

Hermes-vloot audit

Volledige controle van je Hermes Desktop-setup (7 bots + de root Hermes-bot), getoetst aan de officiële best-practice en aan hoe jij werkt. Onderzocht met 4 read-only audit-teams. Elke bevinding heeft een bronpad.

Legenda: [GEFIXT] heb ik al gedaan · [VEILIG] kan ik zo doorvoeren · [JOUW KEUZE] vraagt jouw beslissing.

Kort: het goede nieuws

- Alle bots staan op `openai_runtime: auto` → geheugen werkt weer (de fix van vandaag houdt).
 - Geen dood model meer in de actieve configs (de x-preview-fout is echt weg).
 - App-bots doen geen eigen Telegram (jouw regel wordt gerespecteerd).
 - Actieve secrets staan op 0600, geen hardcoded sleutels, secret-scanner draait.
 - De 7 profielen zijn compleet, geen spookprofielen. De backend draait.
-

1. Veiligheid

[GEFIXT] Wereld-leesbare secret-backups

5-6 backup-bestanden (`.env.env.bak-voor-gbraintoken`) met ~27 secrets, waaronder je Telegram-token, stonden op 0644 (leesbaar voor elk proces). **Dichtgezet naar 0600.**

- [JOUW KEUZE] die Telegram-token heeft een tijd open gestaan. Wil je 'm laten roteren? (Raakt de fleet, dus jouw beslissing.)

[JOUW KEUZE] Dirigent = YOLO op de machtigste bot

`dirigent/config.yaml` : `command_allowlist`: - '*' (alles vooraf goedgekeurd) + `approvals.mode`: `false` . En juist die bot heeft alle MCP's (supabase, github, wordpress, transip, render, cloudflare). Een fout of prompt-injection daar raakt direct productie.

- Advies: '*' eruit, `approvals.mode: smart` voor dirigent. Botst nu ook met je "alleen Jamal doet exit"-regel.

[VEILIG] Vangnet tegen zelf-uitzetten ontbreekt

Er is nergens een `approvals.deny` -lijst. Dat is de enige lijst die óók bij `mode: off` hard blokkeert. Precies het vangnet dat je guard-regel bedoelt.

- Fix: top-level een `approvals.deny` met o.a. `launchctl *disable*`, `launchctl *bootout*`, `*kill* nl.haarvisie.*`, `git push --force*`.

[LET OP] `approvals.mode: false` is geen geldige waarde

De docs kennen alleen `smart`, `manual`, `off`. `false` is een boolean → gedrag onzeker. Vervang door `off` of `smart`.

2. Geheugen, skills en rol-afbakening (jouw overlap-probleem)

[JOUW KEUZE — grootste hefboom] Skills niet afgebakend

Elke bot draagt bijna dezelfde ~19 skill-bundels. Beeld heeft dev-skills, Web heeft beeld-skills, School heeft alles. **Dit is de motor achter het overnemen.**

- Voorstel per discipline (alleen eigen vak aan):
- **Beeld:** creative, media, lovart-api, social-media, haarvisie
- **Web:** software-development, github, devops, mlops, data-science, research, haarvisie
- **School:** haarvisie (podcast), research (licht), note-taking
- **Content (beide):** creative (alleen blog-writing + humanizer), research, media (youtube), social-media, haarvisie (report/podcast)
- **Overall UIT:** smart-home, apple, **email** (mail = alleen Lenovo/Conductor, teamregel), yuanbao
- Aanpak: eerst 1 bot als proef (School heeft het meeste te winnen), testen, dan de rest.

[VEILIG] "Blijf in je vak"-regel ontbreekt in elke SOUL

Geen enkele bot krijgt te horen: doe alleen jouw vak, geef ander werk door via @mention. Bot Mode (@mention) werkt technisch wél (staat default aan), maar de bots weten het niet.

- Fix: een vast "Blijf in je vak"-blok in elke discipline-SOUL, met @mention naar de juiste bot en @dirigent bij twijfel.

[VEILIG] contentcodex heeft géén geheugen

Geen memory-provider, lege memories/. Fix: memory.provider: hermes-mnemosyne + een MEMORY.md/USER.md.

[VEILIG] Dirigent z'n geheugen is 9x te groot

dirigent/memories/MEMORY.md = 20.218 tekens (limiet 2.200). Het hele team-regelboek staat er inline in. Gevolg: elke prompt is opgeblazen én de volgende geheugen-schrijf geeft een fout. Fix: regelboek eruit (staat al in gBrain), MEMORY.md slank maken.

[LET OP] beeld-geheugen bijna vol

MEMORY.md 92%, USER.md 95%. Preventief opschonen.

[JOUW KEUZE] Geheugen-goedkeuring

write_approval staat overal uit (bots schrijven vrij naar geheugen). Jij wilde kunnen goedkeuren. Zetten we memory.write_approval: true aan? Dan review je met /memory pending .

[VEILIG] oxlab + email

oxlab staat in de bot-roster en is @mentionbaar terwijl hij bewust geïsoleerd is → verbergen in de app. De email -bundel staat overal aan terwijl mail teamregel-verboden is → eruit.

3. Modellen, runtime en kosten

[VEILIG] contentcodex mist de codex-veiligheidsremmen

Kale config zonder tool_output -caps (valt terug op 50000/2000 i.p.v. de vloot-standaard 15000/800). Risico op de codex-overflow ("wartaal"). Fix: het codex-blok van web/school overnemen.

[VEILIG] show_cost staat overal uit

Je ziet nergens een dollarbedrag, ook niet bij contentclaude (de enige betaalde bot). Fix: `show_cost: true` op contentclaude.

[VEILIG] contentclaude draait dure Opus ook voor achtergrondwerk

De achtergrond-review + titels draaien op het hoofdmodel = betaalde Opus. Fix: die hulp-taken naar het gratis Codex-model, of de achtergrond-review uitzetten voor deze bot.

4. Installatie en werking

[JOUW KEUZE — belangrijk] Terugkerende fout: "codex model not supported"

Sommige interne hulp-calls sturen het model letterlijk als `codex`, en je ChatGPT-account weigert dat. Gevolg: **beeldanalyse (14x), titels (13x), achtergrond-reviews en compressie vallen stil**. Fix: die hulp-taken een expliciet ondersteund model geven (`gpt-5.6-sol`) i.p.v. de kale `codex`-naam. Dit wil ik graag precies goed doen, dus leg ik het je voor.

[JOUW KEUZE] Install loopt ver achter + update van 21-8 faalde

De installatie is 1421 commits achter, en de laatste update (21-8) mislukte op een git-fetch-fout. Er staan ook 2 lokaal gewijzigde kernbestanden die bij een blinde update geparkeerd worden. Updaten is aan te raden, maar zorgvuldig — niet blind.

[JOUW KEUZE] Codex-login "expired" + gBrain MCP verbindt niet

De "auth failed" is bij Codex vaak eigenlijk credits op, niet echt uitgelogd (bekende valkuil). gBrain MCP (`100.88.240.64:8788`) gaf verbindingfouten. Beide even verifiëren.

[JOUW KEUZE] ~7 GB rommel

Een oude pre-update-backup van 988 MB (21-8), plus veel losse `.bak` / `.tmp`-bestanden. Opruimbaar zodra de volgende update goed staat.

[OK] Cron draait leeg

De enige cron-job (SEO-wachter) staat uit sinds 23-7 (budget op). Weer aanzetten of weggooien = jouw keuze.

Voorgestelde volgorde

1. **Nu al gedaan:** geheugen-runtime fix, comms-regels, security-backups dichtgezet, dood model weg.
2. **Veilig, meteen door te voeren:** "blijf in je vak"-regel, contentcodex remmen + geheugen, dirigent-geheugen slanken, show_cost, contentclaude-aux naar gratis, approvals.deny-vangnet.
3. **Jouw beslissingen:** skill-scoping (proef op School), geheugen-goedkeuring, dirigent-rechten inperken, de "codex model"-fout fixen, Hermes updaten, opruimen, token roteren.

Alles reversibel (backups per wijziging).