

Чек-лист: eval-driven development для Agent Skills

Подготовить скачиваемый PDF-файл `downloads/firebase-agent-skills-eval-checklist.pdf` с этим чек-листом. В XLSX на листе Downloads указать: filename — `firebase-agent-skills-eval-checklist.pdf`; title — «Чек-лист eval-driven development для Agent Skills»; description — «Памятка для подготовки сценариев, eval'ов, трассировки и ограничений полномочий AI-агента»; buttonLabel — «Скачать PDF».

Контракт и набор сценариев

- Зафиксировать ожидаемый результат, запрещённые действия и точки подтверждения.
- Добавить типовые запросы, неоднозначные формулировки, неполный контекст и опасные действия.
- Определить, какие сценарии требуют эскалации человеку.
- Версионировать skill вместе с набором проверок.

Оценка поведения

- Проверить activation нужного skill и отсутствие ложных срабатываний.
- Проверить инструмент, параметры, полномочия и подтверждения.
- Проверить траекторию в многошаговых сценариях.
- Подтвердить итог внешним тестом, policy-check или состоянием изолированной среды.
- Не допускать, чтобы средний score скрывал критическое нарушение.

Наблюдаемость и доступы

- Сохранять контекст запуска, версии модели и skill, список инструментов и trace.
- Исключать секреты и чувствительные данные до сохранения trace.
- Учитывать срок хранения Interaction и параметр `store=false` для stateless-сценариев.
- Разделить подготовку изменения и право применить его в production.
- Ограничить инструменты и исходящий сетевой доступ только необходимыми возможностями.