist Middle (400k/мес.) — наиболее зрелый рынок с наибольшим пулом кандидатов. Срок закрытия — 3–4 недели. Риск: через 1–1.5 года без роста в сторону ML Engineer или DS Lead уходит в более интересный проект.

Рекомендация: прописывать карьерный трек на берегу.

AI Product Manager (586k/мес., Lead) — редкая гибридная роль. На рынке ~80% кандидатов — либо PM без глубокого AI-понимания, либо ML-инженеры без продуктового опыта. Отсев на этапе технического скрининга — 70–80%.

Рекомендация CDTO: не пытаться переучить классического PM — искать профиль с хотя бы 1 годом работы в ML-смежной среде.

Project Manager (420–470k/мес.) — самый быстрый найм в команде (2–3 недели при правильном JD). Критично в JD прописать AI-специфику: опыт с ML-жизненным циклом, понимание неопределённости в исследовательских задачах. Без этого фильтра входящий поток будет нерелевантным.

AI Trainer / Prompt Engineer (150k/мес.) — массовый найм при масштабировании. Для PoC достаточно 1 человека с широким профилем. Главный риск: кандидаты переоценивают свою квалификацию, реальный уровень проверяется только на тестовом задании.

Рекомендация: делать тестовое обязательным.

Выводы для команды в целом

Для HRD: Бюджет PoC (~68 млн ₽/год) — это входной билет, эквивалентный небольшому продуктовому отделу. Главная ошибка — нанимать всю команду одновременно. Оптимальная последовательность: сначала ML TL + 1 MLE + DS + PM (4–5 чел., ~35 млн ₽/год), затем расширение по итогам первых спринтов. Head of AI лучше нанимать не позже 2-го месяца: без стратегического лидера команда быстро уходит в «вечный PoC».

Переход в Production (+40 млн ₽/год, +6 FTE) требует обоснования через ROI первого PoC. Без этого бюджет не пройдет защиту у CFO.

Рекомендация: ещё на этапе PoC фиксировать baseline-метрики бизнес-процессов, которые будет улучшать AI.

Для CDO: К ФОТ необходимо добавить операционный бюджет: GPU/облако (500k–2М ₽/мес. в зависимости от нагрузки), лицензии на инструменты (MLflow, W&B, Langsmith ~150–300k ₽/год), конференции и обучение (~50–150k ₽/чел./год). Реальная стоимость владения AI-командой в Production — 120–130 млн ₽/год с учётом инфраструктуры.

Гипотезы о соответствии рынку и трендам

Зарплатный рост продолжится. По данным hh.ru и внутренней аналитики, ML-роли растут на 15–25% YoY. Текущие вилки будут актуальны до конца 2026; к 2027 Senior MLE и ML TL могут сдвинуться на 15–20% вверх. Бюджеты стоит закладывать с запасом.

Дефицит Senior+ нарастает. Крупнейшие игроки (Яндекс, Сбер, МТС, VK) активно формируют собственные AI-лаборатории и поглощают доступный пул. Через 12–18 месяцев найм Senior ML без конкурентного employer brand будет занимать 4–6 месяцев.

AI Architect станет обязательной ролью. Сейчас роль редкость — есть у единиц. К 2027, по аналогии с Cloud Architect в 2018–2020, станет стандартом для любой enterprise AI-команды. Компании, которые начнут выращивать этот профиль сейчас, получают преимущество.

Роли Prompt Engineer и AI Trainer эволюционируют. Prompt Engineer либо вырастает в AI Product Designer / Applied AI Engineer, либо поглощается ML-инженером по мере упрощения инструментария. AI Trainer с ростом синтетических данных и автоматической разметки трансформируется в Domain Expert — специалиста с предметной экспертизой, а не просто разметчика.

GPU-инфраструктура становится конкурентным преимуществом при найме. Кандидаты уровня Senior+ всё чаще оценивают не только зарплату, но и доступность вычислительных ресурсов. Компании без собственного GPU-кластера или облачного бюджета проигрывают борьбу за исследователей уже в 2026.



AI

РАЗДЕЛ 04

Практика найма

Реальные метрики найма AI-специалистов, каналы поиска, причины ухода и продажа вакансии.

В ЭТОМ РАЗДЕЛЕ

- 1 Опыт найма MWS AI — метрики 2025
- 2 Мотивация и продажа вакансии



10

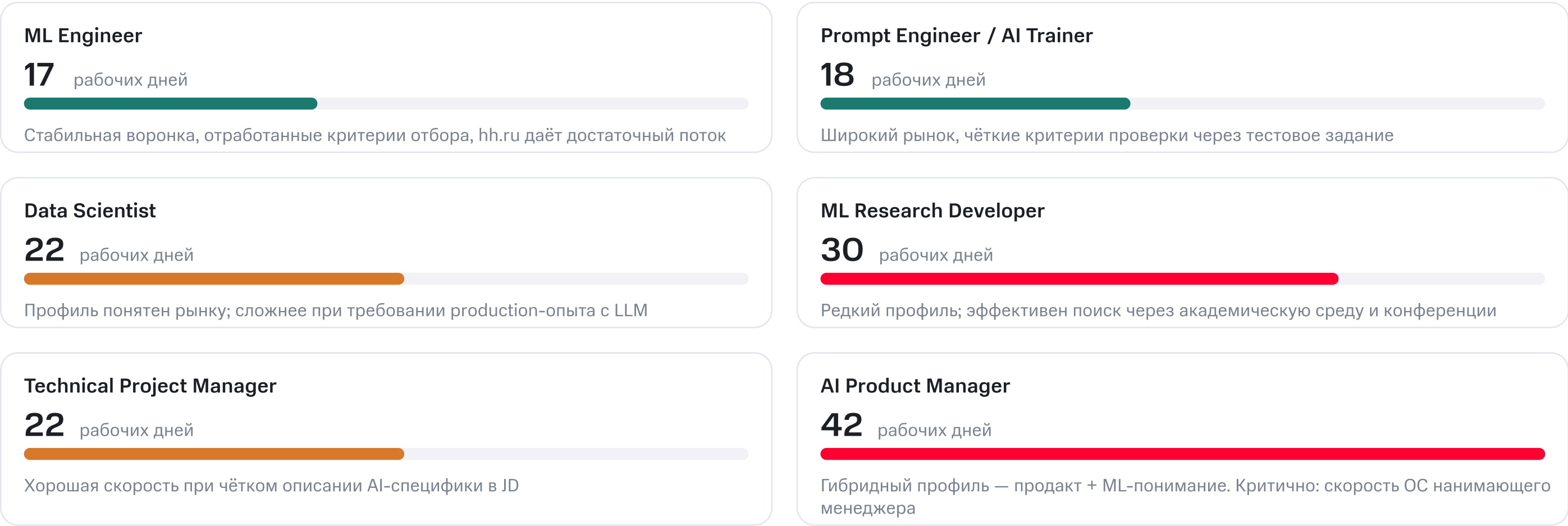
Опыт найма MWS AI — метрики

Найм в AI-команду: что работает на практике

Раздел гайдбука, основанный на реальном опыте найма AI-специалистов за 2025 год. Мы нанимали ML-инженеров, исследователей, продуктовых менеджеров и архитекторов в условиях дефицита рынка — и готовы поделиться тем, что работает, а что только кажется работающим.

1 Сколько времени занимает закрытие вакансии

Средний срок закрытия по компании в 2025 году составил **25 рабочих дней**. Но это среднее скрывает принципиально разный опыт для разных ролей.



Head of AI / CAIO и ML/AI Team Lead

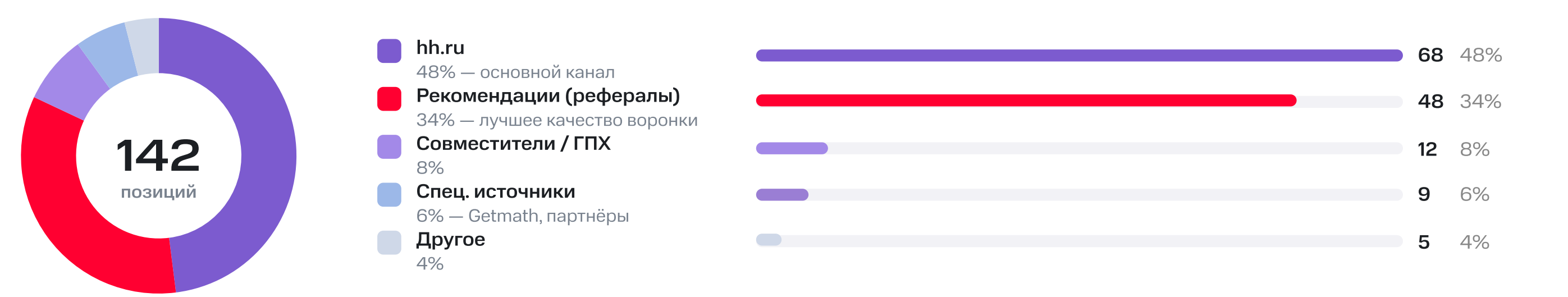
Эти роли выходят за рамки стандартной воронки. Реалистичный срок — 60–90 календарных дней, нередко до 4 месяцев. Поиск ведётся через executive search, профессиональные сообщества, прямой хедхантинг. Срок закрытия почти не зависит от рекрутера — он зависит от того, насколько точно сформулирован профиль и насколько быстро принимаются решения на стороне заказчика.

60–90
календарных дней

Главный враг скорости закрытия — не рынок, а внутренние задержки. В нашем опыте задержки обратной связи от нанимающего менеджера удваивают срок закрытия. На конкурентном AI-рынке кандидат уровня Team Lead или AI Architect крайне редко ждёт ответа дольше 5–7 рабочих дней.

2 Где мы искали — и что сработало

За 2025 год через все каналы закрыто 142 позиции. Распределение по источникам:



hh.ru остаётся основным каналом — но для AI-ролей выше уровня middle его одного не хватает. Реферальная программа дала 34% от всего найма, а для senior-позиций — значительно больше.

Реферальная программа — самый недооценённый инструмент в AI-найме. Рекомендованные кандидаты приходят уже с пониманием культуры, быстрее проходят этапы отбора и реже отказываются от оффера. Для ML Research Developer и AI Architect реферальный канал закрывает вакансии быстрее и качественнее, чем любой job-борд.

2 Где мы искали — и что сработало

За 2025 год через все каналы закрыто 142 позиции. Распределение по источникам:

Какой канал работает для какой роли

Роль	hh.ru	Рекомендации	Конференции / сообщества	Executive Search
ML Engineer	Основной	Доп.	—	—
Data Scientist	Основной	Доп.	—	—
ML Research Developer	Слабо	Основной	Ключевой	—
Prompt Eng. / AI Trainer	Основной	—	—	—
AI Product Manager	Доп.	Основной	Доп.	—
AI Architect	Слабо	Основной	Доп.	—
ML/AI Team Lead	Слабо	Основной	Доп.	Доп.
Head of AI / CAIO	Не работает	Доп.	Ключевой	Основной

3 Откуда приходят кандидаты

ML Engineer · Architect

- Яндекс (YandexGPT, Search, Cloud)
- Сбер AI (GigaChat, SberDevices)
- MTS AI
- VK (МАРУС, Почта)
- Т-Банк (Tinkoff AI)
- Газпромбанк Tech

ML Research Developer

- МФТИ (лаборатории ML)
- ВШЭ (факультет CS)
- ИТМО (AI-направление)
- МГУ (мехмат, ВМК)
- Сколтех
- Getmath / AIChallenge-участники

AI Product Manager

- Авито (ML-продукты)
- Ozon Tech
- Wildberries Tech
- СберМаркет / Delivery Club
- Яндекс (продуктовый трек)
- 2ГИС / Дубль ГИС

Prompt Eng. · AI Trainer

- Лингвистические факультеты МГУ, РГГУ
- Технические писатели IT-компаний
- Контент-менеджеры с AI-опытом
- Переводчики и редакторы
- Фриланс-разметчики (Толока и др.)

Head of AI · Team Lead

- Яндекс (руководители ML-направлений)
- Сбер AI (Chief AI Officer уровень)
- Mail.ru / VK (AI-лиды)
- Rambler / RuTube
- Спин-оффы и AI-стартапы экосистем

Technical PM · Аналитики

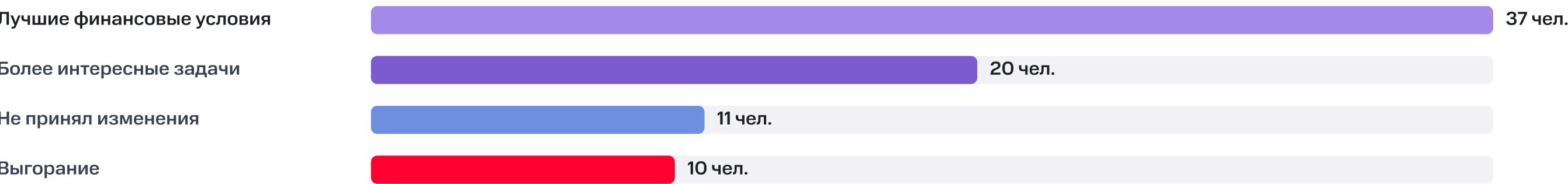
- Проектные офисы крупных IT-компаний
- Системные интеграторы (ЛАНИТ, ЕРАМ)
- FinTech: Тинькофф, Альфа
- Госсектор (Постелеком, Росатом Digital)
- Консалтинг: Kept, ex-McKinsey Digital

Важный контекст: компании-доноры — это одновременно ваши конкуренты за кандидата. Специалист из Яндекса или Сбера привыкает к определённомu уровню инфраструктуры, задач и компенсации. Если ваше предложение не соответствует хотя бы двум из трёх — привлечение займёт значительно больше времени, чем закладывает стандартный план найма.

4 Почему AI-специалисты меняют работу

По данным exit-интервью за 2025 год: из 147 ушедших сотрудников 37% назвали финансовые условия главной причиной. Это самый высокий показатель среди всех категорий. Но мотивация к смене работы (то, что заставляет кандидата ответить на звонок рекрутера) — несколько другое.

Причины ухода — весь AI-трек, 2025



Данные exit-интервью, 2025. Всего 147 уволившихся.

ML Engineer · Data Scientist · AI Trainer

- Компенсация выше рынка — основной триггер отклика на холодный контакт
- Чёткие задачи в production — не «помогать исследователям», а делать реальный продукт
- Качественная инфраструктура: GPU-кластеры, нормальный MLOps-стек
- Прозрачный грейдинг и рост без «подожждём годик»

ML Research Developer

- Академическая свобода и время на «высокорисковые» исследования
- Возможность публиковаться (NeurIPS, ICLR, EMNLP)
- Работа рядом с сильными людьми из топ-вузов и open-source
- Конкурентная зарплата — но не главный фактор принятия оффера

AI Product Manager

- Строить AI-продукт с нуля, а не поддерживать legacy
- Прямой контакт с командой разработки без бюрократии
- Культура data-driven решений — не просто декларируемая
- Компенсация + equity / бонус за product-метрики

Head of AI / ML Team Lead

- Реальный мандат и поддержка C-level — не «декоративная» должность
- Возможность выстроить команду самостоятельно, без унаследованных проблем
- Публичная позиция: конференции, сообщество, видимость на рынке
- Компенсация в диапазоне 600к–1,5М + долгосрочный стимул

Разница между исследователем и инженером в части мотивации принципиальна: инженер идёт туда, где лучше платят и интереснее задачи. Исследователь идёт туда, где дадут время думать и возможность публиковаться. Один оффер не закрывает обе потребности.

5 Что мы вынесли из года найма — для вас

- 01 / Самое узкое место — не рынок, а скорость решений внутри.** Дефицит AI-специалистов реален, но большинство потерь кандидатов происходит не из-за рынка, а из-за того, что нанимающий менеджер не смотрит резюме три дня или не может выйти на встречу неделю. На конкурентном рынке это стоит закрытой вакансии.
- 02 / Два-три ключевых донора важнее широкой воронки.** Лучшие ML Engineers приходят из Яндекса, Сбера и MTS AI. Лучшие исследователи — из МФТИ и ВШЭ. Знание этих маршрутов и точечная работа с ними эффективнее, чем массовый парсинг hh.ru.
- 03 / Реферальная программа должна работать с первого дня.** 34% найма через рекомендации — это не случайность. Это результат осознанной программы, в которой сотрудники мотивированы рекомендовать и понимают, каких людей вы ищете.
- 04 / Удержание дешевле найма — особенно в AI.** Замена уволившегося senior ML Engineer стоит 3–5 месяцев закрытия + онбординг. Знание причин ухода (финансы и интересность задач) позволяет работать превентивно — до того, как человек начал смотреть рынок.

11

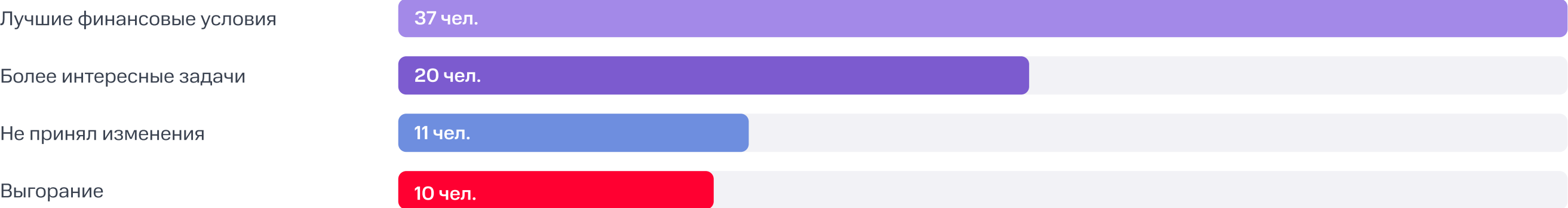
Мотивация и продажа вакансии

Как продать вакансию ML-разработчику, который работает в крупной tech-компании

Данные exit-интервью говорят прямо: деньги — причина ухода №1, но не главный мотив перехода. Это разница, которую большинство нанимающих менеджеров не улавливают — и проигрывают борьбу за оффер.

Что говорят данные exit-интервью

Причины ухода из AI-подразделений, 2025 (147 чел.)



Источник: HR-отчёт 2025, exit-интервью

Ключевое наблюдение. Деньги — причина ухода. Интересные задачи и смысл — причина прихода. Это разные психологические триггеры. Бить в деньги при рекрутинге из крупной tech-компании — проигрышная стратегия: вы всегда заплатите больше или потеряете кандидата через 6 месяцев тому, кто заплатит ещё больше.

Три профиля — три разных разговора

ML Research Developer

из крупной tech-экосистемы

- Хочет публиковаться и видеть своё имя в SOTA
- Устал от бюрократии и конкуренции внутри большой команды
- Хочет видеть реальный эффект своей работы на производстве
- Ценит интеллектуальную свободу и время на «странные» гипотезы

ML Engineer (Senior)

из продуктовой tech-компании

- Устал поддерживать чужой legacy-код
- Хочет строить с нуля, а не дорабатывать
- Ищет роль, где его решения видны и имеют вес
- Хочет ближе к бизнесу: понимать зачем, а не только как

AI/ML Team Lead

из крупной tech-экосистемы

- Хочет настоящий мандат, а не декоративную должность
- Хочет строить команду самостоятельно, без унаследованных проблем
- Устал от политики и конкуренции внутри большой компании
- Ищет публичность и влияние за пределами одной экосистемы

Кейс: ML Research Developer → промышленная компания

Что кандидат думает о промышленном AI, и почему он ошибается

Типичный страх кандидата

«Там скучно. Автоматизация конвейеров — это не research. Я буду деградировать. Там нет сильных людей рядом. Данные плохие, задачи простые.»

Реальность, о которой он не знает

Промышленные данные — это мультисенсорные временные ряды, которых нет в open-source. Задачи дефектоскопии, предиктивного обслуживания, оптимизации плавки — это frontier research в прикладном смысле.

Как работать с этим?

Что продаём исследователю

- Уникальные данные, которых нет нигде: сенсоры, химия плавки, вибрации — это не MNIST
- Публикации реальны: промышленный AI — активно цитируемая область в NeurIPS и ICML
- Строит первым: нет унаследованного кода, нет «так исторически сложилось»
- Видит результат: модель реальное производство измеримый экономический эффект
- Автономия: нет очереди из 20 research-инженеров за одним GPU-кластером

Что продаём ML-инженеру

- Greenfield: строит MLOps-стек с нуля под свои стандарты, не допиливает чужое
- Близость к бизнесу: понимает, зачем строит — в ROI, в тоннах стали, в % брака
- Масштаб: production на реальных промышленных нагрузках — не pet-проект
- Влияние: архитектурные решения его, а не комитета из 15 человек
- Рост в уникальный профиль: Industrial AI Engineer — дефицитный профиль уже сейчас

6 нефинансовых рычагов убеждения

01 Уникальность данных

Промышленные данные недоступны в open-source. Исследователь, поработавший с мультисенсорными временными рядами с реального производства, получает эксклюзивную экспертизу, которую не воспроизвести в крупной tech-компании.

«У нас есть 10 лет сенсорных данных с 3 заводов. Нигде на рынке этого нет.»

02 Видимость результата

В большой tech-компании результат растворяется в продукте. В промышленности — модель снизила брак на 4%, сэкономила X млн рублей. Это измеримо, цитируемо, вписывается в резюме как конкретный кейс.

«Ваш код будет работать на реальном производстве, а не в A/B-тесте на 1% аудитории.»

03 Строить, а не поддерживать

Для человека, которому надоел legасу, главный аргумент — чистый лист. Нет накопленного техдолга, нет «так исторически сложилось», нет комитетов согласования каждого архитектурного решения.

«Вы — первый. Архитектура будет вашей.»

04 Публикации и конференции

Промышленный AI активно публикуется. ICML, NeurIPS, KDD принимают прикладные работы с реальными данными. Исследователю важно, что компания поддерживает публикации, а не засекречивает всё.

«Мы готовы публиковать результаты. У нас уже есть опыт совместных статей с профильными вузами.»

05 Автономия и доверие

В больших экосистемах у сильного специалиста — очередь за ресурсами, согласование каждого эксперимента, конкуренция с коллегами. В новой AI-функции — прямая линия к руководству, собственный бюджет, своя команда.

«Вы не ждёте, пока кто-то освободит GPU. Вы сами решаете, что запускать.»

06 Редкий профиль = рыночное преимущество

Industrial AI Engineer — один из самых дефицитных профилей к 2027 году. 2–3 года опыта в промышленном AI дадут кандидату уникальную позицию на рынке, которую не получить, оставаясь в общем потоке продуктовых ML-команд.

«Через 3 года ваше резюме будет стоить значительно дороже, чем если вы останетесь.»

Как строить разговор с кандидатом

Шаг 1 — Сначала слушать, не продавать (15 мин)

- «Что в вашей текущей работе вас энергизирует, а что забирает энергию?»
- «Какую задачу вы хотели бы решить, но не можете в текущей роли?»
- «Что для вас важнее: глубина в одной теме или ширина применения?»
- Не перебивайте. Фиксируйте — именно его словами опишете вакансию на следующем шаге.

Шаг 2 — Отзеркалить его мотиваторы на вашу задачу

- Если устал от legасу → «У нас нет ничего написанного. Вы — архитектор.»
- Если хочет публиковаться → «Мы закладываем бюджет на 1–2 конференции в год и поддерживаем статьи.»
- Если хочет видеть результат → «Ваша модель в production = реальный завод, реальные деньги, реальные люди.»
- Если устал от конкуренции внутри → «Здесь вы не один из десяти, а единственный. С прямым выходом на CDO.»

Шаг 3 — Закрыть страхи напрямую

- «Данные плохие?» → Покажите примеры: объём, разнообразие, уникальность. Дайте попробовать на пилоте.
- «Нет сильных людей рядом?» → Назовите имена. Если их нет — честно скажите, кого планируете нанять и когда.
- «Буду деградировать?» → Покажите roadmap: что будет через 6, 12, 24 месяца. Какие задачи появятся.
- «Компания не понимает AI?» → Назовите конкретного C-level-спонсора и его мандат. Без него — вакансия мертва.

Что помогает успешному офферу

Слушать

Сначала узнать, что движет кандидатом — только потом продавать

Конкретно

Данные, задачи, имена, цифры — вместо «интересно» и «стабильно»

Быстро

48 часов ответ, 5 дней решение. Senior не ждёт дольше

Антипаттерны — что убивает оффер

«Мы дадим больше денег»

Сильный кандидат из крупной tech-компании понимает: если единственный аргумент — деньги, через полгода появится тот, кто заплатит ещё больше.

→ Деньги закрывают порог. Смысл и задачи — удерживают.

«У нас интересные задачи»

Абстрактное утверждение без конкретики. Каждый рекрутер так говорит. Без примеров, данных и имён — это пустые слова.

→ Конкретная задача + конкретные данные + конкретный результат.

«У нас стабильность»

Для исследователя и сильного инженера «стабильность» = стагнация. Этот аргумент работает против вас в этом сегменте.

→ Говорите про рост, масштаб и уникальность — не про безопасность.

Медленный процесс

Кандидат из крупной tech-компании ведёт 2–3 процесса одновременно. Молчание больше 5–7 дней = потеря. Всегда.

→ SLA: 48ч ответ, 5 дней решение после финала.

Продавать то, чего нет

«У нас уже есть отличная AI-инфраструктура» — если это неправда, кандидат узнает на первой неделе и уйдёт.

→ Честность про текущее состояние + ясный план, что будет.

Нет C-level спонсора

Если нанимающий менеджер не может назвать конкретного топ-менеджера, который «болеет» за AI — умный кандидат откажется. Он знает, как умирают AI-инициативы без поддержки сверху.

→ Перед наймом — получите видимый мандат от CDO или CEO.

Что помогает успешному офферу

Слушать

Сначала узнать, что движет кандидатом — только потом продавать

Конкретно

Данные, задачи, имена, цифры — вместо «интересно» и «стабильно»

Быстро

48 часов ответ, 5 дней решение. Senior не ждёт дольше

Главный принцип: вы не переманиваете из крупной tech-компании деньгами — вы предлагаете то, чего в крупной tech-компании нет: уникальные данные, видимый результат, строить с нуля и стать первым в нише. Это другой разговор.

Ключевые выводы

- 1

Деньги закрывают порог, но не удерживают

Финансовые условия — причина ухода №1 по данным exit-интервью. Но они же — наименее устойчивый мотив при переходе: кандидат, пришедший только за деньгами, уйдёт к следующему, кто предложит больше. Нефинансовые мотиваторы — уникальность задачи, автономия, видимый результат — формируют лояльность, которая не продаётся на аукционе офферов.
- 2

Промышленные и «нетеховые» компании имеют реальные конкурентные преимущества в найме AI

Уникальные данные, которых нет в открытом доступе, — сильнейший аргумент для исследователя. Возможность строить с нуля — для инженера. Прямая связь между работой модели и измеримым бизнес-результатом — для обоих. Это то, чего принципиально нет в крупных tech-экосистемах, где специалист — один из многих и результат его работы растворяется в продукте. Компании, научившиеся это формулировать, выигрывают в найме не за счёт бюджета, а за счёт ценностного предложения.
- 3

Один разговор не работает для всех профилей

ML-исследователь, ML-инженер и Team Lead принимают решение о переходе по принципиально разным критериям. Исследователю важны публикации и свобода экспериментов. Инженеру — архитектурная автономия и масштаб. Лидеру — мандат и возможность строить команду. Попытка продать одно и то же всем трём — гарантированный провал. Рекрутер и нанимающий менеджер должны сначала выяснить, что движет конкретным человеком, и только потом выстраивать оффер.
- 4

Честность про текущее состояние — это конкурентное преимущество, а не слабость

Сильный кандидат умеет читать между строк и проверяет заявления на первой же встрече с командой. Компании, которые честно описывают текущее состояние AI-функции («мы только начинаем, данные сырые, инфраструктуры нет — но вот что мы строим и вот кто за этим стоит») вызывают больше доверия, чем те, кто продаёт несуществующую инфраструктуру. Кандидат с опытом предпочтёт строить с нуля при честном руководстве, чем войти в обманутые ожидания.
- 5

Без C-level спонсора AI-вакансия не продаётся

Опытный AI-специалист видел достаточно AI-инициатив, умерших через год из-за отсутствия поддержки сверху. Если нанимающий менеджер не может назвать конкретного топ-менеджера, который лично стоит за AI-трансформацией, — умный кандидат откажется. Наличие видимого спонсора в лице CDO или CEO — это не формальность, а сигнал серьёзности намерений. Это нужно решить до начала найма, а не в процессе.
- 6

Скорость решений — часть оффера

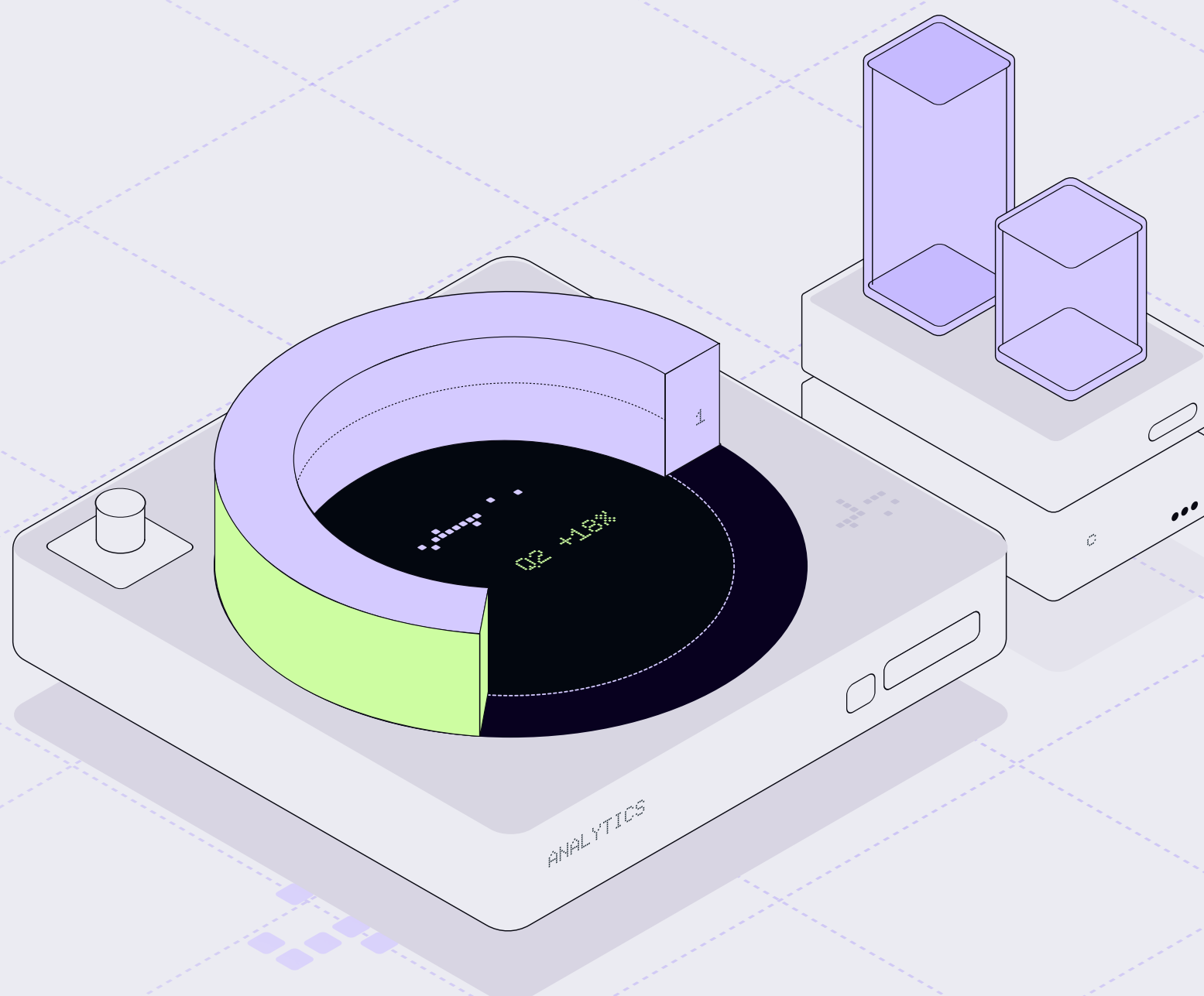
Senior AI-специалист из сильной tech-компании ведёт несколько процессов одновременно. Молчание больше 5–7 рабочих дней на любом этапе — это не нейтральное событие, это сигнал: «мы не умеем принимать решения быстро». Именно с такой скоростью кандидат будет работать после выхода — так он думает. Установленный внутренний SLA на ответ (48 часов на первичную реакцию, 5 дней на финальное решение) — это не про вежливость, это про конкурентоспособность оффера.

Тренды

Куда движется рынок AI-найма до 2029 года: спрос, компетенции, агентные системы и выводы для ЛПР.

В ЭТОМ РАЗДЕЛЕ

- 1 Тренды рынка AI до 2029
- 2 Тренды — инфографика
- 3 Выводы для ЛПР



12

Тренды рынка до 2029

Шесть ключевых трендов российского рынка AI-специалистов: что происходит сейчас и каким будет горизонт через три года.

Тренд 1. Рынок охладел, но не ослаб — и это надолго

По данным HeadHunter, в 2025 году количество IT-вакансий сократилось примерно на 20%, а количество резюме выросло более чем на четверть.¹ Это не кризис — это нормализация после перегрева 2021–2023 годов. Рынок окончательно переходит от модели «рынка кандидата» к «рынку работодателя» в массовом сегменте, оставаясь дефицитным только в узких нишах.^{2 3}

Эффект сейчас. Компании нанимают точно, требования к кандидатам выросли, конкуренция за junior-позиции огромная. В России сейчас работает около 1,1 млн IT-специалистов — рост на 14% за год. При этом 64% компаний фиксируют дефицит кадров, а реальная нехватка оценивается в 500–700 тысяч человек. Разрыв в 10 раз между зарплатой junior и Lead с AI-опытом — это новая норма.⁴

Через 3 года. Биполярность усилится: рынок mid-уровня — конкурентный, рынок senior AI-специалистов — дефицитный хронически. Компании, не выстроившие систему внутреннего роста, будут платить всё больше за внешний найм.

Тренд 2. AI-компетенции становятся базовым требованием везде

В 2025 году российские работодатели столкнулись со смещением спроса в сторону кандидатов, умеющих работать с AI. Пункт «базовое знание AI» встречается уже в каждой второй вакансии.^{5 6}

Во втором квартале 2025 года доля крупных российских компаний, использующих AI-решения, выросла до 43% — год назад было 28%. Это не стартапы и не эксперименты: AI становится производственным инструментом в финтехе, промышленности, медицине.⁷

Эффект сейчас. Самые высокие зарплаты в IT — у специалистов по искусственному интеллекту и DevOps-инженеров. На этих направлениях сохраняется острая нехватка экспертов. Переход с middle на senior увеличивает доход на 70–100% — это главный карьерный рычаг.⁸

Через 3 года. AI-неграмотность станет профессиональным барьером наравне с незнанием SQL. Роль «разработчик, не умеющий работать с AI-инструментами» исчезнет не потому, что людей заменят, а потому что рынок перестанет такие роли финансировать.

Тренд 3. Агентный AI — от пилотов к инфраструктуре

2025-й стал годом, когда AI-агенты вышли из экспериментов. Около 87% разработчиков по всему миру применяют AI в рабочих процессах.⁹ Мы перешли порог от экспериментов к инфраструктуре.¹⁰ По прогнозам MTS AI, объём российского рынка AI-ассистентов для разработчиков в 2025 году составит 17,4 млрд рублей.¹¹

2026 год компании называют годом агентов: главный тренд — усиление мультиагентных систем. Для полноценной автоматизации точечных внедрений уже недостаточно — нужны комплексные решения. Разрыв между компаниями, начавшими в 2023 году, и остальными углубляется.¹²

Эффект сейчас. Для HR и AI-консалтинга это означает: клиент приходит не за одним ML-инженером — он хочет агентную систему под конкретный бизнес-процесс. Спрос смещается от «постройте нам модель» к «автоматизируйте нам процесс». Компании используют агентов, которые надстраиваются поверх существующих ERP и CRM — без замены текущих систем.¹³

Через 3 года. Ключевые запросы при найме IT-специалистов смещаются: раньше нужно было знание синтаксиса языков, сейчас — умение проектировать, настраивать и интегрировать AI-инструменты. Роли будущего — инженер-оркестратор агентов, специалист по контролю качества AI-систем, архитектор мультиагентных пайплайнов.¹⁴

Тренд 4. Импортозамещение создаёт устойчивый спрос на специфические профили

Государство сдвинуло крайний срок перехода на российское ПО на 2028 год, активно обсуждаются оборотные штрафы за использование иностранного ПО. Это создаёт устойчивый, не зависящий от конъюнктуры спрос на конкретные категории специалистов. Разработчики 1C и администраторы российских СУБД — среди главных бенефициаров этой политики.^{15 16}

Спрос на 1C-разработчиков вырос на 25% за два года. Специалист по Astra Linux стал золотым единорогом. Тот, кто умеет мигрировать с Oracle на PostgreSQL, может выбирать из десятка предложений.⁴

Эффект сейчас. Для AI-команд это имеет прямое значение: все LLMOps и MLOps-стеки строятся на российских облаках (VK Cloud, Yandex Cloud, SberCloud), а не на AWS/GCP. Навык работы с суверенной инфраструктурой — конкурентное преимущество, которое пока недооценивается.

Через 3 года. К 2028 году рынок специалистов по российскому enterprise-стеку будет перегрет. При этом компании, которые начали строить AI-команды сейчас на отечественной инфраструктуре, получают преимущество перед теми, кто откладывал.

Тренд 5. Гибридизация профессий — смерть «чистых» ролей

Гибридизация IT-профессий — тренд, который будет только усиливаться: разработчик с навыками интеграции моделей, инженер с опытом автоматизации с помощью AI-агентов, аналитик с MLOps-компетенцией.¹⁴

В вакансиях для Project Manager и Delivery Manager начали встречаться требования, связанные с вайб-кодингом и быстрым прототипированием. AI не заменяет команды, а меняет роли: рутинные задачи всё чаще уходят к AI, а человек переходит на контроль качества его работы.

Эффект сейчас. Technical PM без понимания ML-стека, Data Scientist без инженерных навыков, AI PM без способности читать код — эти профили становятся неконкурентоспособными быстро. Компании платят не за должность, а за способность человека работать на стыке.

Через 3 года. Специалист, свободно владеющий AI, работает продуктивнее — и стоит дороже. Через 3 года компании, заменившие часть штата AI-усиленными специалистами, получают структурное преимущество в скорости и себестоимости.

Тренд 6. Рекрутинг сломался — и чинится медленно

AI вместе с ATS-фильтрами стал коварным инструментом в руках работодателя. Массовый отсев кандидатов привёл к ответному удару: кандидаты используют AI для прохождения технических собеседований.¹⁷

Значительная часть вакансий теперь закрывается не через job-борды, а через нетворкинг и рекомендации. Из-за этого появляются тысячи нерелевантных резюме, и качество воронки падает.¹⁸

Эффект сейчас. Рекрутеры, работающие только с hh.ru, теряют лучших кандидатов. Реферальная программа и прямой поиск дают принципиально иное качество воронки. Это подтверждает и наш собственный опыт: 34% всех найма в 2025 году — рекомендации.

Через 3 года. HR-функция разделится на два принципиально разных трека: массовый найм (полная автоматизация через AI-скрининг) и экспертный найм (отношения, нетворкинг, прямой поиск). Рекрутер, умеющий только первое, будет заменён агентом. Умеющий второе — станет дороже.

Главный вывод для клиента: эксперты «Яков и партнёры» и Яндекса оценили ожидаемый экономический эффект от AI в 7,9–12,8 трлн рублей в год к 2030 году — до 5,5% прогнозного ВВП. Компании, которые начнут строить AI-команды в 2025–2026 годах, войдут в этот рост с 2–3-летним преимуществом перед теми, кто ждёт.¹⁹

Источники

1. hh.ru — «Рынок труда в 2025 году: почему стало труднее искать работу»

2. sense-it.ru — «Зарплаты IT-специалистов в РФ: основные тренды I квартала 2026»

3. elbrusboot.camp — «IT-рынок 2026: секреты найма и 5 стратегий»

4. vc.ru — «Парадокс IT-рынка России 2025»

5. mfl.ru — «Дефицит кадров и продавцы в топе: как рынок труда пережил 2025 год»

6. forbes.ru — «Спрос на сотрудников с навыками применения ИИ вырос на треть»

7. comnews.ru — «Российские компании стали чаще внедрять инструменты ИИ»

8. 1000bankov.ru — «Российский IT без перегрева: сколько теперь платят»

9. 4pda.to — «87% разработчиков то и дело прибегают к помощи ИИ»

10. habr.com — статья по агентному AI (1059506)

11. cnews.ru — «MTS AI начинает продажи внешним клиентам»

12. generation-ai.ru — «Recap 2025»

13. bitrix24.ru — «AI-агент Bitrix24: автоматизация»

14. techinsider.ru — «Профессии, которые не исчезнут: каким будет IT-рынок после 2030»

15. course-itrecruiter.ru — «Рынок труда IT 2026»

16. vedomosti.ru — «Переход значимых объектов КИИ на отечественный софт отложили на восемь лет»

17. incrussia.ru — «AI-HR: AI-сотрудник»

18. kinzhal.media — «Супер-отклик»

19. yakovpartners.ru — «AI 2025»

13

Тренды — инфографика

IT и AI-рынок: тенденции, эффекты и горизонт 3 лет

Визуальная сводка ключевых тенденций рынка AI-найма.
Открытые источники 2025–2026: hh.ru · Forbes · Ведомости · Habr · TAdviser · McKinsey · Gartner · Deloitte · MTS AI.

Тренд 1 — Охлаждение рынка

- 34%

IT-вакансий за год

Число открытых позиций упало — но это не кризис, а нормализация после перегрева 2021–23

hh.ru, 2025

+ 28%

Резюме на рынке

Конкуренция среди соискателей выросла, особенно в junior-сегменте: 2 300 откликов на 1 вакансию

hh.ru · vc.ru, 2025

10x

Разрыв зарплат junior → Lead AI

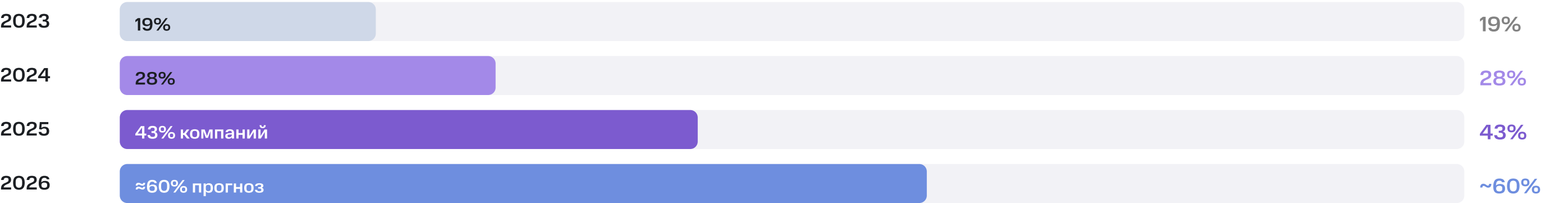
Junior: 60–85 тыс. ₽ · Lead с AI-опытом: до 700 тыс. ₽. Рынок поляризован

vc.ru, Sep 2025

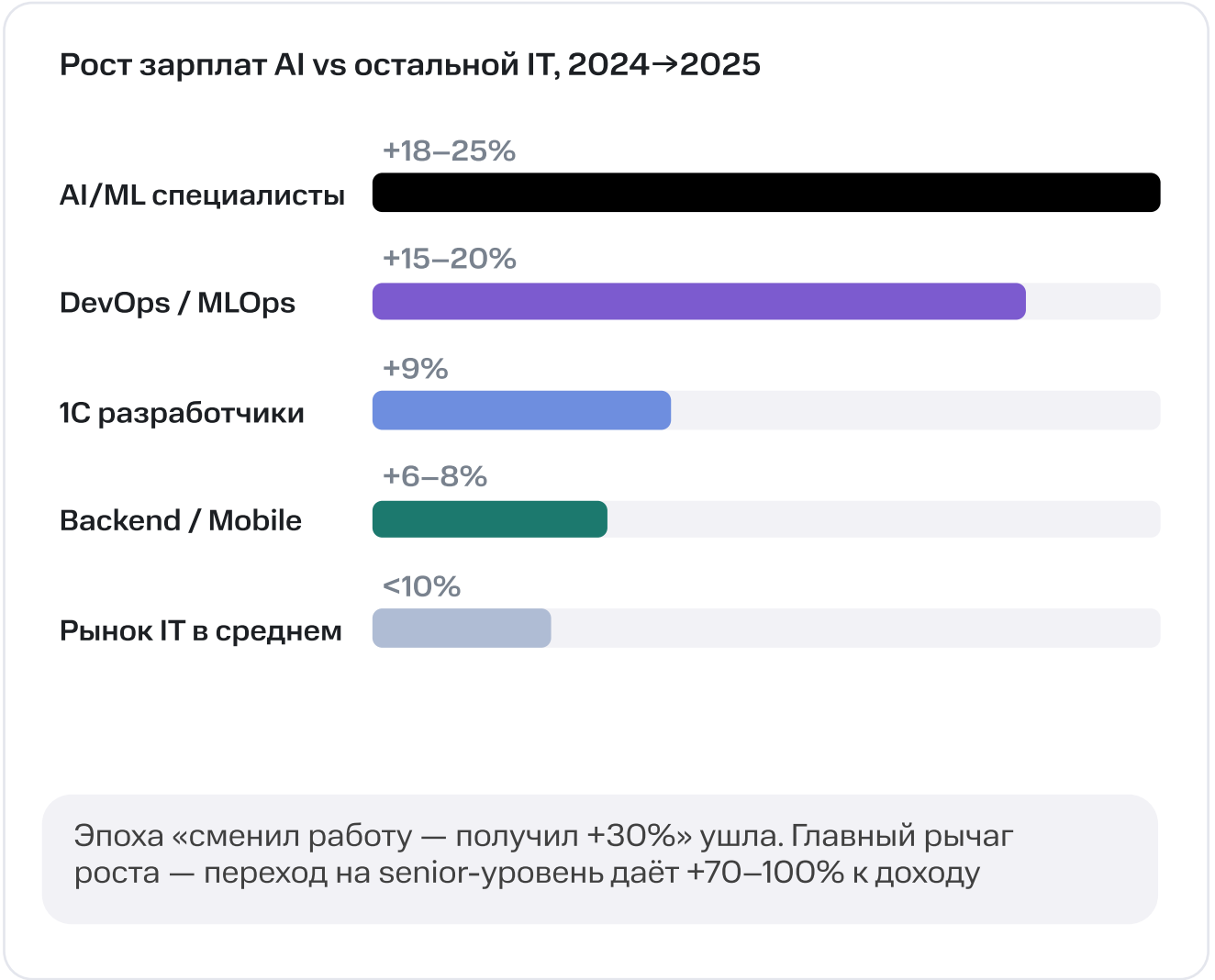
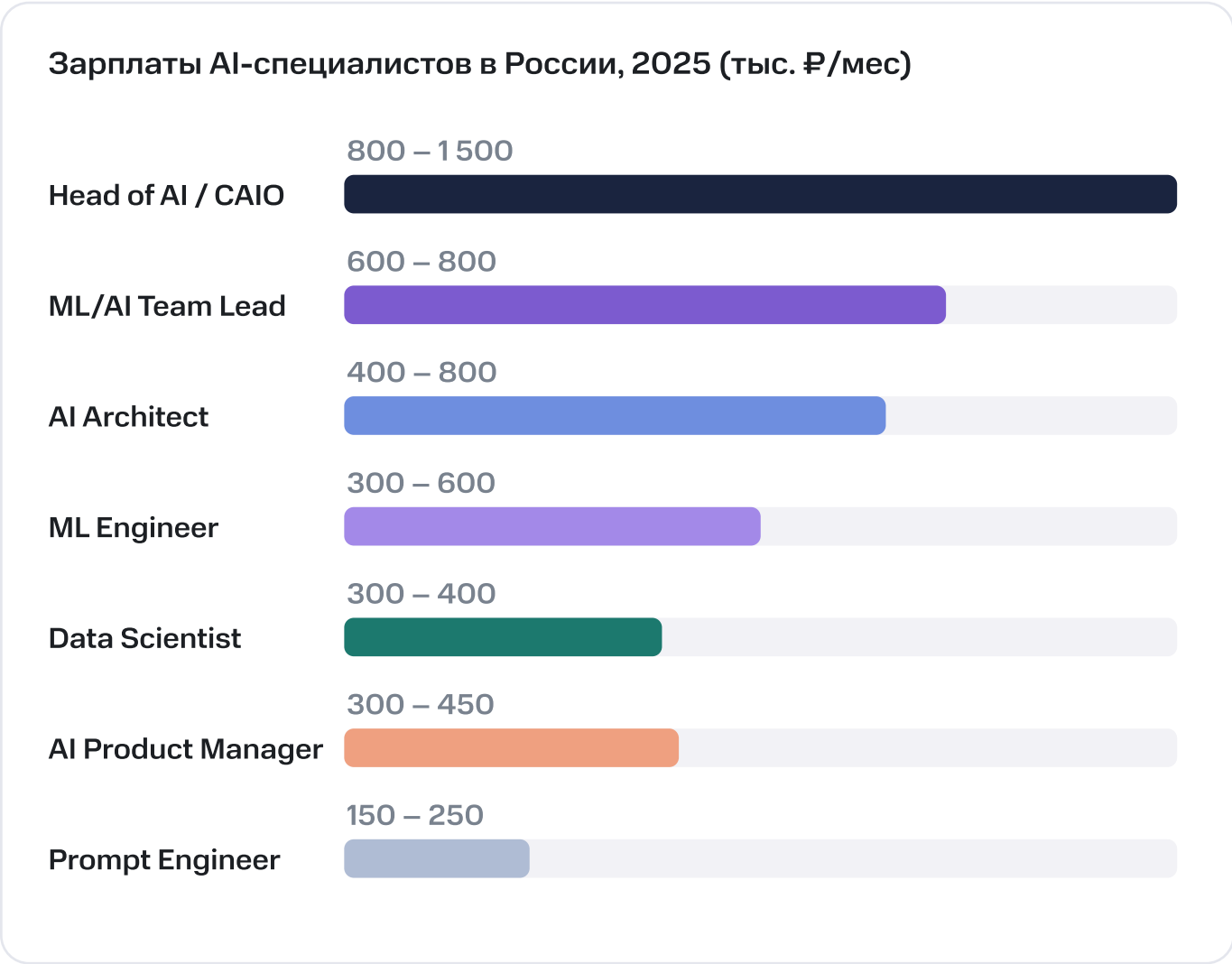
Через 3 года: рынок mid-уровня останется конкурентным, рынок senior AI-специалистов будет дефицитным хронически. Реальная нехватка IT-кадров уже сейчас оценивается в 500–700 тыс. человек и может достичь 1 млн к 2027 году.

Тренд 2 — AI как базовый навык

Доля крупных российских компаний, использующих AI-решения



Источник: Habr, октябрь 2025



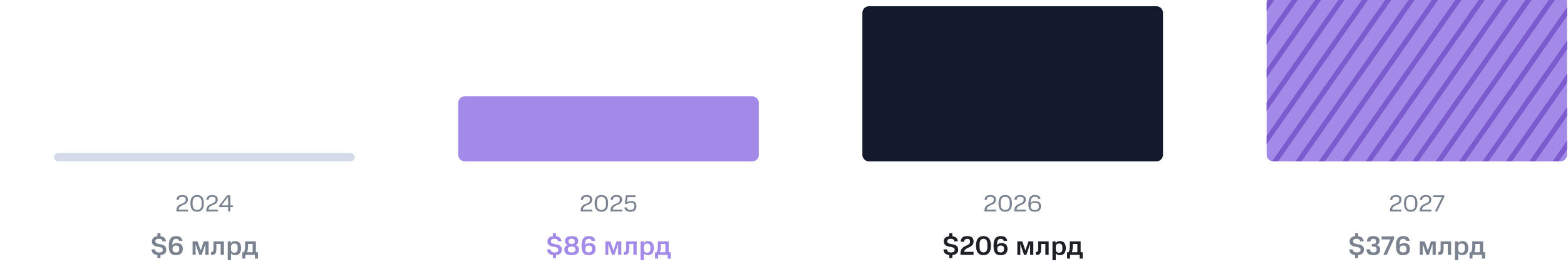
ТРЕНД 3 — АГЕНТНЫЙ AI: ОТ ПИЛОТОВ К ИНФРАСТРУКТУРЕ

Эволюция AI-агентов

От чатботов к автономным мультиагентным системам



Расходы компаний на AI-агентов — мировой рынок (млрд \$)



Источник: Gartner, 2025. 2027 — прогноз

Для AI-трансформации клиента: компании-лидеры (Авито, 2ГИС, Магнит, Альфа-Банк) уже собрали первые измеримые результаты. Те, кто ещё не начал — пытаются догнать. Разрыв между группами будет расти ежеквартально.

ТРЕНД 4 — ИМПОРТОЗАМЕЩЕНИЕ: УСТОЙЧИВЫЙ СПРОС ДО 2028

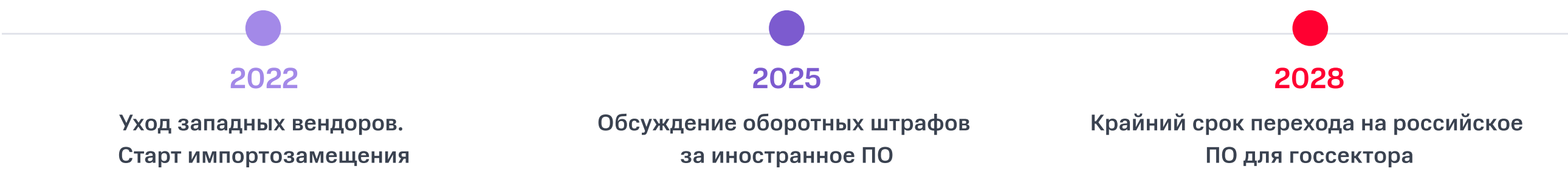
PostgreSQL
вместо Oracle / MS SQL
+25%
спрос на DBA, 2 года

Astra Linux
вместо Windows Server
×3
к зарплате vs рынок

1C разработка
вместо SAP / Oracle ERP
+9%
зп mid за полгода

VK / Yandex Cloud
вместо AWS / GCP
+14%
рост рынка ПО, 2025

Дедлайны и регуляторные триггеры

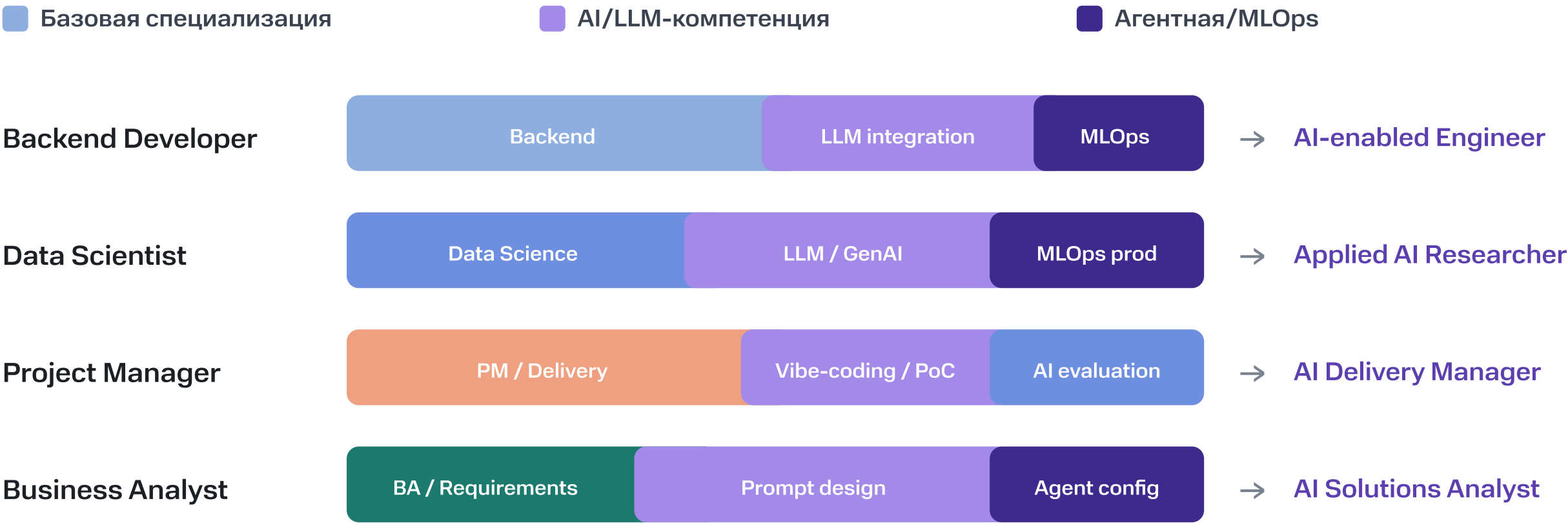


Для AI-команд: весь MLOps и LLMOps строится на суверенном стеке — VK Cloud, Yandex Cloud, SberCloud. Навык проектирования AI-систем на отечественной инфраструктуре становится конкурентным преимуществом уже сейчас.

ТРЕНД 5 — ГИБРИДИЗАЦИЯ: СМЕРТЬ «ЧИСТЫХ» РОЛЕЙ

Трансформация ролей к 2027–28

Что добавляется к базовой специализации — и как меняется название роли



«Специалист с AI работает за двоих и стоит дороже» — уже не метафора. Через 3 года компании, нанявшие 5 AI-усиленных специалистов вместо 10 обычных, получают структурное преимущество.

Что сломалось

- AI-автоотклики создают тысячи нерелевантных резюме
- ATS-фильтры отсекают хороших кандидатов по ключевым словам
- Кандидаты используют AI для прохождения технических собеседований
- hh.ru работает для 15% вакансий, остальные — через нетворкинг
- Senior-кандидат не ждёт ответа дольше 5–7 рабочих дней

Что работает

- Реферальная программа — 34% всего найма (наш опыт 2025)
- Конференции: AI Journey, AIDAO, ICLR, NeurIPS
- Прямой поиск в академической среде (МФТИ, ВШЭ, ИТМО)
- Executive search для Head of AI и Team Lead
- Скорость решений: 5 дней max от резюме до оффера для senior

Через 3 года: HR-функция разделится на два трека. Массовый найм — полностью автоматизирован через AI-агентов. Экспертный найм (AI-специалисты, team lead, архитекторы) — отношения, нетворкинг, прямой поиск. Рекрутер, умеющий только первое, будет заменён агентом.

Итог

ТЕНДЕНЦИЯ	ЭФФЕКТ СЕЙЧАС	ЧЕРЕЗ 3 ГОДА	РИСК ДЛЯ AI-КОМАНДЫ
Охлаждение рынка	Легче нанять mid, сложнее senior	Senior-дефицит нарастает, нехватка до 1 млн	Не выстроена система выращивания
AI — базовый навык	В каждой 2-й вакансии требование AI-грамотности	«Без AI» = нет на рынке; зарплатный разрыв ×10	Нанимают «классиков» без AI-апгрейда
Агентный AI	Первые production-внедрения у лидеров	\$376 млрд рынок; мультиагенты как стандарт	Нет архитекторов агентных систем
Импортозамещение	Суверенный стек обязателен уже сейчас	Дедлайн 2028, оборотные штрафы	Команда не знает российской инфраструктуры
Гибридизация ролей	Чистые роли теряют ценность	5 AI-усиленных > 10 обычных	Нет стратегии апскилинга команды
Сломанный рекрутинг	hh.ru работает только для mid/junior	HR делится: массовый (агент) vs экспертный (человек)	Нет реферальной программы и нетворкинга

7,9–12,8

трлн ₽ в год к 2030

Ожидаемый эффект AI для экономики России
Яков и партнёры · Яндекс, декабрь 2025

5,5%

прогнозного ВВП

Вклад AI к 2030 году по базовому сценарию

2025

оптимальное время входа

Компании, начавшие в 2023–25,
собирают результаты.

14

Выводы для ЛПР

Что каждый из шести трендов означает для HR-директора и директора по цифровой трансформации — и какие решения стоит принять уже сейчас.

Тренд 1. Охлаждение рынка

Что происходит. Число IT-вакансий упало на 34%, резюме на рынке стало больше на 28%. Это создаёт иллюзию, что нанимать стало легче.

Для HRD. Иллюзия справедлива только для junior и mid-уровня. Рынок senior AI-специалистов остался дефицитным и стал жёстче — потому что крупные игроки (Яндекс, Сбер, MTS AI) в этот период не сокращали найм в AI, а наращивали. Если ваша компания замедлила найм AI-специалистов в 2024–2025 годах, рассчитывая «нанять потом подешевле» — это ошибка, которая обнаружит себя в 2026–2027 году ростом стоимости найма и удлинением сроков закрытия.

Для CDO. Охлаждение рынка — окно возможностей для выращивания. Сейчас относительно доступны сильные mid-специалисты с потенциалом роста в senior. Компании, которые создадут внутренние треки развития AI-компетенций в 2025–2026 году, получают через 2–3 года собственных сеньоров по цене их онбординга, а не рыночного найма.

Стратегический вывод. Не откладывайте найм ключевых AI-ролей в ожидании «когда рынок ещё остынет». ML/AI Team Lead и AI Architect — это 60–90 дней закрытия в лучшем случае даже сейчас. Если вы начнёте искать, когда появится задача, вы опоздаете минимум на квартал.

Тренд 2. AI как базовый навык

Что происходит. 43% крупных российских компаний используют AI-решения в production. В каждой второй вакансии — требование AI-грамотности. Зарплатный разрыв между junior и Lead с AI-опытом достиг 10×.

Для HRD. Ваши профили вакансий, написанные до 2023 года, скорее всего не отражают реальные требования к AI-компетенциям. Если в JD для Data Scientist нет требований к LLM, RAG и production MLOps — вы нанимаете специалиста прошлого поколения. Пересмотр профилей вакансий и системы грейдирования — задача не HR-администрирования, а конкурентного позиционирования на рынке.

Для CDO. Запрос «нам нужен Data Scientist» без уточнения AI-стека означает, что вы получите специалиста, который умеет строить классические модели, но не умеет работать с LLM в production. Эти два профиля сейчас стоят принципиально по-разному и закрывают разные задачи. Размытое техзадание на найм — это прямые потери: человек выйдет на работу и обнаружится несоответствие через 2–3 месяца.

Стратегический вывод. Обновите матрицы компетенций и профили вакансий прямо сейчас. Разрыв между «специалистом с AI» и «специалистом без AI» будет только расти — и этот разрыв уже влияет на offer-left в пользу конкурентов.

Тренд 3. Агентный AI

Что происходит. 2026 год называют годом агентов. Мировой рынок AI-агентов: \$86 млрд в 2025 году → \$376 млрд к 2027-му. Компании-лидеры уже строят мультиагентные системы в production.

Для HRD. В ближайшие 1–2 года появится запрос на роли, которых сейчас нет в классических штатных расписаниях: инженер-оркестратор агентных систем, специалист по AI Governance, эксперт по безопасности LLM-систем. Сейчас их нет на рынке в виде готовых профилей — значит, нанять «с рынка» не получится. Единственный источник — выращивание изнутри или переквалификация существующих специалистов. Если HR-стратегия не предусматривает этого, к 2027 году компания окажется перед структурным дефицитом.

Для CDO. Агентный AI меняет вопрос с «нам нужен ML-инженер» на «нам нужен человек, который умеет проектировать автономные системы, работающие без постоянного надзора». Это требует иного уровня инженерной зрелости и принципиально других архитектурных решений. Если ваш технический стек и люди не готовы к этому переходу — конкуренты, которые готовятся сейчас, оторвутся на дистанции в 2–3 года.

Стратегический вывод. Начните формировать внутреннюю экспертизу в агентных системах уже в 2025–2026 годах. Конкуренция за этих специалистов к 2027 году будет острее, чем конкуренция за LLM-инженеров сегодня.

Тренд 4. Импортозамещение

Что происходит. Дедлайн перехода на российское ПО для госсектора — 2028 год. Обсуждаются оборотные штрафы. Спрос на специалистов по российскому enterprise-стеку растёт быстрее, чем их предложение.

Для HRD. Для компаний, работающих с госсектором или в регулируемых отраслях, это означает: специалисты с экспертизой в VK Cloud, Yandex Cloud, Astra Linux, отечественных СУБД и импортозамещённом MLOps-стеке стоят сейчас дорого — и будут стоить ещё дороже к 2027–2028 году. Если вы не формируете этот пул сейчас, вы будете нанимать в условиях острого дефицита и платить за срочность.

Для CDO. AI-команда, умеющая работать только на AWS/GCP/Azure, — это операционный риск для бизнеса уже сегодня. Не только потому что дедлайны 2028 года реальны, но и потому что всё enterprise-внедрение AI в России идёт через российскую инфраструктуру. Если в вашей команде нет компетенции по суверенному AI-стеку — вы не можете участвовать в значительной части рынка.

Стратегический вывод. Оцените текущую компетенцию вашей AI-команды в части российской инфраструктуры. Это не абстрактный риск — это критерий допуска к сделкам уже сейчас.

Тренд 5. Гибридизация ролей

Что происходит. «Чистые» роли теряют ценность. Backend-разработчик с LLM-интеграцией, Data Scientist с MLOps-навыками, PM, умеющий прототипировать — это не исключение, а новая норма. Специалист с AI работает за двоих и стоит дороже.

Для HRD. Оценка кандидатов по старым профилям ведёт к найму людей, которые через год–два окажутся ниже рынка по компетенциям. Система аттестации и грейдирования должна включать AI-компетенции как отдельное измерение. Если в вашей компании нет внутренних программ AI-апскиллинга, люди уйдут туда, где они есть.

Для CDO. Прагматичный вывод прост: 5 AI-усиленных специалистов продуктивнее и дешевле в обслуживании, чем 10 обычных. Это не значит сокращать команду — это значит принять другую логику найма: меньше людей, выше требования, выше компенсация, выше отдача. Если ваш бюджет планируется в логике «нанять больше людей, когда задач станет больше» — к 2027–2028 году эта модель будет структурно проигрывать конкурентам, работающим в логике AI-усиления.

Стратегический вывод. Запустите внутреннюю программу AI-апскиллинга раньше, чем это сделают ваши конкуренты. Обучить своего хорошего инженера работе с LLM-инструментами — дешевле и быстрее, чем нанять готового специалиста с рынка.

Тренд 6. Сломанный рекрутинг

Что происходит. До 85% вакансий закрывается через нетворкинг, не через job-борды. AI-автоотклики создали тысячи нерелевантных резюме. Senior-кандидат не ждёт ответа дольше 5–7 рабочих дней.

Для HRD. Реферальная программа — не приятный бонус, а критическая инфраструктура найма AI-специалистов. 34% лучших наймов приходят через рекомендации. Если программы нет или она не работает — вы системно проигрываете конкурентам в доступе к лучшим кандидатам. Второй приоритет — скорость: внутренние согласования дольше 5–7 дней на любом этапе приводят к прямым потерям senior-кандидатов. Это не гипотеза, это задокументированный паттерн.

Для CDO. Нанимающий менеджер, который смотрит резюме раз в неделю и «не может сейчас» — самый дорогостоящий участник воронки найма. Стоимость потери одного AI Architect или ML Team Lead из-за задержки ответа — это 2–3 месяца повторного поиска плюс потеря первого кандидата конкуренту. Это решается организационно: SLA на обратную связь как внутренний стандарт.

Стратегический вывод. Установите внутренний SLA на обратную связь по AI-вакансиям: 48 часов на первичный ответ, 5 рабочих дней на решение после финального интервью. Для senior-ролей — это не опция, это гигиена.

Общий вывод для ЛПР

Все шесть трендов указывают в одном направлении: окно для формирования AI-команды с разумными затратами закрывается. Компании, которые начали в 2023–2025 годах, входят в рост рынка с готовой командой и накопленной экспертизой. Компании, которые начнут в 2027–2028 году, будут нанимать в условиях острого дефицита, высокой конкуренции и сформировавшегося рынка — с принципиально другими затратами на найм и удержание.

Три решения, которые стоит принять прямо сейчас

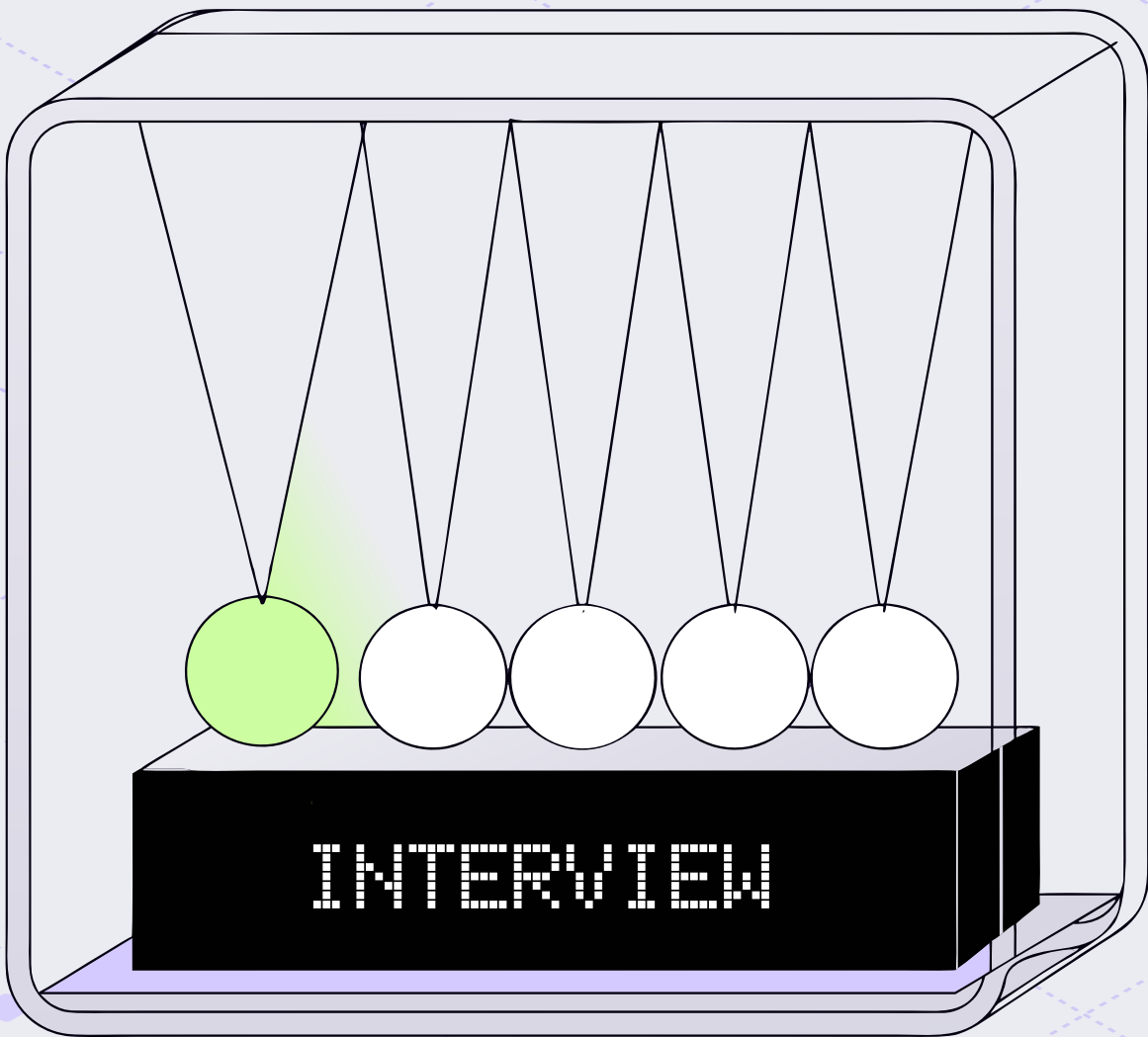
- 1 · Обновить профили вакансий и систему грейдирования с учётом AI-компетенций — до того, как начнётся следующая волна найма.
- 2 · Запустить реферальную программу и установить SLA на обратную связь — это самые дешёвые инструменты повышения эффективности найма из доступных.
- 3 · Определить стратегию выращивания vs найма для ролей, которые к 2027 году появятся на рынке в дефицитном режиме (агентные инженеры, AI Governance, MLOps на суверенном стеке).

Интервью

Глубинные интервью с техническим и продуктовым лидерами AI-направления и их структурный анализ.

В ЭТОМ РАЗДЕЛЕ

- 1Интервью с СТО — Артур Сахаров
- 2Интервью с СРО — Максим Волошин
- 3Интервью с САЮ — Даниил Киреев
- 4Анализ интервью



15

Интервью с СТО · Артур Сахаров

Вопрос 01 · Рынок

Как ты оцениваешь рынок AI-специалистов? Что изменилось за последний год?

Главный сдвиг. Раньше competitive advantage был в модели. Сегодня — в архитектуре решения, eval-дисциплине и умении встраивать AI в реальные бизнес-процессы с адекватной стоимостью.

Что изменилось за год

Год назад

Учились писать «правильные промпты». Промпт-инжиниринг как отдельная роль — в моде. Модель = competitive advantage. Спрос на «промптеров из редакторов».

Сегодня

Claude и ChatGPT сами генерируют хорошие промпты лучше человека. Роль «чистого промптера» уходит. Нужны люди, строящие агентов, графы вызовов, оркестрирующие модели.

Пять ключевых изменений

Промпт-инжиниринг потерял актуальность. Роль «чистого промптера» уходит. Модели сами понимают плохо написанные запросы — техническое мастерство в формулировке промптов перестало быть дифференцирующим навыком.

Выросли ожидания от качества сервисов. Пользователи сравнивают внутренние AI-решения с ChatGPT: задержки меньше секунды, любые файлы, идеальный UX. Корпоративные системы пока проигрывают — и это создаёт давление на AI-команды.

Спрос сместился на AI-инженеров, а не MLE. Нужны люди с техническим бэкграундом и опытом программирования логики — строящие агентов, графы вызовов, оркестрирующие модели. Не «промптеры из редакторов».

Инфраструктура стала узким местом. DevOps и безопасность не готовы к AI. Корпоративные политики блокируют Claude Code, доступ к моделям, работу с данными. Без подготовленной инфраструктуры AI-команда не взлетит.

Разработчики массово переходят на AI-ассистентов. Если инженер с зарплатой X тратит на токены менее 30–40% от X — он работает неэффективно. AI-ассистенты должны быть встроены в процесс разработки как CAD-системы в инженерии.

Главный вопрос сегодня — не «какую модель взять», а «как построить пайплайн, который работает быстро, предсказуемо и дёшево».

Вопрос 02 · Роли

Какие роли нужны в офисе AI у клиента?

Важно: не пытайтесь создать «отдельную AI-disrupt-команду как во времена IT-департаментов». Это переходная структура. Состав команды радикально зависит от выбранного пути.

Сценарий А — продуктовая компания встраивает AI в команды

Роль	За что отвечает
Head of AI (lite)	Стратегия, приоритеты, отношения с вендорами
AI-инженеры (2–3)	RAG, агенты, eval, интеграции. Поддержка внутренней платформы
Embedded AI-инженеры	Внутри бизнес-команд, сидят рядом с пользователями и продактами
DevOps / Platform	Инфраструктура, доступ к моделям, безопасность
Evaluation Engineer	Golden sets, регрессия, детекция дрейфа качества
Образовательная функция	Обучение сотрудников работе с AI-инструментами

Сценарий В — покупка платформы

Не нужны: разработчики самой платформы.

Нужны: AI-инженеры для кастомизации и интеграции · PM / бизнес-аналитики, умеющие формулировать задачи для AI · инфраструктура (данные, безопасность).

Ключевое про платформу: платформа — это не «проигрыватель агентских сценариев», это набор инструментов для работы с данными, обучением, развёртыванием, поддержкой. Кастом — только под уникальные бизнес-процессы, которых у вас скорее всего нет.

Сценарий С — «делаем всё сами с нуля»

Практически никогда не нужно. Вы отстанете от рынка быстрее, чем соберёте стратегию. Исключения: foundation lab (единицы в мире) или объём задачи такой, что зрелые рыночные решения не подходят (редко). **В 90% случаев: возьмите готовую платформу и сфокусируйтесь на бизнес-логике.**

Три категории людей в компании

Категория	Кто это	Что нужно
Строят AI-сервисы	AI-инженеры, MLE	Глубокая техническая экспертиза
Используют AI-ассистентов	Обычные разработчики	Доступ к Claude Code, Copilot и аналогам; привычка оркестрировать агентов
Обычные сотрудники	Маркетологи, юристы, аналитики	Навыки работы с AI-инструментами, обучение

Вопрос 03 · Требования и матрица зрелости

Какие требования к команде клиента?

Самое важное: не думайте, что вы наймёте «супер-руководителя AI» — и всё поедет. Это миф. 70% выпускников программ цифровой трансформации меняют компанию в течение 2 лет, потому что в старой компании невозможно ничего изменить.

Что важнее правильной роли

- Выдать руководителю реальные полномочия — не декларативные
- Пересмотреть KPI и процессы — руководитель AI должен быть готов менять топ-менеджмент и цели, а не только «руководить разработкой»
- Инвестировать в обучение существующих сотрудников — не только нанимать новых
- Принять, что часть людей не впишется в новую реальность

Матрица зрелости компании

L1	Хаотичные эксперименты	Кто-то пробует ChatGPT, нет стратегии, нет ответственных, нет бюджета
L2	Отдельные пилоты	1–3 PoC запущены, но нет процесса оценки и внедрения. Большинство компаний — здесь или между L2 и L3
L3	Платформа + интеграция	Куплена платформа, есть AI-инженеры, eval. Довели хотя бы один продукт до прода / клиентов
L4	AI в каждой команде	Продуктовые команды сами встраивают AI-фичи, инженеры используют лучшие инструменты по умолчанию
L5	AI как стандарт	Без AI невозможно работать. Это часть культуры, а не отдельная инициатива

→ Большинство компаний находятся между L2 и L3. Цель — не прыгнуть сразу на L5, а пройти каждый уровень осознанно, не пропуская этапы.

Вопрос 04 · Скрининг и собеседование

Как проверить, что кандидат реально соответствует?

Главный принцип: проверяйте не технические детали, а способность внедрять изменения и управлять неопределённостью.

На что смотреть в резюме

- Похожий опыт в похожей компании. Размер, индустрия, стадия зрелости — важнее, чем просто «работал с LLM»
- Трек-рекорд изменений, а не присутствия. Не «работал», а «изменил». В чём именно?
- Как долго жил в предыдущих местах. AI-сектор лихорадит — частые прыжки это флажок, а не показатель широкого опыта

Ключевые вопросы на собеседовании

«Расскажи, что не получалось — и как ты с этим справлялся». Если кандидат говорит только об успехах — скорее всего не доводил дело до конца.

«Как ты встраивал изменения в компанию, которая не хотела меняться?». Ключевой вопрос для руководителя AI-инициатив. Нет примера — нет опыта внедрения.

«Кого ты вырастил? Кто пошёл за тобой?». Проверьте по LinkedIn. Сильные люди идут за лидером — это проверяемо.

Самая частая ошибка найма: нанять технического гения без навыка внедрения изменений и опыта успешного вывода продуктов на рынок. Он построит красивую систему / PoC-гипотезы — которыми никто не сможет пользоваться.

Вопрос 05 · KPI

KPI у команды PoC и у команды промышленного внедрения

95% PoC не доходят до production. Не потому что технически провалились — а потому что при попытке вывести в прод экономика ломается, качество падает, архитектура не масштабируется.

Команда PoC / Discovery

Что мерить

Количество проверенных гипотез
Скорость от идеи до «работает / не работает»
Kill rate 40–60% — это норма, не провал

Что не мерить

Uptime и стабильность
Latency
Соответствие SLA

Команда Production

Что мерить

Качество модели (eval drift)
Reliability: uptime, latency,
SLA Unit economics: стоимость одной транзакции (токены + инфраструктура)
Adoption: DAU, MAU

Что не мерить

Количество новых гипотез
Скорость PoC-итераций

4 причины, почему PoC не доходит до прода

- 01 · **Качество не масштабируется.** На 10 сценариях работает, на 100 — уже нет. А те первые десять начинают работать хуже.
- 02 · **Добавление ломает старое.** Архитектура не выдерживает расширения сценариев. Каждое новое требование ломает то, что работало.
- 03 · **Экономика ломается.** Inference становится слишком дорогим для того качества и скорости, которые необходимы для финансового эффекта.
- 04 · **Инфраструктура не готова.** Мониторинг, безопасность, SLA. Месяцами не можем запустить даже небольшие продуктовые сценарии.

Вопрос 06 · Руководитель AI-направления

Три главных качества для Head of AI

Готовность менять топ-менеджмент и процессы

Это не технический лид. Это человек, который будет «морковкой спереди и сзади» для других руководителей. Без этого качества AI-трансформация не происходит

Понимает: успех не в модели, а во внедрении

Он должен говорить «нет» в 10 раз чаще, чем «да». Приоритизация и отказ от неперспективных направлений — ключевой навык

Способность привлекать таланты

Сильные люди идут за лидером, а не за брендом. Это проверяется: «Пошёл бы ты к нему снова?»

Как проверить на собеседовании

- «Расскажи, как ты менял компанию, которая не хотела меняться» — нет конкретного примера = нет опыта трансформации
- «Назови проекты, которые ты закрыл — и почему» — умение закрывать неперспективное важнее умения запускать всё подряд
- Reference-звонки: с бывшими подчинёнными — «Пошёл бы ты к нему снова?», с руководителями — «Что получилось, что не получилось?»

На позицию Head of AI лучше не нанимать с рынка, а найти кого-то, кто уже успешно внедрял изменения и руководил проектами внутри компании или в близком контексте.

Главные инсайты и взгляд в будущее

Четыре инсайта — прямо сейчас

- Модель «одна центральная AI-команда» умирает. AI становится функцией внутри продуктовых команд — так же, как произошло с выделенными IT-департаментами
- Консалтинг «дадим 200 страниц рекомендаций и уйдём» не работает. Нужно идти вместе с клиентом и делать руками
- Желание нанять «супергероя» и забыть про AI — иллюзия. Трансформация требует участия топ-менеджмента и инвестиций в обучение
- Профессии не исчезнут — изменятся. Как дизайнеры перешли с бумаги на цифру, так аналитики, юристы и маркетологи начнут использовать AI-инструменты как часть работы

Взгляд на 5–10 лет

- AI встроен во все бизнес-процессы — как сегодня интернет и специализированный софт
- Команды меньше, но каждый сотрудник продуктивнее. Начинающих позиций станет значительно меньше
- Отдельной профессии «AI-инженер» скорее всего не останется — станет частью роли разработчика или продакта
- Централизованные AI-отделы исчезнут — останутся только платформенные команды и центры компетенций, как сейчас в продуктовой разработке

Вопрос 08 · Три совета

Компания строит AI-функцию с нуля — три совета

1 · Не начинайте с найма «звезды». Сначала поймите, готовы ли вы меняться как компания. Иначе звезда уйдёт через год без результатов — а цикл проверки длиной в год это очень дорого. Готовность компании к изменениям важнее любого кандидата.

2 · Начните с обучения и инфраструктуры. Дайте сотрудникам инструменты и доступ, научите использованию. Пусть сами пробуют. AI-культура растёт снизу — и это единственный способ, который работает устойчиво.

3 · Примите, что 80% работы — не в самой умной модели, а вокруг неё. Данные, интеграции, безопасность, экономика, adoption. Без этого AI не работает — какой бы умной ни была модель.

Бонус-совет №4. Не пытайтесь делать всё сами. Готовые платформы уже решили 90% технических проблем. Ваша задача — бизнес-логика и интеграция, сквозное целеполагание и готовность драйвить изменения. Технология — это инструмент, а не цель.

16

Интервью с СРО · Максим Волошин

Вопрос 01 · Рынок

Рынок AI-специалистов. Что изменилось за последний год?

Рынок расслоился. Сверху — ожесточённая война за топ-5% (зарплаты ×2–3 за год). В середине — массовый поток «AI-инженеров», которые научились звать LLM по API. Реальный дефицит — не в моделях, а в людях, которые умеют доводить AI до production: eval-дисциплина, агентные системы, продуктивное чутьё.

7 изменений за последний год

- 1 · Сдвиг с выбора модели к архитектуре решения.** Ключевой вопрос: «Как спроектировать пайплайн, уровни абстракции, взаимодействие агентов». Системное мышление и бизнес-реинжиниринг важнее тонкого тюнинга отдельных LLM.
- 2 · Промпт-инжиниринг автоматизируется.** Современные инструменты генерируют и улучшают промнты лучше большинства людей. Роль «чистого промптера» уходит — ценность смещается к пониманию бизнес-задачи и целостного решения.
- 3 · Сдвиг с тренировки на применение.** Тренировка foundation-моделей — роскошь 5–10 лабораторий в мире. В корпоративе 90% задач — application layer: RAG, агенты, fine-tuning.
- 4 · Появилась роль AI Product Engineer (он же Forward Deployed).** Гибрид: инженер + продакт + частично консультант. Сидит рядом с заказчиком, итерирует с пользователем. За год стала самой востребованной и самой дефицитной.
- 5 · Агенты и multi-agent — отдельная специализация.** Год назад это был research, сегодня — production-skill. Люди, умеющие проектировать · eval · observability агентных пайплайнов, на вес золота.
- 6 · «Чистый Data Scientist» уходит.** Notebook-ML без production-навыка — позиция уровня аналитика. Сильные DS перешли либо в MLE, либо в AI Engineering.
- 7 · Российский контекст.** Отток продолжается, но внутри страны вырос крепкий middle через переобучение (DS · MLE · AIE). Слабость — мало людей с опытом промышленных multi-tenant deployments.
- Главный сдвиг:** ценность сместилась с «кто умеет обучать» на «кто умеет довести модель до бизнес-эффекта и удержать качество в проде». Это меняет всё — от резюме-скрининга до структуры команды.

Вопрос 02 · Роли

Какие роли нужны в офисе AI у клиента?

Ключевое: состав команды радикально зависит от сценария зрелости.
Перепутать сценарий — значит нанять либо избыточно (дорого, медленно), либо недостаточно (демо без production).

Сценарий А — In-house разработка AI-продуктов (промышленные решения)

Клиент строит свой продукт с AI внутри — нужна полная цепочка от данных до прода.

Роль	За что отвечает	Ценность
CAIO / Head of AI	Стратегия, портфель кейсов, ехес-sponsorship, привлечение талантов	Без него — «AI как проект IT»
AI Product Manager	Use case discovery, приоритизация, value tracking	Соединяет AI с P&L
AI / ML Architect	Платформа, MLOps-фундамент, выбор моделей	Задаёт пол качества, решает build vs buy
Senior ML Engineers (2–4)	Lifecycle моделей: train, serve, monitor	Core delivery muscle
AI Engineers (2–3)	RAG, агенты, prompt-инженерия, интеграции	Скорость до первого результата
Data Engineers	Пайплайны, feature stores, контракты данных	Без них — AI на песке
MLOps / Platform	Inference infra, observability, cost control	Делает production возможным экономически
Evaluation Engineer	Eval-наборы, метрики качества, regression-test	Самая недооценённая роль 2026 года
Applied Scientist (1)	Fine-tuning, custom модели под домен	Только если домен требует
AI Safety / Compliance	Red team, ответственность модели, регуляторика	Критично в банках, медицине, госе
Embedded Domain Experts	Носители знания процесса, валидаторы	Без них модель учится на мусоре

Сценарий В — PoC + внедрение силами партнёра

Клиент закупает экспертизу, но не «отдаёт мозги наружу». Внутри нужен крепкий заказчик.

Роль	За что отвечает	Ценность
Head of AI (lite)	Бизнес-кейс, portfolio, отношения с вендорами	Лицо, принимающее решения
AI PM / Use Case Owner	Discovery, формулировка задач, измерение value	Не даёт вендору сделать «красиво, но не туда»
AI-fluent Solution Architect	Требования, выбор вендоров, интеграции	«Технический переводчик» с подрядчиком
Data Steward / Owner	Доступ к данным для партнёра, права	Без этого пилот стоит 6 месяцев
Integration / Change Lead	Передача в эксплуатацию, change management	Здесь умирает 70% пилотов
Business Owner (must-have)	Adoption, ROI tracking, защита у спонсора	Без него — нет внедрения, только отчёт
AI Procurement Partner	SLA, ценообразование, exit-стратегия	Защита от vendor lock-in

Самая частая ошибка найма: клиент не нанимает «крепкого заказчика» — и через год обнаруживает полную зависимость от вендора во всём: от данных до решений.

Сценарий С — Свои агенты для PoC + поддержка моделей в проде

Реалистичный сценарий для большинства компаний в 2026.

Роль	За что отвечает	Ценность
Head of AI	Бэклог кейсов, talent, отношения наверх	Один человек. Без него не взлетит
AI Product Manager	Discovery, приоритизация, метрики	1 на 2–3 кейса
AI Engineers (2–3)	RAG, агенты, prompt-eng, eval	Core. Не MLE — другой профиль
Integration Engineer / Техлид	Интеграции с корп. системами, оркестрация AI-инстансов на платформах	Часть ручной разработки замещается платформами
Data Engineer (light)	Подключение источников, чистка, индексация	Иногда shared
LLMOps / Platform Engineer	Деплой, мониторинг, стоимость токенов	Прод без него — бомба замедленного действия
Evaluation Engineer / QA	Golden sets, regression, hallucination tracking	Самая критичная новая роль
Annotation / Eval Curator	Eval-датасеты, разметка edge cases	Может быть part-time или вендорится

Минимальная жизнеспособная команда: 5–7 человек.
Меньше — нет покрытия eval/ops. Больше — преждевременная оптимизация.

Вопрос 03 · Требования + матрица зрелости

Требования к команде. Матрица зрелости AI-программы.

Требования по сценариям

Сценарий А · IN-HOUSE INDUSTRIAL Опыт production AI на масштабе Сильный MLOps и data foundation Зрелые code review & eval практики Опыт безопасности AI: ABAC, audit Минимум 2 уровня сениоров в team	Сценарий В · РОС + ПАРТНЁР Бизнес-переводчики сильнее инженеров Умение писать ТЗ для AI Vendor management и SLA-навык Data governance — кто что отдаёт Adoption-функция — отдельная роль	Сценарий С · СВОИ АГЕНТЫ + OPS AI Engineering, не классический ML Eval-дисциплина с первого дня Прагматизм: build vs buy на каждом шаге Cost-awareness: токены — это P&L Скорость итерации важнее красоты
--	---	--

Матрица зрелости AI-команды (self-assessment)

	LO Discover	L1 Pilot	L2 Productize	L3 Scale	L4 Transform
Стратегия & портфель	Идеи на бумаге	1 пилот в активе	2–3 кейса в проде	Портфель с приоритизацией	AI как OS компании
Талант	1 энтузиаст	Малая команда, single-thread	Кросс-функциональные pods	Career ladders, retention	Внутренняя школа AI
Данные	Хаос, ручная сборка	Источники ad-hoc	Контракты данных, каталог	Self-service для AI-команд	Real-time, feature store
Платформа / MLOps	Ноутбуки	Один inference endpoint	CI/CD моделей, мониторинг	Multi-tenant, cost control	Полная observability
Eval & качество	Глазами	Manual spot-check	Golden sets, regression	Online eval, A/B на качестве	Self-improving eval loops
Risk / Safety	Не думали	Чек-листы	Policy + audit logs	Red team, incident response	AI risk фрейм как у банков
Adoption / value	Демо в PowerPoint	Один use case в работе	Измеримый ROI ≥ 1 кейса	Value tracking по портфелю	AI в KPI бизнес-единиц

Большинство российских корпоратов в 2026 — между L1 и L2. Прыжок через ступени не работает: попытка идти из L1 сразу в L3 ломается на отсутствии eval-дисциплины и data foundation.

Вопрос 04 · Скрининг и собеседование

Как проверить, что кандидат реально соответствует?

Главный фокус при скрининге — не технические детали, а shipping-опыт, eval-дисциплина и способность работать с неопределённостью.

На что смотреть в резюме в первую очередь

- **Что довёл до прода, а не «работал над».** Конкретика: какая модель, какой масштаб, какая метрика сдвинулась.
- **Метрики бизнеса, не только технические.** «Снизил latency на 40%» хорошо. «Принёс ₺X выручки / сэкономил Y часов» — лучше.
- **Tenure ≥ 18 месяцев на ключевых ролях.** 4 работы за 2 года в hype-зоне — флажок (но не приговор).
- **Внешние артефакты.** Open source, посты, выступления, Kaggle, eval-наборы. Сигнал любопытства и реального вклада.
- **Для лидеров — кого они вырастили.** Найди этих людей в LinkedIn и спроси напрямую.

Красные флаги в резюме

- **Buzzword bingo без специфики:** «LLM, RAG, agents» без названий моделей и метрик.
- **3+ работы за год.**
- **«Sole AI architect» в большой компании** — обычно неправда.
- **Только PoC,** нигде нет «launched» или «in production».
- **Нет метрик и мониторинга** в описании проектов.
- **Только академический бэкграунд** без shipping-опыта.

Ключевые вопросы на собеседовании

«Расскажи про последний AI-проект — как принимал решения, что пошло не так, что бы сделал иначе.» Здесь видна глубина за 15 минут.

«Покажи свой последний фейл.» Кандидаты, которые рассказывают только успехи — либо junior, либо лгут. Зрелость = умение разобрать провал.

Eval-вопрос: «Откуда ты знал, что модель работает в проде?» Ответ «по фидбеку пользователей» — слабо. Есть golden set, regression, calibration — сильно.

Cost & trade-off: «Сколько ваш сервис стоил в токенах? Как принимали решение между fine-tune и prompt?»

«Что в твоём подходе изменилось за последний год?» AI движется быстро — закостеневших нанимать опасно.

Лучший способ проверки из доступных: платный двухнедельный paid trial на реальном кусочке работы. Стоит дороже собеседования, но дешевле плохого найма в 50 раз. Особенно для senior+ ролей.

Культурный фит (часто забывают)

- Готовность сидеть с пользователем, а не только за ноутбуком
- Толерантность к ambiguity — в AI часто «мы пока не знаем, как правильно»
- Скорость итерации важнее желания идеального решения

Вопрос 05 · KPI

KPI у команды PoC и у команды промышленного внедрения

Это две функции с противоположной экономикой. PoC-команду нельзя мерить как production. Перепутаешь — получишь либо «вечные пилоты», либо «парализованную бюрократию».

Команда PoC / Discovery

Что мерить

Гипотезы, проверенные и закрытые за квартал
Time-to-value-signal: недель от идеи до «работает / не работает»
Стоимость проверенной гипотезы (P/гипотеза)
Kill rate 40–60% — здоровая норма
Insights, переданные другим командам
Stakeholder confidence score на каждом этапе

Что не мерить

Uptime и стабильность
Latency p95
Покрытие тестами

Команда Production

Что мерить

Quality SLO: eval drift не падает ниже X
Reliability: uptime, latency p95, error rate
Adoption: DAU/MAU, % покрытия аудитории
Business outcome: P / часов сэкономлено / NPS lift
Unit economics: cost per inference, токены/запрос
Incident response: MTTD / MTTR

Что не мерить

оличество новых гипотез
Velocity новых фич
«Сколько демо показали»

Портфельная логика PoC: венчурная модель

Корпоративный венчур

Выделенный бюджет (например, 100 млн на 1–3 года)
Цель: x10 проектов под x5 эффект — с пониманием, что часть «умрёт»
Kill rate 40–60% — норма
Управление как портфелем, не как набором проектов.

Переход PoC → продакт

PoC — быстро, дёшево, измеримо
После валидации PMF — стабильность, latency, качество, масштабируемость
Без подтверждённого PMF — не переходить к «Series A»
R&D и production — разные команды и цели.

Вопрос 06 · Руководитель AI-направления

Три главных качества для Head of AI

Ключевое: это самая важная роль во всём процессе. Ошибка здесь умножается на всё остальное. Лучше 6 месяцев искать сильного Head of AI, чем за 6 недель нанять «технического руководителя AI-направления» и переделывать функцию через год.

Бизнес-переводчик

Умеет объяснить AI финансовому директору в его терминах — и финансы инженеру. Без этого качества Head of AI превращается в технического спорщика, которого никто не слушает на правлении.

Проверка: «Объясни последний проект как для CFO за 5 минут». Если начинает с архитектуры — нет.

Portfolio judgment

Умеет выбрать 2–3 кейса из 30 предложений и безжалостно закрыть остальные. Большинство AI-программ умирают не от плохого исполнения, а от того, что делают всё одновременно.

Проверка: «Расскажи проект, который ты закрыл, и почему». Нет таких — критично.

Talent magnet

Сильные люди хотят с ним работать. AI-таланты — дефицит, и они идут не на бренд компании, а на лидера. Без этого даже идеальная стратегия умирает в найме.

Проверка: reference-звонки с 2–3 бывшими подчинёнными — «Пошёл бы ты к нему снова?»

Бонус-качество, если повезёт:

комфорт с ambiguity и готовность вести команду через «мы пока не знаем». В AI это половина работы.

Откуда приходят сильные Head of AI

- Трек 1 · Портфельное управление. Опыт инновационных / трансформационных портфелей. Говорят языком бизнеса и метрик, управляют стейкхолдерами.
- Трек 2 · Консалтинг → C-AIO. Партнёры и senior-менеджеры крупных консалтинговых компаний, CDTO. Структурируют проблемы, работают с CEO.
- Трек 3 · R&D / Техлид / PM. Вход из R&D или разработки. Критерий: управление командой и связь с деревом бизнес-метрик, а не только техническая экспертиза.

Наблюдения из практики и видение будущего

Инсайты из практики

- **Самая дорогая ошибка:** нанять технического суперзвезду без продуктового чутья. Построит красивую систему, которой никто не пользуется. ₽20–40М в воздух за год.
- **Дешёвая ошибка, которую легко чинить:** junior без eval-дисциплины. Учится за квартал.
- **Внутренняя мобильность бьёт внешний найм.** Доменный эксперт, выучивший AI, — сильнее, чем AI-эксперт, учаший домен. Особенно в банке, телекоме, нефти.
- **Forward-Deployed модель работает.** Инженер, физически сидящий с пользователем 50% времени, делает больше, чем 3 инженера в отдельной команде.
- **Tenure ≠ качество.** 18 месяцев на правильной команде с правильными задачами бьёт 5 лет в спящем месте. Стек обновляется каждые 6 месяцев.
- **Tenure ≠ качество.** 18 месяцев на правильной команде с правильными задачами бьёт 5 лет в спящем месте. Стек обновляется каждые 6 месяцев.
- **Co-located pods > centers of excellence.** PM + AIE + дизайнер + доменный эксперт в одной комнате — в 3 раза быстрее, чем централизованный AI-центр.
- **Eval — это product moat.** Не модель, не данные — именно команда, умеющая измерять качество, делает разницу между демо и production.

Взгляд на 5–10 лет

Роли, которые вырастут	Роли, которые исчезнут или сожмутся
AI Product Engineer — самая массовая роль Evaluation Engineer — отдельная дисциплина AI Safety / Trust Engineer — alignment в эксплуатации AI Operations / SRE для AI Agent Orchestrator / Workflow Designer AI-fluent Domain Expert — каждый аналитик / врач / юрист Model Economics Analyst — токены как P&L CAIO — стандартная C-level роль	Чистый Prompt Engineer — уйдёт внутрь модели Классический DS с ноутбуками — сольётся с BI Low-level MLE для тренировки — только в foundation-labs Manual annotators в массовом виде «AI Strategist»-консультанты без рук — рынок разделит делателей от ораторов

Форма команды

- **Маленькое ядро + большая периферия.** Core AI-команда — 5–15 человек на крупный бизнес. Но AI-augmented остальная компания — больше. Каждый PM, аналитик, инженер использует агентов ежедневно.
- **Outsourcing-динамика переворачивается.** Рутинный ML автоматизируется. Ценность остаётся в judgment, integration, eval, safety — то, что нельзя отдать наружу.
- **Команды по доменам, а не по моделям.** Сегодня нанимают «NLP-инженера», завтра — «AI-инженера для клиентского сервиса».

Компания строит AI-функцию с нуля — главные советы

- 1 · Сначала Head of AI — потом всё остальное.** Не нанимай ни одного ML-инженера до того, как нашёл сильного лидера направления. Без него команда соберётся «не та», под «не те» кейсы. Через год — переделка с нуля. Сколько искать: 3–6 месяцев — норма. Hire slow on this role.
- 2 · Один кейс с ₽₽ за 90 дней — не три.** Не «AI-стратегия». Не «портфель из 5 пилотов». Один use case с измеримым бизнес-эффектом за 90 дней. Победил — расширяй. Сразу три — провал в трёх местах одновременно. Критерий выбора: высокая частота, измеримая ценность, доступные данные.
- 3 · Eval-мускул с первого дня.** Если ты не умеешь измерять качество — ты не можешь шипать в production. Точка. Evaluation Engineer — первая инвестиция после Head of AI и до полноценного MLE. Минимум: golden set из 100 примеров + регрессионный гейт.
- Бонус-совет №4.** Не нанимай MLE на работу AI Engineer и наоборот. Это разные профессии. Перепутаешь — потеряешь обоих через 6 месяцев.
- Бонус-совет №5.** Разделяй R&D и production. R&D проверяет гипотезы в изоляции, production — «заказчик» задач. Правило 70/30 по времени работает, но крупные новшества — через отдельную ветку. Смешивать — значит получить ни исследований, ни стабильного прода.

Лучший AI-старт, который я видел: 1 Head of AI + 1 AI PM + 2 AI Engineers + 1 Eval Engineer = 5 человек, один кейс, измеримый эффект за 4 месяца. Дальше — масштабирование на доказанной формуле.

17

Интервью с SAIO · Даниил Киреев

Вопрос 01 · Рынок

Как ты оцениваешь рынок AI-специалистов? Что изменилось за последний год?

Главный сдвиг Рынок меняется вслед за технологиями, а в технологиях вырос уровень абстракции. Раньше модели обучали под себя — сегодня фундаментальные модели «из коробки» настолько хороши, что многие «костыльные» роли отмирают, а сам способ работы с технологией меняется.

Уровень абстракции вырос

Раньше

Нужно было самому обучать модели, кастомить узкие модели под задачу. Много ML-исследователей. Промпт-инжиниринг — отдельная востребованная роль.

Сегодня

Фундаментальные модели «из коробки», качество выросло. Промпт-инжиниринг — базовый навык: модель сама доуточняет запрос (метапромтинг). Фокус — на сборке работающих пайплайнов.

Как трансформируются роли

- 1

Промпт-инженеры → AI-инженеры

«Чистый» промптинг становится базовым навыком, как владение Word. AI-инженер берёт готовые модели без дообучения и собирает их в работающий pipeline, который приносит ценность.
- 2

ML-исследователи → ML-инженеры

Задачи внутри компаний всё чаще переходят от исследователей к инженерам. Исследователи нужны там, где делают сами фундаментальные модели или глубокую кастомизацию.
- 3

Размывание границы «ресёрчер / инженер»

Крупные игроки уходят от деления: в Anthropic — Member of Technical Staff, в xAI — «все инженеры». Грань размывается и смещается в сторону инженеринга.
- 4

AI-тренеры и разметчики

Обучать модели нужно реже, но там, где нужны данные для дообучения, нужна разметка — AI-разметчики и AI-тренеры (более профессиональная роль для работы с текстом).
- 5

MLOps / LLMOps

Не всем доступен SaaS по API — многим нужен on-premise, чтобы данные не утекали. Нужны те, кто развернёт и настроит framework.
- 6

Данные: Data Steward / Data Owner

Для эффективных экспериментов данные должны быть качественно структурированы и храниться правильно. Появляются роли с ответственностью за данные.

Из хороших новостей: большинство новых ролей — производные от прежних.
Промпт-инженеру несложно стать AI-инженером — внутри есть матрица перехода. Главный процесс рынка — рескиллинг.

Рынок готов не в полной мере: старые роли уходят, люди перестраиваются в новые.
Но спрос растёт быстрее, чем рескиллинг успевает покрывать, — людей приходится дополнительно добирать с рынка.

Старые роли уходят, новые — производные от них. Ключевой вызов рынка сегодня — не просто найм, а рескиллинг.

Вопрос 02 · Роли

Какие роли нужны в команде у клиента?

Ключевая идея Состав команды — это градиент: чем глубже клиент идёт в AI, тем больше ролей. Базовый слой — AI-инженер, MLOps/LLMOps, дата-инженер и AI-product-менеджер. Глубже добавляются ML-инженер, ещё глубже — ML-исследователь.

Три сценария клиента

Сценарий	Ключевые роли
In-house разработка	Базовый слой: AI-инженер · MLOps/LLMOps · дата-инженер · AI-product-менеджер. Глубже — ML-инженер; ещё глубже — ML-исследователь (для специфичных областей, напр. поиск месторождений). AI-тренеры/разметчики — когда нужна разметка данных для обучения.
PoC силами интегратора	«Клей» на продуктовом уровне (AI-product-менеджер с фокусом на рынок/продукт + приёмка результата и метрики) · «клей» на технологическом уровне (AI-архитектор: встраивание инициатив в базовую IT-архитектуру) · слой данных (Data Owner).
Свои PoC + поддержка в проде	Сценарий интегратора плюс ownership: MLOps/LLMOps для поддержки инфраструктуры. Дальше — по глубине, в пределе доходит до полноценной in-house команды.

Кто управляет командой

Идеальный универсал (бизнес + продукт + технологии) встречается редко — это большая удача. Реалистичная конструкция растёт вместе с глубиной:

- Базовый слой — дуо: AI-Product + Tech
- Глубже (ML-инженеры/исследователи) — триумвират: Product + Tech Lead + ML
- Вся организация — CAIO: связывает бизнес, продукт и технологии (как минимум на уровне метрик). Нужен при росте числа

Это градиент от базовой интеграции до глубокой in-house команды. На старте логично быть ближе к «левому полюсу» — к интеграторам, не делая большую ставку, — и по мере роста AI-культуры смещаться вправо. Скорость зависит от качества реализации инициатив.

Чем глубже AI, тем больше ролей. Начинать стоит с лёгкого состава ближе к интеграторам и наращивать команду по мере зрелости.

Какие требования к команде клиента? Матрица зрелости с точки зрения СТО / CAIO

Базовое требование Требования растут соразмерно ролям и глубине. Неизменный фундамент — слой данных: без цифровизации AI не поедет. Garbage in — garbage out: плохие данные на входе — плохой результат на выходе.

Данные и инфраструктура

Слой данных обязателен. Если его нет — сначала цифровая трансформация (из бумаги в цифру), и только потом AI вокруг этих данных.

Инфраструктура — по необходимости. On-premise и контроль → нужна своя инфраструктура. Облако или разворачивание силами интегратора → требований меньше. Требования растут вместе с ролями: появился ML-инженер → нужна инфраструктура для обучения; появился ML-исследователь → больше железа под эксперименты.

Матрица зрелости: от гипотезы до портфеля

L1	База — метрики и методология экспериментов. Мышление на уровне метрик: какие бывают, почему accuracy нельзя использовать на разбалансированной выборке. Корректная методология: train/test, недопустимо «обучаться на тесте». Must-have.
L2	LLM и агентные системы. Понимание, как устроены LLM, чем базовая ML-модель отличается от агента, зачем нужны мультиагентные системы. Достаточно, чтобы оценивать архитектуру решений.
L3	Машина экспериментов и trade-off'ы. Цепочка экспериментов как проверка гипотез, stage-gate-модель (рано закрывать неудачные, переключать ресурсы). Ощущение trade-off'ов на всех уровнях: latency / throughput / стоимость токена, качество против цены, GPU / TPU / LPU.
L4	Управление портфелем и культура. Насмотренность и критеризация инициатив по фреймворку, оценка вендоров. Предпринимательский, венчурный mindset (выстреливает ~5–10% инициатив, как стартапы). Change management и культура — должны прорасти вниз по команде.

Софт-скиллы и культура (C-level). Любое внедрение AI — это change management: люди сопротивляются новому, ждут «магии». Задача — культура, где хотят пользоваться AI эффективно и не боятся ошибаться (guardrails: например, систему нельзя заставить отправить персональные данные вовне). Предпринимательский mindset прорастает от CAIO вниз по иерархии.

Требования растут вместе с глубиной, фундамент — данные. Зрелость идёт от базы (метрики, эксперименты) до управления портфелем (насмотренность, венчурный подход, культура).

Вопрос 04 · Скрининг и собеседование

Как проверяешь кандидата: на этапе резюме и на собеседовании?

Подход На резюме — быстрый просмотр по «тегам» и влиянию на результат через метрики. На собеседовании — разбор любимого проекта через рамку trade-off'ов, затем современные инструменты и софт-скиллы.

Скрининг резюме

Green flag

Влияние на результат с метриками: «внедрил RAG → precision/recall выросли», а не просто «качество выросло».

Актуальный технологический стек.

Инженерный фундамент (для CAIO — приятный плюс, не строгое требование).

GitHub и реальные проекты — приятный позитивный сигнал.

Red flag

Устаревшие технологии и теги — мышление старыми подходами.

Результаты без метрик, «модель классная, и всё».

Несоответствие уровню зрелости компании (в т.ч. overqualified — будет скучно).

Отдельные компании не выделяю (BigTech — приятно, но не решающе). Для C-level опыт важнее образования: путь, который человек прошёл, важнее изначальной базы. Главное — уровень соответствия зрелости компании, а не абстрактные «флаги».

Собеседование

Любимый проект (любимая часть)

Глубоко закопаться и разобрать trade-off'ы. Red flag — если человек не мыслит альтернативами (взял модель на 100 млрд параметров «потому что классная», не думая о скорости и стоимости). Важно чувствовать, действительно ли человек горел проектом.

System design

Задача, близкая к реальным, — слушать рассуждения через ту же рамку trade-off'ов.

Современные инструменты

Чем пользуется (ChatGPT / Claude / Gemini), как и почему; знает ли MCP. Нормально не пользоваться продвинутым, если может это обосновать, — но базово ориентироваться обязан.

Софт-скиллы

Как принимал решения и обрабатывал возражения — снизу, сверху, от коллег. Data-driven или авторитарно. Умение быть услышанным и слушать: CAIO — «клей» между бизнесом и технологиями.

Глубина инженерного бэкграунда

Если заявлен — несколько минут вопросов. Дает позитивный сигнал, отсутствие — не негативный.

Мне крайне важно, чтобы человек мыслил trade-off'ами и не был зашорен. В разборе любого реального проекта это чувствуется очень хорошо.

На резюме — метрики и актуальность стека. На интервью — любимый проект через рамку trade-off'ов, владение инструментами и софт-скиллы.

Вопрос 05 · KPI

Чем должны отличаться KPI команды PoC и команды промышленного внедрения?

Суть различия PoC измеряет скорость и качество экспериментов. Production — надёжность, юнит-экономику и реальную денежную ценность. Это бизнесовая роль: мы делаем бизнес, а не науку.

Команда PoC — что мерить

Конверсия гипотез в следующую стадию (hit-rate) — на заранее заданном уровне (у нас ~60–80%).

Баланс надёжность — инновационность: рискованные гипотезы тоже нужны.

Time-to-value — скорость перехода между стадиями.

Стоимость эксперимента.

Команда Production — что мерить

SLA / SLI и доступность: система не падает и быстро отвечает.

Юнит-экономика — стоимость решения одной задачи.

Качество результата выше, чем на MVP (ближе к «сотой»).

Реальная денежная value — сколько решение экономит или приносит.

Мы делаем бизнес, в конце концов, а не науку. Все эксперименты важно смотреть в разрезе денег — на PoC это стоимость эксперимента, в проде это уже денежная ценность для бизнеса.

PoC: скорость, качество и стоимость экспериментов. Production: надёжность, юнит-экономика и реальная денежная ценность.

Вопрос 06 · Руководитель AI-направления

Три главных качества руководителя AI-направления (CAIO)

CAIO — это «клей» между бизнесом и технологиями на уровне всей организации: общается с CEO и продуктом, понимает рынок и при этом разбирается в технологиях как минимум на уровне метрик.

1 · Двужычие «бизнес — технологии»

Одинаково качественно говорить на языке бизнеса и технологий. Сюда же — архитектурное видение: helicopter view на все системы с умением погружаться на нужную глубину.

2 · Мышление метриками и trade-off'ами

Качество, скорость, стоимость — умение обрабатывать эти параметры совокупно и принимать решения через них.

3 · Культура и change management

Выстраивать культуру не только в своей команде, но во всей компании: «если культуру не завести — ничего не поедет».

CAIO — клей между бизнесом и технологиями: двужычие, мышление метриками/trade-off'ами и построение культуры.

Вопрос 07 · Инсайты + 5–10 лет

Главные инсайты и взгляд на горизонт 5–10 лет

Главный инсайт Ключевое — нанимать хороших инженеров. Хороший инженер с современными инструментами быстро перестраивается: у нас закрывали направления, но классных инженеров не увольняли — они трансформировали опыт и перешли на соседние задачи.

Кто такой «хороший инженер»

- **Владение инструментами и насмотренность**
Умение качественно использовать и сравнивать инструменты — модели, базы данных. Не «один раз изучил и везде пихаю», а понимание, что с чем сравнивал.
- **Мышление через эксперименты, а не только тесты**
В классической разработке тест прошёл или нет; в ML качество — это, например, 83%. Умение мыслить экспериментом, а не бинарным тестом, — отдельный ценный навык.
- **Фундамент + софт-скиллы**
Понимание базового computer science (как работает БД, ОС, процессор, параллелизм) — тогда любая новая технология осваивается быстро. Плюс софт-скиллы: инженер всё больше проводник между бизнесом и технологиями.

Взгляд на 5–10 лет

- Будут появляться новые роли и новые требования к ним — масштабирование всего, что обсуждали.
- Роль данных станет ещё важнее.
- Появится регуляторика — юристы и специальные комиссии для работы с ней.
- Часть профессий не вымрет полностью: лучшие останутся как «люкс/винтаж»-сегмент (как винил), но их станет меньше — делать ставку на такую профессию рискованно.

Выживут ли компании без AI. Скорее нет: «AI — это новое электричество». Даже без системной работы сотрудники всё равно будут пользоваться AI; системная работа резко повышает вероятность успеха. Без электричества, без компьютеров и без AI компании не выживают.

Главное — нанимать хороших инженеров (инструменты + эксперименты + фундамент + софт-скиллы). Роль данных и регуляtorики растёт, часть профессий уходит в «люкс-сегмент».

Вопрос 08 · Три совета

Три совета компании, которая только начинает строить AI-функцию

- 1 · **Проверить фундамент — данные и цифровизацию.**
Garbage in — garbage out: сначала разобраться с данными и цифровизацией, потом строить AI вокруг них. Это не жёсткий блокер — какие-то инициативы можно прорабатывать параллельно, но базовый слой никуда не денется.
 - 2 · **Не «AI ради AI».**
Мыслить через продукт и потребности, а не через технологии. Идти с двух сторон — от технологий к применению и от потребности к технологии, — но именно синергия продукта и технологий даёт результат.
 - 3 · **Начинать с быстрых побед.**
Low-hanging fruit почти всегда есть. Искать их через системную работу с инициативами (а не «пришла идея — делаем»). Быстрые победы обосновывают бюджет, на который дальше развивается команда.
- Финал — технооптимизм.** Риски «захвата мира» есть, но их вероятность ниже, чем у светлого сценария: человечество достаточно адекватно, чтобы отслеживать риски на каждом этапе развития.

Данные → продукт, а не технологии ради технологий → быстрые победы через системную работу. И ставка на хороших инженеров.

18

Три взгляда — одна картина: технический, продуктовый и стратегический угол

Одна тема, три эксперта с разным бэкграундом. Не противостояние — а объёмная картина: технический, продуктовый и стратегический взгляды дополняют друг друга. Вместе они дают полную карту для тех, кто строит AI-команду.

СТО · Интервью №1 — Артур Сахаров. СТО, MWS AI

СРО · Интервью №2 — Максим Волошин. СРО MWS AI

САЮ · Интервью №3 — Даниил Киреев. САЮ, MWS AI

Где взгляды совпадают — консенсус трех экспертов

Промпт-инжиниринг как самостоятельная роль умирает.

СТО и СРО независимо пришли к одному выводу: модели сами генерируют хорошие промпты лучше большинства людей. Ценность смещается к архитектуре решения, eval-дисциплине и умению встраивать AI в бизнес-процессы.

САЮ подтверждает: промпт-инженеры трансформируются в AI-инженеров, а сам навык становится базовым — «как владение Word». Это не мнение одного человека, а рыночный факт 2025–2026 года, который называют все трое.

Фундамент — данные, инфраструктура и eval — узкие места, о которых молчат.

Сахаров: «DevOps и безопасность не готовы к AI, без подготовленной инфраструктуры AI-команда не взлетит». Волошин: «Eval Engineer — первая инвестиция после Head of AI».

САЮ добавляет нижний слой: «garbage in — garbage out» — без качественно структурированных данных не поедет ничего; для on-premise нужны MLOps/LLMOps. Одна проблема — три угла зрения.

«Нанять звезду и забыть» — иллюзия.

СТО и СРО категоричны: без готовности компании меняться, без реального мандата сверху, без инвестиций в обучение сотрудников любой найм провалится.

САЮ формулирует это как дисциплину: любое внедрение AI — это change management. Нужна культура, где не бояться ошибаться (guardrails), а предпринимательский настрой прорастает сверху вниз.

PoC и Production — две функции с противоположной логикой.

СТО и СРО называют одни и те же причины, почему PoC не доходит до прода: экономика ломается при масштабировании, архитектура не выдерживает, инфраструктура не готова. PoC-команду нельзя мерить как production.

САЮ задаёт точные метрики: PoC измеряет скорость и качество экспериментов (hit-rate, time-to-value, стоимость эксперимента), Production — SLA/SLI, юнит-экономику и реальную денежную ценность.

Руководитель AI-направления — самая важная роль во всём процессе.

Сахаров: «лучше найти кого-то внутри». Волошин: «лучше 6 месяцев искать сильного, чем за 6 недель нанять слабого». Три качества совпадают с точностью до формулировок.

САЮ описывает суть роли: это «клей» между бизнесом и технологиями. Его три качества — двуязычие «бизнес — технологии», мышление метриками и trade-off'ами, построение культуры — почти дословно совпадают с двумя другими.

Централизованная AI-команда — переходная структура.

СТО и СРО видят одно будущее: AI встраивается в продуктовые команды как функция, а не живёт в отдельном бункере. Аналогия с IT-департаментами использована обоими независимо.

САЮ описывает это как градиент: от лёгкой интеграции с партнёром до глубокой in-house команды; число ролей и глубина растут вместе с числом инициатив.

Как три взгляда дополняют друг друга

Три взгляда описывают одну реальность с разных сторон — как план здания, схема коммуникаций и финансовая модель проекта: каждый необходим, ни один не заменяет другой. СТО видит инфраструктуру, системы и архитектуру. СРО видит людей, процессы и измеримый бизнес-эффект. САИО связывает технологии со стратегией и деньгами — отвечает за то, чтобы «как построить» и «ради чего» сходились в единую экономику.

Тема	Технический угол — СТО	Продуктовый угол — СРО	Стратегический угол — САИО
Найм: на что смотреть	Способность внедрять изменения и управлять неопределённостью — трек-рекорд изменений, не присутствия.	Shipping-опыт и eval-дисциплина — «что довёл до прода». Paid trial как лучший способ проверки.	Разбор любимого проекта через рамку trade-off'ов: red flag — кто не мыслит альтернативами. Плюс мышление метриками и уровень соответствия зрелости компании (overqualified — тоже риск).
Структура команды	Стратегический выбор пути: AI встраивается в команды или создаётся конструкторское бюро.	Три сценария (in-house / партнёр / агенты) с ролями и таблицами под каждый.	Команда — это градиент глубины: от интегратора до in-house. Управление растёт от дуо (Product + Tech Lead) к триумvirату (+ ML Lead), а на уровне организации — САИО как связующее звено.
Минимальная команда	Важна структура и готовность компании, а не магическое число. Предостерегает от найма «не тех».	Конкретный ориентир: 5–7 человек — минимум для покрытия eval/ops.	Базовый слой: AI-инженер · MLOps/LLMOps · дата-инженер · AI-product-менеджер. Начинать ближе к интеграторам и наращивать по мере роста AI-культуры.
Узкое место рынка	Инфраструктурный кризис: DevOps не готов к AI, корпоративные политики блокируют инструменты.	Кадровое расслоение: война за топ-5%, поток «API-caller'ов» в середине.	Фундамент данных (garbage in — garbage out) и рескиллинг: новые роли — производные от старых, но спрос растёт быстрее, чем рынок успевает переучиваться.
КРИ и измерения	Экономический взгляд: стоимость инференса, unit economics, эффективность инженера через токены.	Портфельный взгляд: kill rate, stakeholder confidence, MTTD/MTTR.	PoC: hit-rate гипотез (~60–80%), time-to-value, стоимость эксперимента. Production: SLA/SLI, юнит-экономика, денежная ценность. Портфель — как венчур: выстреливает ~5–10% инициатив.
Работа с клиентом	Консалтинг «200 страниц и уйдём» не работает — нужно делать руками вместе с клиентом.	Инструменты для самодиагностики: матрица зрелости, self-assessment, чек-листы.	«AI ради AI» не работает — идти от продукта и потребностей. Начинать с быстрых побед через системную работу с инициативами, а не «пришла идея — делаем».
Роли будущего	AI-инженер и AI Architect как ключевые технические роли. Evaluation Engineer — отдельная и недооценённая.	AI Product Engineer (Forward Deployed) — гибрид инженера, продакта и консультанта.	AI-инженер как базовая роль, Data Owner/ Steward за данные, ML-инженер и ML-исследователь — по глубине. Дальше — регуляторика: юристы и комиссии под новые требования.

Три взгляда — это не конкуренция, а объёмная картина. СТО видит инфраструктуру, СРО — людей и бизнес-эффект, САИО — стратегию и связь технологий с деньгами. Компания, у которой при решениях о найме и построении AI-функции есть все три угла зрения, допускает значительно меньше дорогостоящих ошибок.

Соответствие трендам до 2029 года

Агентный AI \$376 млрд рынок к 2027 Переход от «умения писать промпты» к «умению строить агентные системы» — все трое описывают как уже произошедший факт 2025–2026. Волошин добавляет роли будущего (Agent Orchestrator, AI SRE), Сахаров указывает на инфраструктурный барьер входа. Киреев: спрос смещается к мультиагентным системам, а AI-инженеры собирают готовые модели в работающий pipeline без дообучения. Прогноз до 2029 подтверждён с трёх сторон.	Все согласны
Гибридизация ролей 5 AI-усиленных > 10 обычных Волошин: «Forward Deployed Engineer — гибрид инженера, продакта и консультанта». Сахаров: «разработчики с AI работают эффективнее». Киреев: промпт-инженер → AI-инженер, размывание границы «ресёрчер/инженер»; хороший инженер с современными инструментами быстро перестраивается. К 2027–28 это станет новой нормой найма.	Все согласны
Охлаждение рынка –34% вакансий, сохр. дефицит senior Сахаров: война за топ-5% продолжается несмотря на охлаждение. CPO-добавлениеВолошин: tenure ≥ 18 месяцев как сигнал качества, внутренняя мобильность бьёт внешний найм. Киреев: ключ — рескиллинг; новые роли производны от старых, но людей нужно добирать и с рынка. К 2028 разрыв junior/senior усилится до 15–20х.	СТО-акцентCPO-добавление
Eval как конкурентный ров Новая дисциплина «Evaluation Engineer — самая недооценённая роль 2026 года» (Волошин). «Competitive advantage сегодня — в eval-дисциплине» (Сахаров). Киреев: «машина экспериментов» со stage-gate и мышление на уровне метрик — must-have; без этого эксперименты не масштабируются. К 2027 eval выделится в отдельную дисциплину наравне с DevOps.	Все согласны
Импортозамещение Дедлайн 2028 Сахаров: корпоративные политики блокируют Claude Code и доступ к моделям, работа на суверенном стеке — ограничение среды. Киреев: для on-premise нужны MLOps/LLMOps, которые развернут суверенный стек; без слоя данных миграция бессмысленна. К 2028 AI-команда без компетенций в российской инфраструктуре — структурный барьер.	СТО-акцент
Смерть центров компетенций AI в каждой команде Все трое используют аналогию с IT-департаментами: AI становится функцией команды, а не отдельным подразделением. Киреев: на уровне организации появляется CAIO, а предпринимательская культура прорастает вниз по иерархии. Разногласие в скорости, но не в направлении.	Все согласны

4 итоговых вывода

- 1

Технология больше не является конкурентным преимуществом

Все три эксперта независимо пришли к одному выводу: преимущество сместилось с «лучшей модели» на архитектуру решения, eval-дисциплину и способность встраивать AI в реальные бизнес-процессы. Киреев усиливает это тезисом о росте уровня абстракции: фундаментальные модели «из коробки» обнулили ценность ручного обучения. Вопрос «какую LLM взять» — второстепенный; первичный — «есть ли люди, умеющие измерять качество в production и управлять экономикой инференса».
- 2

Правильный найм — это вопрос готовности компании, а не рынка кандидатов

Все трое согласны: дефицит реален, но большинство провалов происходят из-за неготовности компании. Нет мандата сверху — звезда уйдёт; нет eval-инфраструктуры — инженер не сможет работать; нет готовности пересмотреть KPI — руководитель AI станет декоративным. Киреев добавляет управленческую рамку: внедрение AI — это change management и культура, где не боятся ошибаться. До 2028 это разделит компании на тех, кто выстроил готовность, и тех, кто ищет «правильного кандидата» вместо правильной среды.
- 3

Разрыв между PoC и Production — организационная проблема, не техническая

Причина того, что PoC не доходит до прода, почти никогда не техническая: экономика ломается, архитектура не масштабируется — это следствия решений на этапе PoC. Решение — с первого дня разделить Discovery и Production, дать им разные метрики и команды. Киреев задаёт эти метрики предметно: PoC — скорость/качество/стоимость экспериментов; Production — SLA, юнит-экономика, денежная ценность, а портфель ведётся с венчурным mindset.
- 4

Данные — фундамент, бизнес — цель, быстрые победы — точка входа

Отдельный стратегический акцент CAIO, дополняющий технический и продуктовый взгляды. Сначала — слой данных и цифровизация («garbage in — garbage out»), иначе AI не на чем строить. Дальше — мыслить через продукт и потребности, а не «AI ради AI». А начинать стоит с быстрых побед через системную работу с инициативами: они обосновывают бюджет, на который дальше развивается команда.
- 5

Окно оптимального входа закрывается — и все три эксперта видят это одинаково

Несмотря на различия в акцентах, все приходят к одному стратегическому выводу: компании, начавшие строить AI-функцию в 2023–2025 годах, входят в рост с 2–3-летним преимуществом. К 2027–2028 специалисты станут дороже, конкуренция острее, требования к инфраструктуре жёстче. Киреев формулирует это предельно прямо: «AI — это новое электричество», и компания без него проиграет так же, как когда-то проиграли бы без компьютеров. Самое дорогое решение — не начинать.

Итог: СТО смотрит на системы и инфраструктуру, CPO — на людей, процессы и измеримый эффект, CAIO — на стратегию, данные и связь технологий с деньгами. Там, где их взгляды совпадают, — это не мнение, а сигнал рынка: смерть «чистого» промпт-инжиниринга, фундамент из данных и eval, правильный руководитель AI, разделение PoC и Production, готовность компании меняться. Именно это стоит делать в первую очередь — начиная с быстрых побед.

19

Источники исследования

Гайдбук сочетает первичные данные MWS AI с анализом открытых российских и зарубежных источников. Каждый ключевой вывод подкреплён не менее чем двумя независимыми источниками; там, где данные расходятся, фиксируются оба взгляда.

Внутренние данные MWS AI

HR-аналитика MWS AI

Закрытые позиции в AI-командах за 2025–2026 гг.: сроки закрытия по ролям, конверсия на каждом этапе воронки, эффективность каналов поиска, распределение офферов.

Exit-интервью

Причины смены работы, мотиваторы перехода и оценка условий труда AI-специалистов.

Глубинные экспертные интервью

СТО, СРО и CAIO компаний с реальным опытом построения AI-команд — рынок, роли, скрининг, KPI, мотивация, ошибки.

Данные рынка труда

hh.ru

«Рынок труда в 2025 году: почему стало труднее искать работу»

Динамика вакансий и резюме, рост конкуренции соискателей.

vc.ru

«Парадокс российского ИТ-рынка 2025»

Численность и дефицит ИТ-кадров, разрыв зарплат junior/senior, спрос на 1С.

Московские новости

«Как рынок труда пережил 2025 год»

«Базовое знание ИИ» как требование во всё большем числе вакансий.

Inc. Russia

«AI в найме: AI-агенты и AI-сотрудники»

Аналитика рынка AI

«Яков и Партнёры» × Яндекс

«Искусственный интеллект в России — 2025»

Экономический эффект AI и его вклад в прогнозный ВВП к 2030 году.

ComNews

«Российские компании стали чаще внедрять инструменты ИИ»

Рост доли компаний, использующих AI (опрос «Систем Икс»).

CNews

«MTS AI начинает продажи внешним клиентам ИИ-ассистента»

Отраслевые и деловые медиа

Forbes

«Спрос на сотрудников с навыками применения ИИ вырос осенью на треть»

AI-навыки как всё более частое требование в вакансиях.

Ведомости

«Переход значимых объектов КИИ на отечественный софт»

Сроки перехода критической инфраструктуры на российское ПО.

Habr «От чат-бота до ИИ-агента: 13 проектов российских компаний» Проникновение AI и агентных систем в российских компаниях.
ТехИнсайдер (TechInsider) «Профессии, которые не исчезнут: каким будет ИТ-рынок после 2030» Роли будущего и гибридизация ИТ-профессий.
4PDA «87% разработчиков то и дело прибегают к помощи искусственного интеллекта»

Профильные платформы и публикации

Битрикс24 «Как подружить AI-агента с Битрикс24» AI-агенты поверх существующих ERP и CRM без замены систем.
SENSE IT «Зарплаты ИТ-специалистов в РФ: основные тренды I квартала 2026» Макроэкономика и зарплатные тренды ИТ-рынка.
Elbrusboot «ИТ-рынок труда 2026: секреты найма и 5 стратегий» Нормализация ИТ-рынка и переход к «рынку работодателя».
1000bankov «Российский ИТ без перегрева: сколько теперь платят разработчикам» Зарплатные рычаги: переход с middle на senior.
Generation-ai «Тренды ИИ 2026 и итоги внедрения AI в российских компаниях за 2025» «Год агентов»: усиление мультиагентных систем.
Course-itrecruiter «Рынок труда ИТ 2026: охлаждение спроса и новая реальность для HR» Импортозамещение (2028) и спрос на 1С и российские СУБД.
Kinzhal «Как откликнуться на вакансию, чтобы вас заметили» Слом классического рекрутинга: нетворкинг и рекомендации.
Neogenda Изменение ролей под влиянием AI: вайб-кодинг в требованиях к РМ, переквалификация функций.

Период исследования: 2025–2026. Источники сгруппированы по типу данных; внутри групп — по вкладу в материал гайдбука.

ИИ-трансформация: системный подход от стратегии до эффектов в P&L

Центр ИИ-трансформации MWS AI — партнёр по ИИ-трансформации корпораций. Помогаем пройти путь от идеи до промышленного запуска: выбрать сценарии, просчитать эффект, закрыть риски, реализовать эффекты и масштабировать их.

Наш подход сочетает консалтинг от оценки ИИ-зрелости и реализации ИИ стратегии до управления ИИ проектами на потоке от идеи до масштабирования. Обучение команд и развитие ИИ компетенций встроено в каждый этап.

Наш подход

Шаг 01
Находим, где ИИ даст эффект
Выбираем до 10 сценариев, считаем потенциал. Финансовая модель — у клиента.

Шаг 02
Доказываем на практике
Пилот на реальных данных, замер результата, обоснование для масштабирования.

Шаг 03
Масштабируем
Внедряем в работу компании, помогаем выстроить внутренний ИИ-центр.

Итог — не презентация, а **зафиксированный эффект в деньгах** и команда, готовая масштабировать ИИ самостоятельно.

Наши направления

01 Стратегический консалтинг

- ИИ-стратегия, дорожная карта, портфель инициатив с расчётом ROI
- Оценка ИИ-зрелости, диагностика процессов, данных, команды, инфраструктуры
- Портфель инициатив, приоритизация по ROI
- Офис трансформации, KPI-модель, governance

02 Развитие ИИ-компетенций

- Центр ИИ-компетенций, AI-команда, матрицы, онбординг
- Обучение топ-менеджеров, управление ИИ-инновациями
- Обучение ИИ-команд, ML-инженеры, data scientists, prompt engineers
- ИИ-культура и AI adoption

для кого

HRD / HR BP
Кого нанимать для ИИ-команды, в какой последовательности и с каким бюджетом.

CDTO / CIO
Стратегия и архитектура ИИ-функции без ошибок, которые стоят месяцев и бюджетов.

CEO / руководители
Измеримый ROI от ИИ, технология как конкурентное преимущество.

БИЗНЕС-ЭФФЕКТЫ

−30% операционных затрат в отдельных функциях

+5% к EBITDA за счёт ИИ в ключевых процессах

×1,5–2 ускорение time-to-market новых продуктов

−40–60% длительность PoC и пилотов

Чем мы отличаемся

Полный цикл без посредников
От стратегии до продакшна — собственная платформа и LLM. Конкуренты работают через партнёров.

KPI-модель: делим риски с клиентом
Мотивация — KPI клиента. Масштабируем только кейсы, прошедшие скоринг эффективности.

Опыт МТС — в ваш бизнес
Более 2 млрд ₽ экономии в МТС. Обкатанная методология — не теория.

On-premise, 152-ФЗ, ФСТЭК
Данные в контуре клиента. Собственные LLM и GPU Cloud.

Реальный опыт

400+
экспертов в ИИ-технологиях

>₽2,5 млрд эффект в 1-й год
Экосистемный бизнес · стратегия + обучение

8 ИИ-инициатив в работу
Банк · центр компетенций + платформа

100+
кейсов внедрения в корпорациях

10 тыс + руководителей прошли наши программы обучения по ИИ
Авиакомпания · стратегия + сессии

7 активов · 20+ инициатив с ROI
Холдинг · оценка ИИ-зрелости активов

Состав команды

Над гайдбуком работала команда MWS AI: HR-консалтинг, исследование рынка и экспертиза в построении AI-команд.

Руководитель проекта



Наталья Рейслер

Руководитель Отдела подбора персонала MWS AI

Квалификация

- С отличием закончила Факультет государственного управления МГУ имени М.В. Ломоносова.
- Весь карьерный путь в области подбора персонала: от ресерчера до руководителя отдела.
- Научные публикации на тему мотивации персонала.

Профессиональные компетенции

- Подбор ML-специалистов и команд в области AI, executive search в фармацевтике, FMCG и IT.
- Full cycle talent acquisition в технологическом секторе.
- Исследование динамики рынка труда в России и влияния AI на процессы найма.

Опыт работы

- Руководитель отдела подбора персонала MWS AI.
- Территориальный менеджер по подбору коммерческого персонала, Coca-Cola Hellenic Bottling Company.
- Ресёрчер → Консультант, HAYS: подбор коммерческого и научного персонала для ведущих фармацевтических компаний, специализация на орфанных заболеваниях.

Достижения

- Обладатель двух красных дипломов МГУ имени М.В. Ломоносова.
- Более 20 закрытых C-level позиций в международных фармацевтических компаниях: Allergan, Abbott Laboratories, Astrazeneca, Pfizer, Abbvie и др.
- Стабильно высокие темпы найма с 2023 г.
- Экспертные комментарии в отраслевых медиа (Fortune, Техинсайдер) по теме AI и рынка труда.

Состав экспертного совета



Оксана Хлыстова

Руководитель Центра AI-трансформации MWS AI

Достижения

- Разработка AI-стратегии для одной из крупнейших авиакомпаний.
- Реализация программ по AI для 10 тыс. сотрудников корпораций.
- Лекции и воркшопы для 200+ топ-менеджеров.
- Авторские курсы для вузов (ВШЭ).
- 4 масштабных международных проекта по AI-трансформации: 300+ AI-решений из 37 стран, 70+ бизнес-функций.



Анастасия Зальцман

Директор по персоналу MWS AI

Достижения

- Соавтор книги «Не мешайте людям работать» (издательство Альпина Паблишер, 2025).
- Спикер программ Сколково (CDTO) и Корпоративный университет Сбербанка (Digital Strategy).
- 1 место (2020) и 2 место (2021) премии HR-бренд hh.ru за проекты по HR-аналитике и менторству.



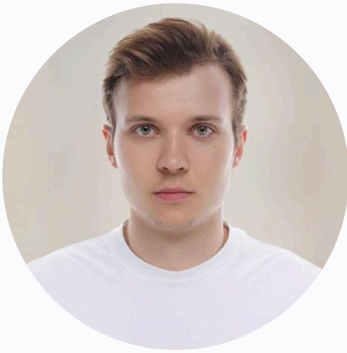
Артур Сахаров
Технический директор MWS AI

- Достижения**
- Спикер программ Корпоративного университета Сбербанка и Сколково.
 - Соавтор книги «Не мешайте людям работать» (издательство Альпина Паблишер, 2025) и одноимённого образовательного проекта.
 - Запуск «ГигаАкадемии» — образовательной экосистемы GigaChat.



Максим Волошин
Директор по продуктам MWS AI

- Достижения**
- Благодарность Президента Российской Федерации за участие в организации саммита G8 в Санкт-Петербурге.
 - Почетные грамоты президента Сбербанка за вклад в реализацию AI-стратегии и развитие инновационных продуктов.
 - Реализация 200+ AI-инициатив в рамках SberDevices, трансформация организации и развитие партнерств.
 - Акселерация 30 AI-стартапов в программах Sber500 и GigaChat.



Даниил Киреев
Директор ML-департамента MWS AI

- Достижения**
- Лауреат рейтинга Forbes «30 до 30» по треку «Наука и технологии».
 - Победа в совместном конкурсе от Яндекс и ВШЭ.
 - Преподаватель в МГИМО и Московском государственном университете имени М. В. Ломоносова.

Команда проекта



Ольга Жигунова
Ведущий менеджер по привлечению и развитию талантов



Алена Остроушко
Ведущий менеджер по привлечению и развитию талантов



Елизавета Правосудова
Ведущий менеджер по привлечению и развитию талантов



Виктория Иванова
Координатор проектов



Сергей Запорожченко
Менеджер образовательных программ



AI

Контакты



hr-guidebook@mts.ai



mws.ai

Обсудим детали и предложим
оптимальное решение под ваши задачи