Fiche Pratique N°88 : Faites tourner une IA sur votre propre ordinateur V1.0

15 nov, 2026 · @Gabriel

Objectif : *installer un assistant d'intelligence artificielle qui fonctionne entièrement sur votre ordinateur, sans compte et sans envoyer vos questions ni vos documents sur Internet, puis choisir un modèle adapté à votre machine.*

- ***Public visé :*** *intermédiaire*
- ***Temps estimé :*** *30 minutes (dont le téléchargement d'un modèle de quelques Go)*
- ***Difficulté :*** ★★☆☆☆ *(Facile)*
- ***Prérequis :*** *un ordinateur de moins de 8 ans avec au moins 8 Go de mémoire vive (16 Go conseillés) et 10 Go d'espace libre ; Linux Mint conseillé (fiches N°13a et N°13b)*
- ***Fiches associées :*** *N°53 (utiliser l'IA sans livrer sa vie privée), N°78 (installer ses logiciels sous Linux Mint), N°74 (traduire sans Google Traduction), N°26 (chiffrer ses fichiers), N°77 (faire durer ses appareils), N°86 (vérifier une information)*

PARTIE 1 : LES BASES (pour débutants)

1. Le problème : vos questions partent chez quelqu'un d'autre

Quand vous utilisez ChatGPT, Gemini, Copilot ou un autre assistant en ligne, chaque question, chaque texte collé et chaque document envoyé part sur les serveurs de l'entreprise, souvent aux États-Unis. Selon les réglages, ils peuvent être conservés, relus ou servir à entraîner de futurs modèles. Pour un courrier personnel, un dossier médical ou un document de travail, c'est un vrai problème.

Situation	Conséquence
Vous faites résumer un compte rendu médical par un assistant en ligne	Vos données de santé sont envoyées à une entreprise étrangère
Vous collez un document de votre association pour le reformuler	Son contenu quitte votre ordinateur
Vous utilisez l'assistant sans compte payant	Vos échanges peuvent servir à entraîner le modèle, selon les conditions
Vous êtes hors connexion (train, montagne)	L'assistant en ligne ne fonctionne pas
Le service change ses prix ou ses conditions	Vous n'avez aucun contrôle

Une image pour comprendre : un assistant en ligne, c'est un écrivain public installé dans un bureau lointain : pour qu'il vous aide, vous devez lui envoyer vos lettres. Une IA locale, c'est un dictionnaire et un correcteur rangés dans votre bibliothèque : ils travaillent chez vous, et rien ne sort de la maison.

Le risque : confier des informations personnelles ou confidentielles à une entreprise que vous ne contrôlez pas.

Le bénéfice : un assistant privé, gratuit, qui fonctionne même sans Internet.

gabriel.ananda@mailo.com
CC BY SA

Fiche Pratique N°88 : Faites tourner une IA sur votre propre ordinateur V1.0

2. Comprendre l'IA locale

Un assistant d'IA est fait de deux éléments : un **logiciel** (la fenêtre où vous écrivez) et un **modèle** (le « cerveau », un gros fichier de plusieurs Go). En local, les deux sont sur votre ordinateur.

Taille du modèle	Mémoire vive nécessaire	Ce qu'on peut en attendre
1 à 4 milliards de paramètres (« 3B »)	8 Go	Reformuler, corriger, résumer un texte court ; rapide
7 à 8 milliards (« 8B »)	16 Go	Bon assistant du quotidien : rédaction, résumés, questions générales
13 à 14 milliards (« 14B »)	16 à 32 Go	Plus précis, plus lent sans carte graphique
70 milliards et plus	64 Go ou plus	Hors de portée d'un ordinateur ordinaire

À savoir : un modèle local est moins puissant que les grands assistants en ligne, et il peut se tromper avec aplomb. Il ne sait rien de l'actualité récente. Vérifiez toujours ce qui compte (fiche N°86).

Avant de commencer

- ☐ Mon ordinateur a au moins **8 Go de mémoire vive** (sous Linux Mint : menu › **Informations système**)
- ☐ J'ai au moins **10 Go d'espace libre**
- ☐ Mon système est **à jour** (fiche N°78)
- ☐ J'ai une **bonne connexion** pour télécharger le modèle (une seule fois)

3. Pas à pas : installer Jan sous Linux Mint

Jan est un logiciel libre (licence Apache 2.0) qui ressemble à ChatGPT, mais fonctionne sur votre ordinateur. Il existe aussi pour Windows et macOS.

1. Ouvrez la **Logithèque**, cherchez **Jan** et installez la version Flatpak (identifiant `ai.jan.Jan`).
    - Vous devez voir : Jan dans le menu des applications.
2. Lancez Jan, puis ouvrez la liste des modèles (**Hub** ou **Modèles**, selon la version). Choisissez un modèle adapté à votre mémoire vive (section 2) : par exemple un **Ministral 3B ou 8B** (Mistral, entreprise française, licence Apache 2.0), un **Qwen** ou un **Gemma** de taille équivalente.
    - Vous devez voir : une barre de téléchargement ; le modèle pèse de 2 à 6 Go environ.
3. Une fois le modèle téléchargé, ouvrez une nouvelle conversation et sélectionnez-le.
    - Vous devez voir : le nom du modèle en haut de la conversation.
4. **Testez sans Internet** : coupez le Wi-Fi, puis posez une question (« Reformule ce paragraphe en plus simple : … »).
    - Vous devez voir : la réponse s'écrire, preuve que tout se passe sur votre ordinateur.

gabriel.ananda@mailo.com
CC BY SA

5. Vérifiez qu'aucun service en ligne n'est utilisé : dans les **Paramètres** de Jan, ne renseignez **aucune clé d'API** de fournisseur extérieur (OpenAI, Anthropic, Mistral…), et désactivez le partage de statistiques s'il est proposé.
    - Vous devez voir : seuls des modèles locaux dans la liste de sélection.

4. À quoi s'attendre et dépannage

Symptôme	Cause possible	Action
La réponse s'écrit très lentement	Modèle trop gros pour votre machine	Prenez un modèle plus petit (3B ou 4B)
L'ordinateur rame ou le ventilateur s'emballe	Le processeur travaille à fond	Normal pendant la réponse ; fermez les autres logiciels
Erreur au chargement du modèle	Mémoire vive insuffisante	Choisissez un modèle plus petit ou une version « quantifiée » (Q4)
Les réponses sont fausses ou inventées	Limite des petits modèles	Vérifiez ; formulez des questions plus précises ; essayez un modèle plus grand
Réponses en anglais	Le modèle répond dans la langue qu'il préfère	Commencez par « Réponds en français. »

PARTIE 2 : ALLER PLUS LOIN (pour intermédiaires et avancés)

5. Discuter avec vos documents (GPT4All)

GPT4All (licence MIT) propose une fonction **LocalDocs** : vous lui indiquez un dossier de documents (PDF, textes, notes), il les indexe, et vous pouvez poser des questions dessus, sans que rien ne sorte de votre ordinateur.

1. Téléchargez GPT4All pour Linux depuis son site officiel, puis installez-le.
    - Vous devez voir : GPT4All dans le menu.
2. Téléchargez un modèle depuis l'onglet des modèles, puis ouvrez **LocalDocs** et ajoutez un dossier.
    - Vous devez voir : l'indexation progresser, puis la collection prête.
3. Dans une conversation, activez la collection et posez une question sur vos documents.
    - Vous devez voir : une réponse qui cite les passages utilisés.

Limites : cela marche bien pour quelques dizaines de documents ; au-delà, les réponses deviennent moins fiables. Vérifiez toujours dans le document d'origine.

gabriel.ananda@mailo.com
CC BY SA

Fiche Pratique N°88 : Faites tourner une IA sur votre propre ordinateur V1.0

6. Pour les utilisateurs avancés : Ollama

Ollama (licence MIT) fait tourner des modèles en ligne de commande et sert de moteur à d'autres logiciels. Installation sous Linux :

```
curl -fsSL https://ollama.com/install.sh | sh
ollama run ministral-3:8b
```

La première commande installe Ollama comme service ; la seconde télécharge le modèle puis ouvre une conversation (tapez `/bye` pour quitter). Vérifiez le nom exact du modèle dans la bibliothèque d'Ollama. Autres commandes utiles : `ollama list` (modèles installés) et `ollama rm nom-du-modèle` (supprimer un modèle).

7. Comparatif

Logiciel	Licence	Points forts	Limites
Jan	Apache 2.0	Interface simple, proche de ChatGPT ; Flatpak sous Linux	Propose aussi des modèles en ligne : à laisser désactivés
GPT4All	MIT	LocalDocs pour interroger vos documents	Interface moins moderne
Ollama	MIT	Léger, moteur pour d'autres outils, ligne de commande	Pas d'interface graphique seule
LM Studio	Gratuit mais propriétaire	Très complet	Code non libre : moins conforme à l'esprit du guide

Les modèles : préférez les modèles à licence ouverte (Apache 2.0, MIT), comme la famille **Ministral 3** de Mistral (3B, 8B et 14B, publiée en décembre 2025 et pensée pour les ordinateurs et téléphones). Certains modèles très connus ont une licence plus restrictive : lisez-la si vous les utilisez pour une association ou un usage professionnel.

À savoir avant de se lancer

Crainte fréquente	La réalité
« Il faut une carte graphique puissante. »	Non : un modèle de 3 à 8 milliards de paramètres tourne sur le processeur, plus lentement.
« C'est aussi bon que ChatGPT. »	Non, c'est moins puissant ; mais suffisant pour reformuler, résumer, traduire et rédiger.
« Mes données partent quand même quelque part. »	Pas avec un modèle local : testez-le en coupant Internet.

gabriel.ananda@mailo.com
CC BY SA

Fiche Pratique N°88 : Faites tourner une IA sur votre propre ordinateur V1.0

Crainte fréquente	La réalité
« Ça va remplir mon disque. »	Un modèle pèse de 2 à 10 Go ; supprimez ceux qui ne servent pas.

Challenge 7 jours

- **Jour 1 :** vérifiez la mémoire vive et l'espace libre de votre ordinateur.
- **Jour 2 :** installez Jan et téléchargez un petit modèle.
- **Jour 3 :** demandez-lui de reformuler un courrier, Wi-Fi coupé.
- **Jour 4 :** comparez deux modèles sur la même question.
- **Jour 5 :** faites-lui résumer un long article.
- **Jour 6 :** essayez GPT4All et LocalDocs sur un dossier de notes.
- **Jour 7 :** listez ce que vous ne confierez plus à un assistant en ligne.

Vous allez constater : pour beaucoup de tâches du quotidien, une IA locale suffit, et vos textes restent chez vous.

En résumé (ce que vous gagnez)

Action	Gagné	Renvoi
Installer Jan	Un assistant privé, sans compte	Section 3
Choisir un modèle adapté	Des réponses à une vitesse raisonnable	Section 2
Interroger vos documents en local	Vos fichiers ne quittent pas l'ordinateur	Section 5
Préférer les modèles à licence ouverte	Un usage libre, y compris associatif	Section 7

À retenir absolument

- **Une IA locale fonctionne sans Internet** : vos questions et documents restent chez vous.
- **8 Go de mémoire vive** pour un petit modèle, **16 Go** pour un bon assistant du quotidien.
- **Pas de clé d'API de fournisseur extérieur** dans Jan si vous voulez rester 100 % local.
- **Vérifiez les réponses** : un petit modèle peut se tromper avec assurance.

Test final :

1. Je connais la mémoire vive de mon ordinateur.
2. J'ai installé un logiciel d'IA locale et un modèle adapté.
3. J'ai obtenu une réponse avec Internet coupé.
4. Je sais quel type de document je garde pour l'IA locale.
5. Je sais supprimer un modèle qui ne sert plus. ✅

gabriel.ananda@mailo.com
CC BY SA

Fiche Pratique N°88 : Faites tourner une IA sur votre propre ordinateur V1.0

Références

- GitHub – Jan, assistant d'IA local (licence Apache 2.0, Flathub ai.jan.Jan)
- Dev.to – GPT4All Review 2026 (licence MIT, LocalDocs, mémoire nécessaire)
- eWEEK – Mistral 3 : Mistral Large 3 et Ministral 3 sous licence Apache 2.0 (décembre 2025)
- Ollama – Bibliothèque de modèles : ministral-3
- Mustafa.net – How to install Ollama and run local LLMs on Linux (octobre 2026)

gabriel.ananda@mailo.com
CC BY SA