Wat kosten mijn automatiseringen en welke kunnen weg

Basis: het complete overzicht van ± 180 automatiseringen (`/tmp/fleet-automatiseringen.md`).
Dit verslag gaat een laag dieper: ik heb de logs en scripts van de MacBook echt gelezen (dat is waar de kosten zitten — 66 taken, waarvan er 19 in het eerdere overzicht als "AI-aanroepend" stonden). VPS/iMac/Sjakie behandel ik korter, zoals gevraagd.

---

1. Samenvatting

Belangrijkste correctie eerst: van de 19 taken die in het eerste overzicht "AI = ja" kregen, riepen er **5 helemaal GEEN AI aan** toen ik het script echt las: `transcript-watchdog` , `vastloper` , `cost-dashboard-report` , `cost-weekly-alert` en `cockpit` .
Die lezen alleen bestandsgroottes, processen of een Supabase-tabel uit. Geen tokens, geen kosten.

Van de resterende taken:
- **Claude/Anthropic (telt mee voor je Max-abo of is echt betaald via API-key)**: ongeveer **50.000 tokens/dag** aan geplande rapport-taken (research-brief, usage-report, actiepunten-check, vragen-sweep) + **tot ~300.000 tokens/dag** aan de conductor-queue bot-verkeer (zie hieronder — dit is kernwerk, geen rapport).
- **Gemini (los, goedkoop, niet je Claude-budget)**: briefings (2x/dag), sjakie-verslagen (1x/dag maar herhaalt zich onnodig, zie kandidaat 2), en incidenteel de AI Report-analyse.
- **Gratis/lokaal**: de dagelijkse podcast-tekst (Ollama) en Ollama zelf.
- **Kimi (eigen abonnement, niet Anthropic)**: kimi-bot en kimi-queue.

Dus: het bedrag dat echt van je Claude Max-quota afgaat voor **rapportage-achtige taken** (niet de bot-communicatie zelf) schat ik op **~50.000 tokens/dag**, oftewel **~1,5 miljoen tokens/maand**. Dit is een SCHATTING — zie sectie 5 voor de aannames.

Weggooi-kandidaten gevonden: 3 (zie sectie 3). Geen van de drie is een bot, poller of watchdog die de vloot draaiende houdt — dus veilig om te bekijken.

2. AI-aanroepende taken op de MacBook — wat ik echt gemeten heb

TAAK	ROEPT AAN	FREQUENTIE	LAATSTE BEWEZEN ACTIVITEIT	OUTPUT GAAT NAAR
conductor-queue	`claude -p --model sonnet` (Max-abo)	poll elke 2 min, AI alleen bij echt bericht	vandaag 21 echte antwoorden (log), begrensd op ~20/dag door loop-guard	Terug naar de vragende bot (Supabase-queue)
daily-research-brief	`claude -p` (Max-abo)	1x/dag 08:00	log 26-7 08:00, gedraaid	Telegram + Obsidian `05_Daily/…-research-brief.md`
usage-daily-report	`claude -p /usage` (Max-abo)	1x/dag 07:00	log 26-7 07:00	Telegram
actiepunten-3daags-check	`claude -p` (Max-abo)	gepland elke 3 dagen	log: 8, 11, 14, dan pas weer 24-7 (gat van 10 dagen — liep niet altijd)	Telegram
vragen-sweep	Anthropic API direct met **eigen key**, model `claude-haiku-4-5` (goedkoopste model) — dus ECHT betaald, niet Max-abo. Skipt zichzelf als er niets openstaat	1x/dag 08:30, alleen als er iets is	log 26-7 08:30	Telegram-archief

TAAK	ROEPT AAN	FREQUENTIE	LAATSTE BEWEZEN ACTIVITEIT	OUTPUT GAAT NAAR
	(`return` in de code)			
jamal-prive-bot	`claude` CLI per binnenkomend Telegram-bericht	on-demand	**log toont alleen verbindingsfouten sinds 24-7, geen bewijs van een echt beantwoord bericht**	Telegram
kimi-queue	`kimi -p` (eigen Kimi-abonnement, NIET Anthropic)	poll elke 2 min, AI alleen bij bericht	vandaag 3 echte antwoorden	Terug naar vragende bot
kimi-bot	Kimi CLI per Telegram-bericht	on-demand	49 logregels vandaag (mix van berichten)	Telegram
briefing-ochtend + briefing-avond	Gemini API (`gemini-2.5-flash`), niet Anthropic	2x/dag (07:30, 19:00)	beide gedraaid vandaag (log 19:02)	Telegram + podcast.haartips.n
sjakie-verslagen-gen	Gemini API	1x/dag 23:15	log 25-7 23:21	JSON naar VPS → sjakie-podcast
daily-podcast	Tekst via **lokale Ollama** (gratis) + Gemini-TTS (volgens het script zelf "~€1-2/mnd")	1x/dag 08:00	log 27-7 08:58	Telegram

TAAK	ROEPT AAN	FREQUENTIE	LAATSTE BEWEZEN ACTIVITEIT	OUTPUT GAAT NAAR
aireport-analyse	Gemini, maar alleen als er een nieuwe AI Report-mail is	elke 30 min gepland, check is IMAP (geen AI)	30 checks/dag, in de laatste 500 logregels **0x** een echte Gemini-aanroep	verslag.haartips.n + Telegram (alleer als er wat is)
ollama-serve	is zelf het lokale model	continu draaiend	actief (proces + log vandaag)	—

---

3. WEGGOOI-KANDIDATEN

Op volgorde van meest zinloos eerst.

1. `uptime-monitor` — waarschijnlijk kapot, niet alleen uit

- Staat gepland om elke 30 minuten te draaien (`StartInterval = 1800`).
- Log is al **12+ dagen stil** (laatste regel 15 juli, vandaag is 27 juli) — bij een taak die elke 30 min zou moeten draaien is dat een sterk signaal dat hij niet meer loopt, niet alleen dat er niks te melden was.
- Ik kon niet zien of hij nog echt in launchd geladen is of stilletjes gestopt is.
- **Besparing**: geen tokens (roept geen AI aan), maar wel opruiming — of weer aanzetten, of het bestand weggooien. Nu doet hij niets van beide nuttig.
- Let op: dit is GEEN AI-kosten-besparing, puur opruiming.

2. `sjakie-verslagen-gen` — geen weggooien, wel repareren (onnodig Gemini-verbruik)

- Dit script heeft **geen voortgangs-check**: bij elke run (elke nacht 23:15) haalt het ALLE Sjakie-berichten ooit op, knipt ze in blokken van 4 dagen, en laat Gemini voor ELK blok — ook blokken van weken geleden die al eerder samengevat zijn — opnieuw een samenvatting schrijven.
- Dat is pure herhaling: dezelfde oude data wordt elke nacht opnieuw naar Gemini gestuurd.

- Dit is geen "weggooien"-kandidaat (de output — Sjakie-podcasts — wordt echt gebruikt), maar wel een makkelijke besparing: een `laatst-verwerkt`-datum bijhouden en alleen nieuwe blokken versturen.
- **Besparing**: bij een groeiende geschiedenis wordt dit script elke maand duurder in Gemini-aanroepen zonder dat de output verandert. Exacte huidige kosten kon ik niet meten (Gemini-verbruik zit niet in de Anthropic-kostenlogs die ik kon inzien).

3. `jamal-prive-bot` — twijfelachtig, mogelijk overbodig geworden

- De enige logregels van de laatste dagen (24 t/m 26 juli) zijn Telegram-verbindingsfouten ("Connection reset", "Read timed out") — geen enkel bewijs van een echt beantwoord bericht.
- In het geheugen staat een aparte notitie "privé → Jay" die erop wijst dat privé-verkeer inmiddels via de Jay-Privé-bot loopt in plaats van deze.
- Er ligt ook een backup-plist met de naam `...bak-key-leak-20260706-1319` naast de actieve
  versie — een aanwijzing dat dit ooit een key-lek had en misschien sindsdien niet meer vertrouwd/gebruikt wordt.
- **Ik weet niet zeker of hij écht dood is** — de foutmeldingen kunnen ook een tijdelijk netwerkprobleem zijn. Vraag het na voor je 'm weghaalt: is dit dezelfde bot als Jay-Privé, of een oude losse versie die vervangen is?
- **Besparing als hij dood is**: minimaal aan tokens (hij wordt toch al niet gebruikt), vooral opruiming + een bron van verwarring minder.

---

4. Stil maar kritiek — NIET weggooien

Deze zag ik ook met (relatief) stille logs of lage AI-aanroep-frequentie, maar ze horen bij de vloot in de lucht houden — nooit uitzetten:

- **hermes-web-poller** — is vandaag (27-7) al bewust uitgezet door iemand in de vloot (bestand hernoemd naar `...plist.UIT-20260727`). Geen actie meer nodig, dit is al gedaan.
  Historisch riep hij trouwens zelden echt AI aan (16 keer in de hele levensduur van het log, ondanks dat hij elke 3 seconden draaide) — hij deed vooral goedkope status-checks.

- **conductor-queue / kimi-queue** — pollen elke 2 minuten maar roepen alleen AI aan bij een echt binnenkomend bericht. Dit is de ruggengraat van bot-naar-bot-communicatie (het Team-Regelboek verplicht dit kanaal). Nooit uitzetten.
- **aireport-analyse** — draait elke 30 min maar doet bijna altijd alleen een gratis IMAP-check zonder AI. Ziet er duur uit door de frequentie, is het niet.
- **ollama-serve** — continu aan, maar gratis en lokaal; nodig voor daily-podcast en andere lokale AI-taken.
- **transcript-watchdog / vastloper** — draaien vaak (elke 5 min) maar roepen GEEN AI aan (zie correctie in sectie 1). Beschermen tegen sessie-crashes. Nooit uitzetten.

---

5. Aannames en onzekerheden

- **15.000 tokens per `claude -p`-aanroep** is de vaste aanname die gevraagd is (input + output + systeemprompt). Ik heb dit niet per taak apart gemeten — het is een ruwe vlag-schatting, geen echt tokenrapport.
- De meeste `claude -p`-aanroepen op de MacBook lopen via het **Max-abonnement**, niet de betaalde API — dus ze kosten geen directe euro's per token, maar tellen wel mee voor je wekelijkse Max-limiet. Alleen `vragen-sweep` gebruikt een echte betaalde API-key (met het goedkoopste model, Haiku).
- Gemini-kosten (briefings, sjakie-verslagen, podcast-TTS, aireport) zitten in een ANDER budget dan je Claude-tokens — ik heb ze niet omgerekend naar euro's omdat ik de actuele Gemini-prijzen niet zeker weet en geen cijfer wil verzinnen. Het enige harde cijfer dat ik vond staat letterlijk in het `daily-podcast.py`-script zelf: "~€1-2/mnd" voor de TTS.
- Voor **conductor-queue** heb ik de 21 aanroepen van vandaag (27-7) als dagcijfer gebruikt. Dat is 1 dag meten, geen gemiddelde over meerdere dagen — kan op andere dagen lager of (tot de ingebouwde limiet van ~20/uur-per-bot, 20 totaal) hoger uitvallen.
- Kan mijn log-check een vals "dood" geven? Ja: als een machine een paar uur uit stond of in slaap was, lijkt een taak stil terwijl hij morgen gewoon weer aanslaat. Bij `uptime-monitor` (12+ dagen stil bij een 30-min-schema) is dat onwaarschijnlijk als verklaring, maar bij kleinere gaten (zoals bij `actiepunten-3daags-check`) hou ik

rekening
met gemiste runs door slaapstand, niet per se een kapotte taak.

- VPS, iMac, Sjakie en Dani's pc heb ik NIET opnieuw diep gecontroleerd — daarvoor geldt nog
  steeds het eerdere overzicht in `/tmp/fleet-automatiseringen.md`, inclusief de daar al genoemde open vraag welke VPS-scripts een Claude-aanroep verbergen zonder het woord
  "claude"/"anthropic" te bevatten.