



# Debrief — qwen/qwen3.6-plus

---

- Instanser: 9
  - Tokens inn: 12,298
  - Tokens ut: 35,606
  - Kostnad: \$0.0734
  - Kjørt: 2026-04-14 01:41:18.162285+00:00
- 

## Sverm-debrief

---

### Konsensus

- **Arkitektur over verktøyjakt:** En strukturert 3–4 agents pipeline med tydelige, ikke-overlappende roller (Scout, Filter, Synth, Archivist) er avgjørende.
  - **Direkte KB-integrasjon:** Eliminer manuell overføring. Push automatisk til Obsidian via REST API, Git-sync eller filsystem-watcher. E-post skal kun være output, aldri mellomlagring.
  - **Aggressiv signal-støy-filtrering:** 80–95 % av innholdet må kastes. Bruk relevansskåring (>70–85 %), avviksdeteksjon mot eksisterende embeddings, eller konsensus-validering.
  - **Data-suverenitet:** Kjør på egen server. Bruk OpenRouter kun som modell-router, men behold rådata, transkripsjoner og vektor-DB lokalt (SQLite/Qdrant/Chroma).
  - **Faset implementering:** Start med 5 kilder, 3 agenter og en 7–14 dagers kalibreringsperiode (shadow-mode) før fullskaladrift.
- 

### Dissens

- **Admin-grensesnitt:** Noen anbefaler lettvekt dashboards (FastAPI/Next.js/Streamlit) for oversikt. Andre advarer mot dette som en vedlikeholds-felle og foreslår CLI/tmux eller ren filsystem-sync.
- **Orkestreringsrammeverk:** Uenighet om LangGraph / CrewAI vs. n8n vs. ren Python med Redis-køer. Flere prioriterer enkelhet og determinisme over komplekse agent-rammeverk.



- **Modellavhengighet:** Noen vil rute alt via OpenRouter for kvalitet. Andre krever lokal fallback (Ollama/Llama.cpp) for kritiske steg for å redusere kostnad, latens og API-sårbarhet.
- **Leveringsfrekvens:** Daglig digest vs. ukentlige tematiske grafer vs. kun hendelsesdrevne avviksvarsler.

---

## Blindsoner avdekket

- **Juridisk/ToS-risk:** Direkte lagring av skrapet X/YouTube-innhold kan bryte plattformvilkår. Løsning: lagre kun hashes, metadata og lenker, ikke råtekst.
- **Kognitiv overmetning:** Den reelle begrensningen er oppmerksomhet, ikke båndbredde. Systemet må beskytte fokus, ikke bare aggregere data.
- **Manglende ontologi:** Uten 5–7 definerte kjerne-dimensjoner (f.eks. orkestreringsmønstre, feiltoleranse, kostnad) vil svermen produsere ustrukturert støy.
- **API-fragilitet:** X og YouTube-API-er er ustabile. RSS-broer, webhooks og lokal caching er nødvendige for resiliens.

---

## Anbefalinger

1. Deploy en 3-agent pipeline ( Scout → Filter/Synth → Archivist ) med Redis-kø og yt-dlp + lokal Whisper. Kjør i shadow-mode i 10 dager.
2. Koble direkte til Obsidian via obsidian-local-rest-api eller Git-sync. Fjern e-post som mellomledd; bruk raymond@sverm.ai kun for <3 min MP3-summaries til pendling.
3. Definer en streng relevans-terksel ( $\geq 0.75$ ) og cosine-similarity mot din eksisterende vault. Kun avvik >15 % eller nye arkitekturer trigger varsler.
4. Implementer automatisk modell-fallback: OpenRouter for syntese, Ollama ( llama-3.1-8b ) ved API-svikt eller høy latency.
5. Bygg en enkel JSON-basert kildegraf som månedlig foreslår 3 nye X/YouTube-kontoer basert på nettverksnaboer. Godkjenn med én tast.