

Чек-лист: как проверить стабильность ИИ-агента

Памятка дублирует практический раздел статьи в компактном формате. Ее можно использовать как стартовый список для аудита агентного сценария перед внедрением или регрессионным тестированием.

Статус повода

- Не называйте issue #1506 планом DeepSeek без подтверждения от команды проекта.
- Проверьте статус issue, labels, assignee, milestone, связанные PR и ответы maintainers.
- Разделяйте заявления разработчика, свойства модели и предложения сообщества.

Логи и сценарии

- Сохраняйте system prompt, пользовательский prompt, schema инструментов, аргументы и результаты вызовов.
- Логируйте чтение и запись памяти, retries, latency, fallback и human handoff.
- Добавьте тесты на потерю роли, смену цели, загрязнение памяти, циклы и восстановление после ошибки.
- Определите ожидаемое состояние для ключевых этапов сценария.

Контроль переменных

- Проводите контрольные и экспериментальные прогоны в одном временном окне.
- Не меняйте одновременно модель, провайдера, prompt, память и оркестратор.
- Не используйте скорость, latency и цену как самостоятельную метрику устойчивости.
- Определите разрешенные инструменты, правила эскалации и аудит действий для enterprise-сценариев.