

Implementatieplan — 3 AI-ontwikkelingen voor Jamals setup

Ik heb je echte setup bekeken.

Je queue is **Supabase agent_messages** (velden: from_agent, to_agent, message, status).

Je bots draaien als **LaunchAgents** (o.a. conductor-autostart, fleet-heartbeat, watchdogs).

Daar sluit dit plan precies op aan.

Eén eerlijkheidsnoot vooraf.

Harde tarieven voor Kling 3.0 en GPT-5.6 verzin ik niet.

Ik zeg per punt wáár je het exacte tarief checkt.

Punt A — GPT-5.6 Terra als goedkoper model voor lichte bot-taken

Wat het oplevert

- Veel bot-taken zijn licht: queue pollen, status melden, korte samenvatting.
- Daar heb je geen duur topmodel voor nodig.
- Terra is ~2x goedkoper dan 5.5, bijna zelfde kwaliteit.
- Op lichte taken = directe besparing op je API-rekening.

Hoe het in je setup past

- Het raakt **niet** je zware bots (Conductor, code-werk blijft Claude).
- Het raakt de **lichte, herhaalde taken**: pollers, heartbeats, korte meldingen.
- Je verandert niet de bots zelf — je zet er een **model-router** tussen.

Belangrijk nu

- GPT-5.6 is ECHT, maar nog **niet algemeen beschikbaar** (eerst partners).
- Je kunt dus nog niet live testen.

- Wel: **nu al de router bouwen**, zodat je klaar bent zodra Terra opengaat.

Concrete stappen

1. Maak een klein routerscript: `~/ .claude/scripts/model-router.py` . *(klein)*
2. Eén functie: `pick_model(taak_type)` → geeft model + provider terug. *(klein)*
3. Drie niveaus, simpel:
 - zwaar (code, redeneren) → Claude
 - middel → later Terra
 - licht (poll/melding) → goedkoopste *(klein)*
4. Bouw een **fallback**: lukt provider niet → val terug op Claude. *(middel)*
5. Voeg een **env-schakelaar** toe: `USE_GPT=0/1` . Zo zet je het uit/aan zonder code te wijzigen. *(klein)*
6. Test eerst op **1 onschuldige taak** (bijv. een heartbeat-tekst), niet op de hele vloot. *(klein)*
7. Log per call: model, tokens, tijd. Vergelijk kosten na 1 week. *(middel)*
8. Pas als Terra opengaat: zet `USE_GPT=1` voor alleen die ene taak. Bevalt het → breid uit. *(klein)*

Kosten-impact

- Bouwen kost niets (alleen tijd).
- Live draaien: bespaart op lichte taken zodra Terra er is.

Prioriteit

- Router bouwen: **nu** (je bent dan voorbereid).
- Terra echt inzetten: **wachten-tot-beschikbaar**.

Risico's / valkuilen

- Niet alles tegelijk omzetten — anders breekt de vloot.
- Begin met één taak.
- Houd Claude als vangnet (fallback verplicht).
- Tarief Terra nog onbekend → check straks op het officiële OpenAI-tariefoverzicht vóór je opschaalt.

Punt B – Kling 3.0 voor realistische haar-video's (@haartipsnl / TikTok)

Wat het oplevert

- Kling 3.0 is nu het beste voor háár: het beweegt natuurlijk mee.
- Native 4K.
- Perfect voor jouw social: een foto die "tot leven komt".

Hoe het in je setup past

- Je gebruikt Kling al via **fal** — top, geen nieuwe tool nodig.
- Beeld blijft via **Gemini** (gemini-2.5-flash-image, curl).
- Workflow = foto met Gemini → video met Kling 3.0.

Concrete workflow (foto → video → social)

1. Maak/kies de **startfoto** via Gemini (curl), brand-stijl: natuurlijk, echte vrouw, effen studio. *(klein)*
2. Verklein binnenkomende beelden tot max 2000px (anders crasht de context). *(klein)*
3. Stuur de foto naar **Kling 3.0 image-to-video** op fal. *(klein)*
4. Houd de beweging **subtiel**: haar dat valt, lichte draai van het hoofd. Geen wilde camera. *(klein)*
5. Genereer kort: 5 sec is vaak genoeg voor een reel/TikTok. Kort = goedkoper. *(klein)*
6. Check het resultaat eerst zelf (visual QA) vóór je iets oplevert — bewijs verplicht. *(klein)*
7. Lever aan met de klikbare fal-output-link erbij. *(klein)*
8. TikTok-links alleen mobiel openen (je eigen regel). *(klein)*

Belangrijke merk-regels (jouw eigen contract)

- Instagram Platina Blond = **nooit zelf posten**.
- Schrijf "tarieven", nooit "prijzen".
- Gebruik altijd het echte logo + merklettertypes.

Kosten-impact

- Kling 3.0 is **duurder per seconde** dan Kling 2.1.

- Audio-aan en Pro-modus kosten extra.
- Bespaar door: korte clips, audio alleen als je het nodig hebt.
- Exact tarief: check per endpoint op fal — link:
<https://fal.ai/models/fal-ai/kling-video/v3/standard/image-to-video>
 (Pro-variant: .../v3/pro/image-to-video)
- Ik noem bewust geen hard bedrag; tarieven wijzigen en ik wil niets verzinnen.

Prioriteit

- **Nu** — je hebt fal al, dit kan vandaag.

Risico's / valkuilen

- Te veel beweging = onnatuurlijk haar. Houd het subtiel.
- Pro + audio loopt snel op in kosten. Begin met Standard, kort.
- Maak eerst 1 testclip vóór een hele reeks.

Punt C — DeepMind veiligheid vertaald naar je bot-leger

DeepMind's roadmap heeft 4 ideeën. Ik vertaal ze naar jouw queue + LaunchAgents.

1. Noodstop per bot

Wat: elke bot apart kunnen stilleggen, zonder de rest te raken.

Hoe het past: je bots zijn losse LaunchAgents. Dat is je natuurlijke aan/uit-knop.

Stappen:

1. Maak een script `fleet-stop.sh` met per bot: `launchctl bootout (stop)` en `bootstrap (start)`. *(klein)*
2. Voeg een **kill-switch via de queue** toe: een rij `to_agent=<bot>, message="STOP"` .
(middel)
3. Elke bot checkt bij elke poll-ronde eerst op een STOP-bericht. Ziet hij STOP → hij stopt netjes. *(middel)*
4. Belangrijk: dit botst niet met je regel "nooit zichzelf afsluiten zonder toestemming" — want **jij of Conductor** geeft de STOP, niet de bot zelf. *(klein)*

Moeite: middel. *Kosten:* geen.

2. Live supervisor-bot

Wat: één bot die meekijkt wat de anderen doen.

Hoe het past: je hebt al `fleet-heartbeat`. Bouw daarop voort.

Stappen:

1. Laat elke bot zijn actie loggen in een tabel `agent_actions` (wie, wat, wanneer). *(middel)*
2. Maak een **supervisor-rol** (kan Conductor zijn, of een aparte lichte bot). *(middel)*
3. Supervisor leest de log elke X minuten. *(klein)*
4. Ziet hij iets vreemds (te vaak, te duur, verboden domein) → meldt het in jouw Telegram (genummerd). *(middel)*
5. Tip: laat de supervisor op een **goedkoop model** draaien (straks Terra) — past mooi bij Punt A. *(klein)*

Moeite: middel. *Kosten:* klein (extra calls), beperkt op goedkoop model.

3. Hoge-risico-acties vooraf laten goedkeuren

Wat: gevaarlijke acties mogen pas ná jouw "ja".

Hoe het past: via je bestaande queue + status-veld.

Stappen:

1. Maak een lijst **hoge-risico-acties**. Denk aan:
 - iets naar buiten sturen (mail/DM)
 - bestanden verwijderen
 - deploy / live zetten
 - geld/dure API-calls *(klein)*
2. Voor zo'n actie schrijft de bot eerst een rij: `status="needs_approval"`. *(middel)*
3. De bot **wacht** — hij voert niets uit zolang status niet "approved" is. *(middel)*
4. Jij krijgt een Telegram: "Bot X wil Y doen. Goedkeuren?" (genummerd). *(klein)*
5. Jij zet `status="approved"` of `"denied"` (via een simpel commando of knop). *(middel)*
6. Pas dan voert de bot uit. Bij "denied" → stopt en meldt het. *(klein)*

Dit dekt meteen 2 van je eigen kernregels af:

- geen cold outreach zonder toestemming
- haarvisie.nl off-limits

Moeite: middel-groot. *Kosten:* geen.

4. Insider-threat-aanpak (een bot die ontspoort)

Wat: bescherming als één bot raar of schadelijk gedrag vertoont.

Stappen:

1. **Secrets afschermen:** bots lezen alleen wat ze nodig hebben. Keychain, nooit in de queue (al je regel). *(klein)*
2. **Geen secrets in logs** — controleer dat de actie-log geen tokens lekt. *(klein)*
3. **Rechten per bot:** Coco/Julio hebben geen deploy-rechten als ze die niet nodig hebben. *(middel)*
4. Supervisor (punt 2) let op afwijkend patroon → automatische noodstop (punt 1) als het echt misgaat. *(middel)*

Moeite: middel. *Kosten:* geen.

Prioriteit Punt C

- Noodstop per bot: **nu** (klein, groot effect).
- Goedkeuring hoge-risico: **nu/binnenkort** (beschermt je grootste risico's).
- Supervisor-bot: **binnenkort**.
- Insider-aanpak: **binnenkort** (deels al gedekt door je regels).

Samenvatting — wat ik je adviseer

Doe nu:

- **B:** eerste Kling 3.0 testclip (je kunt vandaag).
- **C:** noodstop per bot + goedkeuring voor hoge-risico-acties.
- **A:** model-router bouwen (klaar voor Terra).

Doe binnenkort:

- **C:** supervisor-bot + insider-aanpak.

Wacht op beschikbaarheid:

- **A:** Terra echt inzetten zodra OpenAI 'm opent.

Wat ik van jou nodig heb (beslissingen)

1. Mag ik de **model-router** (Punt A) alvast als script neerzetten? (*verandert niks aan draaiende bots, staat uit met `USE_GPT=0`*)
2. Voor welke bot wil je de **noodstop** als eerste testen?
3. Welke acties tel jij als **hoge-risico** (naast: outreach, deploy, verwijderen, haarvisie.nl)?
4. Wil je dat ik **eerst een testclip met Kling 3.0** maak? Zo ja: van welke foto?

— Hermes