

Чек-лист: безопасность пилота ИИ-агента

Сохраняемый PDF должен повторять этот чек-лист без служебных комментариев.

Рекомендуемое имя файла: agent-security-pilot-checklist.pdf. Использование: пройти пункты до выдачи агенту доступа к инструментам и данным.

Идентичности и доступы

- Создать отдельную service identity для каждого агентного сценария.
- Выдать минимально необходимые права вместо широких ролей.
- Не использовать shared credentials и постоянные токены без срока действия.
- Исключить секреты, production-ключи и персональные данные из sandbox, если они не требуются задаче.
- Настроить отдельное подтверждение для операций записи, удаления и изменения прав доступа.

Среда и сеть

- Описать разрешённые tool calls, сетевые маршруты и внешние endpoint до запуска.
- Ограничить исходящий трафик на уровне сети, а не только в prompt-инструкции.
- Проверить package install proxy, кэши, реестры и промежуточные компоненты как часть security boundary.
- Разделить тестовую и production-среды, включая учётные данные и журналы.
- Определить техническую логику остановки агента при выходе за допустимые условия.

Наблюдаемость и реакция

- Логировать вызовы инструментов, API, handoffs и попытки сетевого выхода.
- Связывать действия агента с целью задачи, идентичностью и временем выполнения.
- Защитить журналы от утечки токенов, ключей, файлов и персональных данных.
- Проверить оповещения на попытки привилегированного действия, egress и обращения к секретам.
- Провести учение по сценарию, в котором агент находит обход ограничений.