

Чек-лист: минимальный контур надёжного AI-агента

Используйте этот список перед расширением доступа агента к корпоративным данным или выдачей прав на действия. Для скачивания подготовьте файл [downloads/checklist-ai-agent-production.pdf](#) и добавьте его отдельной строкой на лист Downloads.

Источники и контекст

- Для каждого источника назначен бизнес-владелец и определён разрешённый сценарий использования.
- Заданы каноническая версия записи, правила разрешения конфликтов и SLO свежести.
- Версии документа, индекса, схемы инструмента и embedding-модели фиксируются и доступны для проверки.
- Права пользователя и агента проверяются до выдачи контекста.

Оценка

- Retrieval проверяется на авторитетность, свежесть, доступность и позицию источника в выдаче.
- Ответ проверяется на подтверждённость контекстом, полноту и корректное обозначение неопределённости.
- Траектория проверяется на выбор инструментов, параметры и порядок шагов.
- Оценивается бизнес-исход и возможные побочные эффекты, а не только текстовый ответ.

Наблюдаемость и безопасность

- Для каждого запуска сохраняется трасса с источниками, версиями, tool calls и результатами.
- Технический успех API не подменяется подтверждённым бизнес-исходом.
- Запись во внешние системы закрыта по умолчанию, разрешения выдаются точечно.
- Для высокорисковых операций предусмотрены лимиты, sandbox, stop-условия и human approval.
- Агент умеет безопасно отказаться от действия и передать кейс человеку.

Непрерывная проверка

- Есть регрессионный набор с устаревшими, конфликтующими и неполными данными.
- Набор запускается после изменения модели, промпта, индекса, политики или схемы инструмента.
- Команда регулярно разбирает production-трассы, отменённые действия и обратную связь пользователей.
- Редкие дорогостоящие ошибки анализируются отдельно от средней оценки качества.