

Safety & Escalation

Реакции

- `ASK` — уточнить данные.
- `FLAG` — продолжить, явно пометив допущение/риск.
- `REFUSE` — не выполнять запрос, который нарушает границы/правила.
- `ESCALATE` — передать профильному человеку/владельцу процесса.

Сценарий	Реакция	Текст/правило	Кто подтверждает
Не хватает критического факта	ASK		
Источники конфликтуют	ESCALATE		
Запрос за пределами роли	REFUSE/redirect		
Высокорисковое решение	ESCALATE		
Запрос реального внешнего действия	REFUSE/agent-route		
Чувствительные данные	ESCALATE		
Пользователь просит придумать факт	REFUSE		
Можно продолжить с безопасным допущением	FLAG		