

Чек-лист пилота AI harness

Содержимое этого чек-листа предназначено для отдельного PDF-файла `downloads/deepseek-ai-harness-pilot-checklist.pdf`. В DOCX общий заголовок должен иметь стиль «Заголовок 3», названия блоков «Заголовок 4», а пункты должны быть штатными маркированными списками Word.

Цель и метрики

- Выберите рабочие задачи, а не демонстрационные диалоги.
- Определите критерий успеха для каждой задачи до запуска теста.
- Фиксируйте стоимость, время, число вызовов модели, tool calls, повторы и ручные вмешательства.
- Считайте стоимость успешно завершённой задачи, а не среднюю цену запроса.

Контекст и кэширование

- Разделите постоянные инструкции и динамические данные конкретной задачи.
- Версионизируйте системные правила и проверяйте их на фиксированном наборе.
- Измеряйте отдельно входные токены, выходные токены, повторы и долю cache hit.
- Проверьте, не нарушают ли повторяемость префикса случайные идентификаторы и служебные поля.

Маршрутизация и безопасность

- Сравните дешёвый и более сильный маршруты на одинаковом наборе задач.
- Сохраняйте причину выбора маршрута и итог его выполнения.
- Ограничьте число повторов и бюджет одной траектории.
- Выдавайте агенту только минимально необходимые права.
- Требуйте подтверждения для операций с записью, отправки сообщений и публикации кода.

Надёжность агентного цикла

- Тестируйте многоходовые вызовы инструментов отдельно от обычного чата.
- Проверяйте неполные данные, длинный контекст, ошибки инструментов и обязательную остановку агента.
- Логируйте решения маршрутизатора, вызовы инструментов, ошибки и финальное состояние задачи.
- Добавьте проблемные случаи в регрессионный набор после каждого выявленного сбоя.