LE GUIDE COMPAGNON — Créez votre agent IA

Le compagnon de la vidéo « setup de base » : chaque étape dans l'ordre, avec les phrases exactes à dicter à votre IA de code. Vous ne taperez presque aucune commande — c'est justement le principe.

La règle avant tout

Hermes ne sert à quelque chose que si vous lui donnez **un rôle auquel vous avez pensé à l'avance**. Avant de toucher à un serveur, notez deux ou trois choses que vous faites souvent derrière votre écran et que vous déléguerez bien — sans payer quelqu'un 3 000 EUR/mois pour ça. Gardez cette liste sous les yeux : c'est elle qui donne du sens à tout le reste.

Le budget réel

Poste	Coût	Note
VPS Contabo « Storage 10 » (mon choix)	3,60 EUR/mois	Moins de RAM mais 300 Go SSD + 150 Go NVMe — je suis pas pressé, j'ai de la place
VPS Contabo de base	quelques EUR/mois	8 Go de RAM, 4 vCores, 100 Go SSD — ça suffit largement pour démarrer
LLM débutant	ChatGPT Pro : 20 USD/mois	Le plus simple pour commencer — et vous aurez la création d'images
LLM radin (mon choix)	GLM : ~15 USD/mois	GLM 5.3 Flash — carré et fiable. Alternative : DeepSeek à l'API, quelques centimes
Discord	0 EUR	Là où je parle à mon agent, du téléphone ou du canapé

La feuille de route

#	Étape	Où dans la vidéo
1	Louer le VPS (Ubuntu, Europe)	02:23
2	La clé SSH : votre IA obtient les clés du serveur	05:15
3	Hermes installé par votre IA	08:54
4	Le cerveau : brancher le LLM (abonnement ou clé API)	14:23
5	Le premier agent orchestrateur : LA méthode du prompt	19:22
6	Le fallback : le plan B (optionnel)	24:01
7	Discord : donner une voix à l'agent	24:55
8	Bonus : le bot ne répond pas (le fix)	32:02
9	Bonus : lui laisser des vocaux (transcription locale)	33:26

1 · Le terrain : louer le VPS

Un VPS, c'est un petit morceau d'ordinateur dans un datacenter, allumé 24h/24 et connecté à Internet — votre agent y habite comme dans une maison extérieure à chez vous. Contabo, Hetzner, peu importe : **commencez par le moins cher**, l'upgrade est toujours possible. Mon choix : le Storage VPS — pas le plus rapide, mais je ne suis jamais à court de place. Deux réglages non négociables : **Ubuntu** et **l'Europe**.

Après le paiement, vous recevez un e-mail avec deux informations vitales : **l'adresse IP** (l'adresse de la maison) et **le mot de passe root**. Secrets absolus — on ne les montre jamais. Et à retenir : le VPS ne gère rien tout seul. C'est vous — enfin, votre IA — qui le gérez.

2 · La clé SSH : les clés de la maison

Ouvrez votre IA de code — Claude Code, Antigravity, Codex... — et dites-lui de créer une clé SSH. Le principe : la clé est découpée en deux fichiers. Le **privé** reste chez vous, sur votre ordinateur. Le **public** est déposé sur le VPS (il demandera le mot de passe root une fois, pour vérifier que c'est bien vous). Une fois la clé en place : **fermez Contabo, vous n'y toucherez plus jamais**. Votre IA se connecte au serveur à distance, c'est elle qui tape, vous lisez et vous approuvez.

À DICTER À VOTRE IA DE CODE

```
Je viens de louer un VPS. Aide-moi a creer une cle SSH sur mon ordinateur,
depose la cle publique sur le serveur, puis verifie que la connexion fonctionne.
```

(Même consigne avec les accents : votre IA comprendra dans tous les cas. Ensuite, chaque encadré « à dicter » de ce guide se copie-colle tel quel.)

Sous le capot — si vous préférez comprendre, ou taper vous-même, voici tout ce que votre IA exécute à cette étape (et la suivante) :

```
ssh root@VOTRE_IP                          # 1re connexion
apt update && apt full-upgrade -y          # tout mettre a jour

adduser votre-nom                          # votre utilisateur (jamais root au quotidien)
usermod -aG sudo votre-nom

ssh-keygen                                 # la cle, sur VOTRE ordinateur
ssh-copy-id votre-nom@VOTRE_IP             # la moitie publique, deposee sur le VPS

sudo nano /etc/ssh/sshd_config.d/60-hardening.conf
#    PasswordAuthentication no
#    KbdInteractiveAuthentication no
sudo systemctl restart ssh

sudo ufw allow 22,80,443/tcp               # pare-feu : 3 portes seulement
sudo ufw enable

sudo apt install fail2ban -y               # le gardien qui bannit les insistants
sudo fail2ban-client status sshd           # le compteur tourne deja...
```

La règle d'or du filet : après chaque modification SSH, on teste une nouvelle connexion *avant* de fermer la session en cours. Sinon, un jour, vous restez dehors.

3 · Hermes installé par votre IA

Dans une nouvelle conversation, dictez la demande ci-dessous. Le premier agent est un **agent système** : si votre IA parle de Docker, précisez-lui que pour un premier agent, l'installation classique suffit. Docker, c'est pour plus tard — quand vous voudrez plusieurs agents **cloisonnés**, chacun dans sa bulle. Ce cloisonnement n'est pas un détail : sans lui, des agents qui vivent au même endroit finissent par mélanger les données de vos projets. Et ne vous enflammez pas : **un seul agent d'abord**.

À DICTER À VOTRE IA DE CODE

```
J'ai un nouveau VPS et un acces SSH. Je souhaite installer un agent Hermes.
C'est mon premier agent, il n'y a rien pour l'instant. Donne-moi les grandes
etapes, puis installe-le. Installation systeme classique, pas Docker.
```

Sous le capot — grosso modo, ce que votre IA fait pour vous :

```
curl -fsSL https://hermes-agent.nousresearch.com/install.sh | bash
hermes setup                              # l'assistant de configuration
#    variante --portal : un seul compte pour le modele + les outils web
#    la config vit dans ~/.hermes : config.yaml, memories/, skills/
```

4 · L'âme : brancher le LLM

Sans LLM, pas de réflexion. Deux écoles : l'**abonnement** (connexion OAuth, zéro clé API à gérer) ou la **clé API** (on paie à l'usage, chez DeepSeek ou GLM cela se compte en centimes). Si vous débutez sans opinion : ChatGPT Pro, c'est le plus simple, et vous aurez la création d'images en bonus.

Deux conseils de quelqu'un qui a déjà payé pour rien : si vous comptez **beaucoup tester**, ne mettez pas le modèle le plus cher — prenez un modèle « mini » ou « flash ». Et pas besoin d'un génie : moi j'utilise GLM 5.3 Flash, je lui demande d'être **carré, fiable, organisé**, de travailler en tâche de fond sans me poser dix mille questions. Ça lui va très bien.

À DICTER À VOTRE IA DE CODE

```
Mon Hermes doit utiliser mon abonnement ChatGPT en connexion OAuth.
Donne-moi les etapes.
```

À DICTER À VOTRE IA DE CODE

```
Je veux connecter Hermes avec une cle API DeepSeek (ou GLM, ou OpenRouter).
Donne-moi les etapes, je recupere l'adresse et la cle sur leur site.
```

Le fallback (optionnel) : un deuxième LLM pas cher qui prend le relais si le principal tombe — par exemple DeepSeek ou GLM Flash via OpenRouter, en plus de votre abonnement principal. Obligatoire si votre agent doit rester opérationnel en continu ; inutile pour un premier bricolage.

5 · Le premier agent orchestrateur : LA méthode

Étape la plus importante. Ne laissez pas Hermes créer l'agent « à sa manière » en vous posant ses questions de base — c'est le meilleur moyen d'avoir un agent générique. La méthode qui marche : **préparez le prompt avec une autre IA**, puis collez-le à Hermes.

a. Ouvrez votre IA (Antigravity, Claude, ChatGPT...) et dictez :

À DICTER À VOTRE IA DE CODE

```
Je cree mon premier agent Hermes orchestrateur. Pose-moi toutes les
questions pour preparer sa creation.
```

b. Répondez à ce qui vous parle : votre projet, vos thèmes, vos intentions présentes et futures, l'équipe prévue derrière (rédacteurs ? chercheurs ? développeurs ?). **c.** Puis :

À DICTER À VOTRE IA DE CODE

```
Ecris-moi un prompt complet et precis pour creer cet agent.
```

d. Copiez ce prompt et donnez-le à Hermes via votre IA de code — c'est l'ADN de votre orchestrateur. Ce qu'il doit fixer, avec mes exemples :

À fixer dans le prompt	Mon exemple
Un nom	Le mien s'appelle Hydra — parce qu'il est multitête. C'est presque une tradition
Le domaine	L'éducation — pas le sport, pas la crypto : VOTRE thème
Le style de réponse	Réponses concises, pas de blabla, tableaux dès que possible (sinon on déroule deux minutes pour comparer deux choses)
Les intentions	Ce que vous lui confierez maintenant, et ce qui viendra plus tard

6 · La voix : Discord

Pourquoi Discord ? J'ai testé Telegram, ça ne m'a pas plu ; Discord m'a convaincu dès le premier soir — téléphone, canapé, dictée vocale. C'est un choix, pas une obligation. Une règle par contre : **un serveur Discord dédié**, créé uniquement pour votre agent. Le parcours, portail des développeurs ouvert :

#	Où	Quoi
1	discord.com/developers/applications	Nouvelle application
2	Onglet Bot	Réinitialiser le token (c'est le fonctionnement normal au démarrage), le copier, le coller dans votre IA
3	Bot → Privileged Gateway Intents	Activer les TROIS interrupteurs, comme dans la vidéo
4	Permissions du bot	Administrateur + Save — le serveur ne sert qu'à l'agent, il n'y a rien à casser
5	Dans Discord	Créer votre serveur dédié + un salon #general, clic droit sur le serveur → Copier l'identifiant du serveur (mode développeur activé au préalable)
6	Retour dans votre IA	Token + ID serveur : elle configure Hermes et vous génère le lien d'invitation
7	Le lien d'invitation	Ouvrez, validez — le bot apparaît en ligne dans #general

À DICTER À VOTRE IA DE CODE

```
Mon premier agent est cree, il s'appelle X. Je veux echanger avec lui via
Discord. Voici le token du bot et l'ID de mon serveur : [a coller].
Configure tout et donne-moi le lien d'invitation.
```

Sous le capot — les deux commandes Hermes qui pilotent la passerelle Discord :

```
hermes gateway setup       # assistant guide : Discord, token, salon par defaut
hermes gateway restart     # a chaque changement de configuration
```

Le moment que vous n'oublierez pas : le bot qui passe « en ligne » dans votre salon.

Bonus 1 · Le bot ne répond pas (32:02)

Il est en ligne mais il ignore vos messages ? Pas de panique : il n'écoute simplement pas le salon — ça arrive après certains réglages. Un message, et c'est réglé :

À DICTER À VOTRE IA DE CODE

```
Mon agent n'ecoute pas mes messages sur Discord. Mets ça en place.
```

Bonus 2 · Lui laisser des vocaux (33:26)

Mon usage préféré : une idée vous traverse dans le métro, vous maintenez le micro, vous parlez — à votre arrivée, c'est fait. La transcription tourne **en local sur le VPS** (Faster Whisper, modèle small/fr) : zéro clé API, rien qui parte chez un tiers. Le vocal en retour, live, ce n'est pas supporté — et pas mon délire non plus.

À DICTER À VOTRE IA DE CODE

```
Comment mettre en place les vocaux dans Discord entre nous ?
Je veux lui laisser des messages vocaux, transcrits en local.
```

Et après ?

Vous êtes au niveau zéro : un agent qui attend, Hermes fonctionnel sur le VPS, un cerveau branché, Discord à l'écoute. La suite, elle ne s'installe pas — elle **pousse** : plus vous échangez avec lui, plus il comprend vos habitudes et se les approprie. Skills, second cerveau, mémoire fine : c'est la prochaine vidéo, avec mon setup complet (le lien « setup de base » est en description).

Les pièges vus dans la vidéo

Piège	Parade
Créer un agent sans intention précise	Notez d'abord ce que vous déléguez — sinon vous conclurez que ça ne sert à rien
Laisser Hermes improviser la création de l'orchestrateur	Préparez le prompt avec une autre IA (étape 5), puis collez-le
S'enflammer et créer cinq agents le premier soir	Un seul d'abord. Les autres, plus tard, chacun dans son Docker — sinon les données se mélangent
Le modèle le plus cher pour bricoler	Mini ou flash pour tester ; le gros modèle pour la production
Le bot en ligne mais muet	Il n'écoute pas le salon — bonus 1, un prompt suffit
Un token ou une IP montrés à l'écran	Secret absolu : si ça fuit, on révoque immédiatement et on régénère

Les liens qui servent

Quoi	Lien
Contabo (le VPS de la vidéo)	contabo.com
Hetzner (alternative)	hetzner.com
Hermes Agent — documentation	hermes-agent.nousresearch.com/docs
Hermes Agent — code source	github.com/NousResearch/hermes-agent
Discord — portail développeurs	discord.com/developers/applications
OpenAI (ChatGPT — débutant)	platform.openai.com
OpenRouter (fallback, 1 clé pour tout)	openrouter.ai
DeepSeek (API pas chère)	platform.deepseek.com
GLM / Z.ai	docs.z.ai
Anthropic (Claude)	console.anthropic.com

Checklist de sortie (« mon agent est prêt »)

[ ] VPS commandé : Ubuntu, Europe, plan petit mais évolutif
[ ] Clé SSH en place, connexion testée par votre IA
[ ] Contabo fermé — vous ne toucherez plus au serveur à la main
[ ] Hermes installé en agent système (pas Docker, pour l'instant)
[ ] LLM branché : abonnement OAuth ou clé API — modèle adapté au test
[ ] Orchestrateur créé depuis un prompt préparé (nom, domaine, style)
[ ] Serveur Discord dédié + ID transmis à votre IA
[ ] Les 3 Intents actifs, permissions Admin, token transmis
[ ] Le bot répond dans #general (sinon : bonus 1)
[ ] Fallback en place, si l'agent doit tourner en continu (optionnel)
[ ] Vocaux : Faster Whisper local opérationnel (optionnel)
[ ] La première vraie mission confiée — celle de votre liste du début

Guide généré pour accompagner la vidéo « setup de base » — Mintavocado. Aucun secret inclus : les IP, tokens et clés restent les vôtres.