



Multi-Agent Interview Coach

Мультиагентная система проведения
технических интервью с адаптивной оценкой

Неронов Роман Михайлович

Зачем это нужно

Проблема

Подготовка к техническим интервью —
болезненный процесс:

- Нет объективной обратной связи
- Невозможно практиковаться
реалистично
- Менторы не всегда доступны
- Самооценка субъективна и неточна
- Нет плана развития навыков

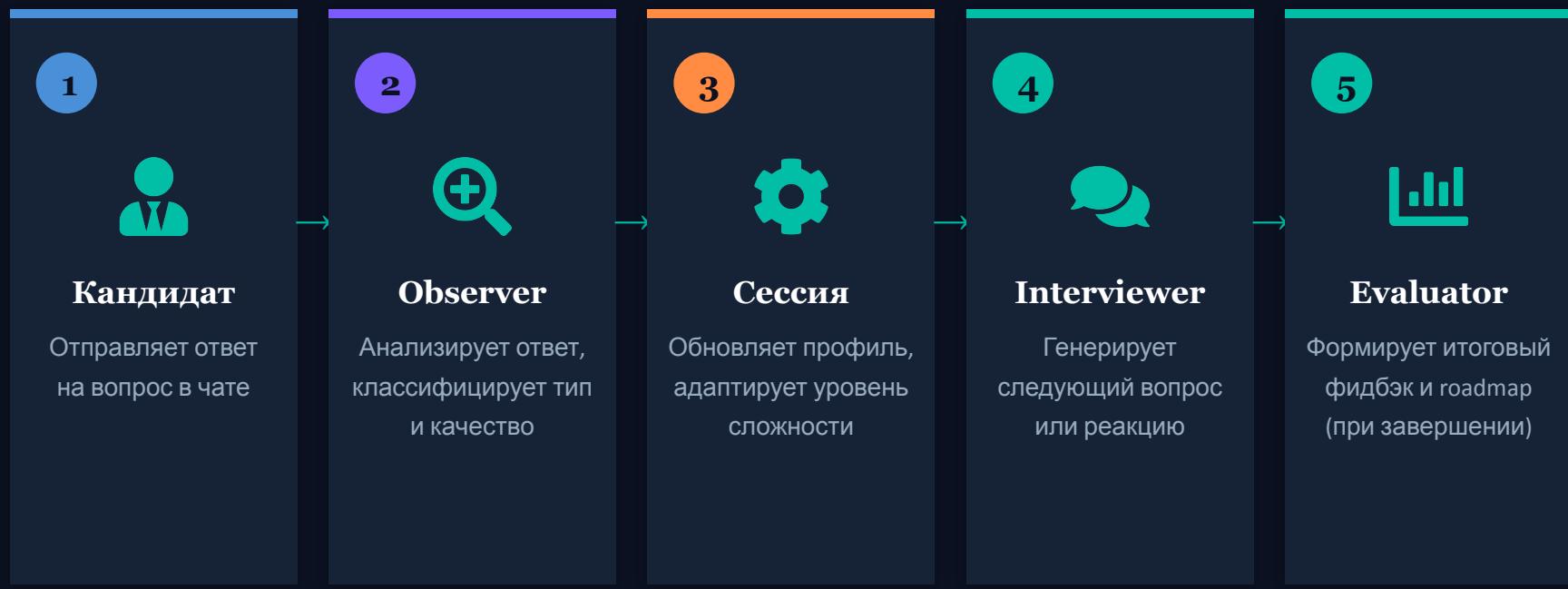
Решение

AI-система для полноценного интервью:

- 3 агента анализируют каждый ответ
- Сложность адаптируется под уровень
- Детекция ошибок и галлюцинаций
- Фидбэк с оценкой навыков
- Персональный план развития

Как работает система

Пошаговый процесс обработки каждого сообщения кандидата



Шаги 1–4 выполняются на каждое сообщение · Шаг 5 — только при завершении интервью

Мультиагентная архитектура

Три специализированных агента обеспечивают глубокий анализ на каждом шаге



Observer

Аналитик

- Классифицирует тип ответа: технический, off-topic, галлюцинация
- Оценивает качество и фактическую корректность
- Извлекает данные о кандидате из текста
- Формирует рекомендации для Interviewer



Interviewer

Интервьюер

- Ведёт диалог, задаёт один активный технический вопрос
- Адаптирует сложность от BASIC до EXPERT
- Реагирует на ошибки, off-topic и галлюцинации
- Вопросы только по стеку кандидата



Evaluator

Оценщик

- Формирует вердикт: грейд и рекомендация по найму
- Анализирует hard и soft skills по данным интервью
- Составляет персональный roadmap развития
- Каждое утверждение подкреплено репликой из диалога

Интеллектуальные возможности

Адаптивная сложность

4 уровня: от BASIC до EXPERT.
Система повышает сложность при серии хороших ответов и снижает при затруднениях

Детекция галлюцинаций

Распознаёт фактически неверные утверждения, сообщает правильную информацию и фиксирует пробелы в знаниях

Активный якорь вопроса

Сохраняет текущий технический вопрос при попытках смены темы, встречных вопросах или бессмыслице



Извлечение профиля

Автоматически определяет имя, позицию, грейд, стек и опыт кандидата из текста для персонализации



Атомарность состояния

Изменения применяются только при успехе всех агентов. При сбое — автоматический откат к предыдущему состоянию



Итоговый фидбэк

Вердикт, анализ hard/soft skills, уверенность оценки и персональный план развития с ресурсами

4

уровня сложности

3

агента в пайплайне

8

типов ответа

5

метрик качества

Технологии и мониторинг



LLM Gateway

LiteLLM Proxy — маршрутизация запросов через OpenAI-compatible API к локальным и облачным моделям



Backend

FastAPI + async Python, Pydantic-валидация, Gradio UI для чат-интерфейса



Инфраструктура

Docker Compose: 6 контейнеров — Gradio, FastAPI, Redis, Nginx, Langfuse, PostgreSQL



Observability

Langfuse (self-hosted) — трейсы сессий, метрики токенов, диагностика LLM-вызовов

Метрики Langfuse

Sessions · Generations · Token Metrics по агентам · Confidence Score · Средние токены за ход · Диагностика ошибок

Спасибо за внимание

Multi-Agent Interview Coach

Неронов Роман Михайлович