

Памятка: как проверять агентные функции с повышенным риском

PDF для [downloads/checklist.pdf](#) должен повторять практический чек-лист из статьи: доступ агента, инструменты и сеть, подтверждения и эскалация, ограничения ярлыка риска.

Опишите доступ агента

- Перечислите все источники данных, которые читает агент.
- Отделите доверенные источники от недоверенных.
- Проверьте доступ к секретам, токенам и внутренним настройкам.
- Определите доступные директории и репозитории.
- Запишите внешние страницы и сервисы, которые агент может открывать.

Ограничьте инструменты и сеть

- Отключите ненужный сетевой доступ.
- Используйте sandbox для кода и файловых операций.
- Ограничьте команды, которые можно запускать без подтверждения.
- Проверьте, может ли агент отправлять данные наружу.
- Настройте минимально необходимые права.

Настройте подтверждения и эскалацию

- Определите операции, которые всегда требуют ручного подтверждения.
- Задайте правила передачи задачи человеку.
- Проверьте журналирование действий агента.
- Разделите черновые рекомендации и реальные изменения в системе.
- Проводите тесты на недоверенном контенте до расширения доступа.

Не переоценивайте ярлык риска

- Не считайте Elevated Risk label технической защитой сам по себе.
- Не снимайте ограничения только потому, что функция официальная.
- Не используйте продуктовые метрики как внешний аудит безопасности.
- Проверяйте настройки доступа после существенных изменений процесса.
- Возвращайтесь к модели угроз, если агент получает новые инструменты.