

Four frontier models, in parallel. From a fraction of a cent.

crosscheck is an MCP server that orchestrates **Anthropic, OpenAI, Gemini, and Grok** through five battle-tested patterns — right inside Claude Code, Cursor, or any MCP client. It runs on **your machine** with **your keys**, and every call returns a per-provider **cost + latency receipt**. Multi-model quality, single-model economics.

FIVE PATTERNS

confer

Ask the panel in parallel. Surface diverging views fast.

debate

Bounded rounds; early-stop on convergence. One clear call.

audit

One proposes, others review. Catches its blind spots.

plan

Collaborative planning, structured for execution.

panel-pick

Router picks the right panel by purpose + reliability.

WHY IT'S EXTREMELY CHEAP ON TOKENS

panel-pick — pay only for the models you need

A smart router selects the smallest panel that clears the quality bar by purpose and historical reliability — lower cost for equivalent answers, no fixed four-model tax on every call.

debate — early-stop when models converge

Rounds auto-halt the moment the panel agrees. A recorded debate converged at round 1 and skipped round 2 outright — you stop paying for tokens that wouldn't change the answer.

create_cheap — cheap tier drafts, one pass reviews

Low-cost models write the code; a single synthesis pass merges and reviews it. A complete, SSR-safe TypeScript hook came back for **\$0.23**.

audit — a full security review for ¢

A production JWT auth-bypass, caught and scored against a rubric, for **\$0.0024** — 308 in / 468 out, 5.2s.

REAL RECORDED RECEIPT · XC CONFER

4 frontier models design a spend dashboard, in parallel				
provider	in	out	cost	wall
anthropic/opus-4.7	163	2,048	\$0.1560	27.8s
openai/gpt-5.5	107	3,997	\$0.0605	66.0s
xai/grok-4	228	520	\$0.0086	6.0s
google/gemini-3.1-pro	104	1,319	\$0.0067	26.1s
total	602	7,884	\$0.2319	66s wall

Four models' worth of judgment — every take, the consensus, schema, gotchas, and a code stub — for **23¢** and one 66-second wall-clock wait (they run concurrently, so you don't pay in time for the extra models). Cost and latency scale with prompt + output size; every call returns its own receipt like this one.

WHERE CROSSCHECK SHINES

► **High-stakes decisions**

Architecture calls, library picks, security reviews — get four independent perspectives with the dissent surfaced, not averaged away.

► **Private by construction**

Server never sees prompts or keys. Content-free telemetry rejects prompt-shaped data at the boundary; keys stay in a local encrypted store.

► **Zero lock-in economics**

\$20/seat/month for the orchestration. Your model spend runs through your own provider accounts — the bill, and the keys, stay yours.

Frank Speiser

frank.speiser@gmail.com

crosscheck · multi-LLM orchestration

\$ npx -y crosscheck-cli@latest install
crosscheckagent.com · first month free · WELCOME100