

Killswitch Gate — PreToolUse + UserPromptSubmit

Killswitch Gate — PreToolUse + UserPromptSubmit

Comprehensive token burn prevention via fail-closed killswitch gate in BOTH hook events.

MC #101650 — PreToolUse Consolidation (2026-05-21)

Verdict: PASS

Task: [P3-4] Hook consolidation: merge PreToolUse matchers, eliminate 6x killswitch + duplicate fires

Summary

`~/.claude/settings.json` had `killswitch-gate.sh` repeated in every PreToolUse matcher group. For `Task`, multiple matcher groups also matched, causing duplicate killswitch execution. The PreToolUse killswitch is now centralized in one universal matcher while specialized gates remain in their original matcher groups.

Change

- Added first PreToolUse entry: matcher `.*` → `bash $HOME/system/hooks/killswitch-gate.sh`.
- Removed `killswitch-gate.sh` from the specific PreToolUse matcher entries:
 - `Bash`
 - `Task|WebSearch|WebFetch`
 - `Task`
 - `mcp__playwright__.*`
 - `Write|Edit|MultiEdit`

- Backed up settings to `~/.claude/settings.json.bak-101650`.
- Restored `uchg` immutable flag on `~/.claude/settings.json`.

Validation

- `settings.json` JSON-valid.
- PreToolUse killswitch occurrences reduced from `5` to `1`.
- Representative matching analysis shows exactly one PreToolUse killswitch for:
 - Bash
 - Task
 - WebSearch
 - WebFetch
 - `mcp_playwright_click`
 - Write/Edit/MultiEdit
- Killswitch OFF smoke: `killswitch-gate.sh` exits `0` with empty stdout/stderr.

Evidence

- `/tmp/101650-hook-inventory.txt`
- `/tmp/101650-post-consolidation-analysis.txt`
- `/tmp/101650-smoke-validation.txt`
- `/tmp/101650-evidence/`

MC #103690 — UserPromptSubmit Gate Addition (2026-06-19)

Verdict: PASS

Task: `killswitch-gate.sh` added to UserPromptSubmit – halt prompts when killswitch engaged

Problem

`killswitch-gate.sh` was registered only in PreToolUse (settings.json), NOT UserPromptSubmit. An engaged killswitch (`~/system/state/killswitch`) blocked tool use but NOT prompt submission — prompts still went through and burned tokens.

Fix

- Added `killswitch-gate.sh` as the FIRST hook in the UserPromptSubmit chain in `~/.claude/settings.json`:
 - Command: `bash $HOME/system/hooks/killswitch-gate.sh`
 - Timeout: `5000`
- `settings.json` is `uchg`-immutable (anti-tamper). Edit procedure:
 - `chflags nouchg ~/.claude/settings.json` → edit → `chflags uchg ~/.claude/settings.json` (re-lock)
 - Back up first

Gate Behavior (Verified)

- Fast-path `exit 0` if `~/system/state/killswitch` absent (fail-open, no stdin parse) → safe in UserPromptSubmit
- When engaged → `exit 2` + stderr `KILLSWITCH:ENGAGED` + UserPromptSubmit JSON `{"hookEventName": "UserPromptSubmit", "permissionDecision": "deny"}`
- Verified via:
 - Direct test: no-killswitch → `exit 0`, engaged → `exit 2`
 - `lint-killswitch-preamble.sh` PASS: "OK (first): UserPromptSubmit[]"

Result

Engaged killswitch now halts BOTH prompts (UserPromptSubmit) and tool use (PreToolUse).

Tools

- Canonical install: `~/system/tools/install-killswitch-settings.sh`
- Lint: `~/system/tools/lint-killswitch-preamble.sh`
- CLI: `~/system/tools/killswitch.sh on|off|status`

Evidence

- `~/.claude/settings.json` (UserPromptSubmit first hook)
- `~/system/tools/lint-killswitch-preamble.sh` PASS
- Direct test: `exit 0` (no killswitch), `exit 2` (engaged)

Revision #2

Created 2026-05-21 12:37:18 UTC by John

Updated 2026-06-19 17:58:53 UTC by John