Secure Development with Claude AI

API Security Checklist · 22 items across 6 categories

22-item pre-deployment checklist

resource.7312.us/guide/secure-claude-dev.html

API KEY SECURITY

- [ ] Store API key in environment variables, never in source code — CRITICAL
- [ ] Ensure API key is never sent to or accessible by the client/browser — CRITICAL
- [ ] Rotate API key immediately if it is ever exposed — CRITICAL
- [ ] Use separate API keys for development, staging, and production — IMPORTANT

INPUT VALIDATION & PROMPT INJECTION

- [ ] Wrap user-supplied content in XML delimiters to separate it from instructions — CRITICAL
- [ ] Validate and sanitize all user input before including it in prompts — CRITICAL
- [ ] Never allow user input to overwrite or extend your system prompt directly — CRITICAL
- [ ] Test prompts with adversarial inputs before shipping — IMPORTANT

DATA MINIMIZATION & PRIVACY

- [ ] Send only the minimum data required for the task — no full database rows — CRITICAL
- [ ] Strip or mask PII, credentials, and secrets before including in prompts — CRITICAL
- [ ] Review Anthropic's data retention and processing policies for your use case — IMPORTANT
- [ ] Log API inputs/outputs only to secured, access-controlled storage — IMPORTANT

LEAST PRIVILEGE & AGENTIC SAFETY

- [ ] Grant Claude only the tools it strictly needs — remove all others — CRITICAL
- [ ] Require human confirmation before any irreversible agentic actions — CRITICAL
- [ ] Set rate limits and usage caps to prevent runaway API costs — IMPORTANT

OPERATIONAL & OUTPUT SAFETY

- [ ] Validate and review model outputs before rendering or acting on them — IMPORTANT
- [ ] Use an allowlist of permitted action types — never a blocklist — IMPORTANT
- [ ] Implement output filtering for use cases that serve sensitive audiences — IMPORTANT
- [ ] Set up monitoring and alerting for unusual usage patterns — GOOD PRACTICE

COST & RATE LIMIT PROTECTION

- [ ] Always set max_tokens explicitly on every API request — IMPORTANT
- [ ] Enforce a hard iteration cap on all agentic loops (e.g. max 10 turns) — CRITICAL
- [ ] Set a monthly spending cap and alert thresholds in the Anthropic console — IMPORTANT