FORGE PC ELITE

# Guide de prise en main
# Pack IA Créatif

Assistant IA privé, génération d'images et de vidéos, transcription audio : tout est installé, configuré et prêt à créer dès le premier démarrage.

Machine montée, testée et configurée à Labège (31) · forgepcelite.fr

# 0 Votre fiche machine

Cette fiche est remplie par notre atelier avant l'expédition. Conservez-la : elle vous sera utile si vous nous contactez.

### Configuration installée

**Modèle de PC :** .....

**N° de commande :** .....

**Carte graphique / VRAM :** .....

**Date de configuration :** .....

**Adresse de l'interface de chat :** .....

**Modèle de chat « rapide » :** .....

**Modèle de chat « polyvalent » :** .....

**Modèle de chat « réflexion / code » :** .....

**Modèle d'images installé :** .....

**Modèle vidéo installé (si compatible) :** .....

**Outil de transcription :** .....

**Configuré par :** .....

# 1 Ce que vous avez entre les mains

Votre PC fait tourner une **intelligence artificielle directement chez vous**, sur sa propre carte graphique. Concrètement :

- **Vos données restent sur votre machine.** Vos questions, vos documents et vos créations ne sont envoyés à aucun serveur extérieur.
- **Aucun abonnement.** Vous pouvez l'utiliser autant que vous voulez, sans quota ni frais mensuels.
- **Elle fonctionne hors ligne.** Une connexion internet n'est utile que pour télécharger de nouveaux modèles ou des mises à jour.

> **Un modèle, c'est quoi ?**
> Un modèle est le « cerveau » de l'IA : un gros fichier (de quelques Go à plusieurs dizaines de Go) qui contient ce qu'elle a appris. Plus un modèle est gros, plus il est capable, mais plus il demande de mémoire sur la carte graphique (la VRAM). Nous avons choisi pour vous des modèles adaptés à votre carte.

# 2 Discuter avec votre IA

## Les logiciels installés

| Logiciel | Son rôle |
|---|---|
| **Ollama** | Le moteur qui fait tourner les modèles en arrière-plan. Il démarre tout seul avec Windows : son icône (une tête de lama) apparaît près de l'horloge. |
| **Open WebUI** | L'interface de discussion, dans votre navigateur, qui ressemble aux assistants IA en ligne. C'est elle que vous utiliserez au quotidien. |
| **LM Studio** (selon configuration) | Une application de bureau tout-en-un, pratique pour tester et télécharger facilement de nouveaux modèles. |

## Premier démarrage

1. Double-cliquez sur le raccourci **« Assistant IA »** du bureau. Votre navigateur s'ouvre sur l'interface (adresse notée sur votre fiche machine).
2. Lors de la toute première ouverture, l'interface vous demande de **créer un compte**. Ce compte est uniquement stocké sur votre PC : il ne s'agit pas d'une inscription en ligne. Notez bien votre mot de passe.
3. En haut de l'écran, **choisissez un modèle** dans la liste déroulante.
4. Écrivez votre question en bas, puis appuyez sur **Entrée**. La réponse s'affiche au fur et à mesure.

> **Le premier message est un peu plus long**
>
> Au premier message, le modèle se charge dans la mémoire de la carte graphique, ce qui prend quelques secondes. Les réponses suivantes sont beaucoup plus rapides.

## Quel modèle choisir ?

| Type de modèle | Idéal pour |
|---|---|
| **Rapide** | Questions simples, reformulation, traduction, résumés courts. Répond quasi instantanément. |
| **Polyvalent** | L'usage de tous les jours : rédaction d'e-mails et de courriers, explications, idées, analyse de documents. |
| **Réflexion / code** | Problèmes complexes, raisonnement en plusieurs étapes, programmation, formules Excel. Plus lent, mais plus précis. |

Les noms exacts des modèles installés sur votre machine figurent sur votre fiche machine (page 2).

# 3 Bien utiliser votre IA

## Obtenir de bonnes réponses

- **Donnez du contexte** : « Rédige un e-mail poli à un client qui paie en retard, ton professionnel, 5 lignes maximum » fonctionne bien mieux que « écris un e-mail ».
- **Précisez la forme attendue** : liste à puces, tableau, nombre de mots, niveau de langue.
- **Affinez la réponse** : « plus court », « plus formel », « donne trois autres idées »… L'IA garde en mémoire le fil de la conversation.
- **Une conversation par sujet** : ouvrez une nouvelle discussion (bouton « Nouvelle conversation ») quand vous changez de sujet. Les réponses seront plus pertinentes.

## Travailler sur vos documents

Dans Open WebUI, cliquez sur le **+** à gauche de la zone de saisie pour joindre un fichier (PDF, Word, texte…), puis posez votre question : « Résume ce contrat », « Liste les dates importantes de ce document »… Le fichier reste sur votre machine.

> **Les limites à connaître**
>
> - Une IA peut **se tromper avec assurance** (on parle d'« hallucination »). Vérifiez toujours les chiffres, dates, citations et informations juridiques, médicales ou financières importantes.
> - Elle **n'a pas accès à internet** : elle ne connaît pas l'actualité et ses connaissances s'arrêtent à la date d'entraînement du modèle.
> - Les modèles locaux sont un peu moins puissants que les plus grands services en ligne, mais ils sont privés, gratuits à l'usage et illimités.

# 4 Ajouter des modèles et entretenir votre installation

## Télécharger un nouveau modèle

**Méthode simple (LM Studio)** : ouvrez l'onglet de recherche, tapez le nom d'un modèle (par exemple « Mistral » ou « Qwen »), puis cliquez sur Télécharger. LM Studio indique si le modèle est compatible avec votre carte.

**Avec Ollama** : ouvrez le Terminal Windows et tapez par exemple :

```
ollama pull mistral-small
```

Le modèle apparaît ensuite automatiquement dans la liste d'Open WebUI. Pour voir les modèles installés : `ollama list`. Pour en supprimer un : `ollama rm nom-du-modele`.

## Quelle taille de modèle pour votre carte ?

| Mémoire de la carte (VRAM) | Taille de modèle confortable* |
|---|---|
| 16 Go | Jusqu'à environ 14 milliards de paramètres (« 14B ») |
| 24 à 32 Go | Jusqu'à environ 32 milliards de paramètres (« 32B ») |
| 48 à 64 Go | Jusqu'à environ 70 milliards de paramètres (« 70B ») |
| 96 Go et plus | Très gros modèles, plusieurs utilisateurs en même temps |

* Pour des versions « compressées » (quantifiées, souvent notées Q4 ou Q5), les plus courantes. Un modèle trop gros pour la carte fonctionne quand même, mais devient beaucoup plus lent.

## Mises à jour

- Ollama et LM Studio se mettent à jour d'eux-mêmes ou vous proposent la mise à jour.
- Pensez à mettre à jour le **pilote NVIDIA** quelques fois par an via l'application NVIDIA.
- De nouveaux modèles sortent très régulièrement : n'hésitez pas à en essayer, ils sont souvent meilleurs à taille égale.

## En cas de souci

| Symptôme | Solution |
|---|---|
| Les réponses sont très lentes | Le modèle est probablement trop gros pour votre carte : choisissez-en un plus petit. Fermez aussi les jeux ou logiciels 3D ouverts. |
| L'interface affiche une erreur de connexion | Vérifiez que l'icône Ollama est bien présente près de l'horloge. Sinon, lancez « Ollama » depuis le menu Démarrer, ou redémarrez le PC. |
| Aucun modèle dans la liste | Ollama n'est pas démarré, ou aucun modèle n'est installé : voir ci-dessus. |
| Mot de passe de l'interface oublié | Contactez-nous, nous vous guiderons pour le réinitialiser. |

# 5 Générer des images avec ComfyUI

**ComfyUI** est le logiciel de référence pour créer des images (et des vidéos) par IA. Son interface fonctionne avec des « blocs » reliés entre eux, appelés **nœuds**, qui forment un **workflow**. Pas d'inquiétude : nous avons préparé des workflows prêts à l'emploi, il suffit de modifier le texte.

## Créer votre première image

1. Lancez **ComfyUI** depuis le raccourci du bureau. Patientez quelques secondes le temps du démarrage.
2. Ouvrez le menu **Workflow** (ou le panneau « Workflows » à gauche) et choisissez un workflow préparé, par exemple « Image – texte vers image ».
3. Dans le bloc **Prompt positif**, décrivez l'image souhaitée, de préférence en anglais : `a red sports car on a mountain road at sunset, photorealistic`.
4. Cliquez sur **Run** (ou « Queue »). L'image apparaît dans le bloc d'aperçu au bout de quelques secondes.

## Les réglages utiles

| Réglage | À quoi il sert |
|---|---|
| **Prompt positif** | Ce que vous voulez voir : sujet, style, éclairage, cadrage (« portrait », « vue aérienne », « style aquarelle »…). |
| **Prompt négatif** | Ce que vous ne voulez pas voir (« flou », « texte », « déformé »…). Certains modèles récents n'en ont pas besoin. |
| **Largeur / hauteur** | La taille de l'image. Restez proche de 1024 × 1024 pour commencer, ou 1344 × 768 pour un format paysage. |
| **Seed** | Le « numéro » de l'image. Même seed et même prompt donnent la même image ; changez-le pour obtenir des variantes. |
| **Steps** | Le nombre d'étapes de calcul : plus il y en a, plus l'image est fignolée, mais plus c'est long. Gardez la valeur préréglée. |
| **Batch size** | Le nombre d'images générées d'un coup. |

Vos images sont enregistrées automatiquement dans le dossier **output** de ComfyUI (un raccourci « Mes images IA » est placé sur le bureau). Une astuce : faites glisser une image générée dans ComfyUI pour retrouver exactement le workflow et les réglages qui l'ont créée.

## La vidéo (selon votre carte)

Si votre carte graphique le permet, un workflow « Vidéo » est également fourni. Il fonctionne de la même manière : un texte (ou une image de départ) génère quelques secondes de vidéo. Comptez de une à plusieurs minutes de calcul par clip selon la durée et la résolution.

**Message « out of memory » ?**

La carte graphique manque de mémoire. Réduisez la taille de l'image, le nombre d'images (batch size) ou, pour la vidéo, la durée et la résolution. Fermez aussi les autres logiciels gourmands, et l'interface de chat si un modèle y est chargé.

## 6 Transcrire l'audio avec Whisper

**Whisper** transforme la parole en texte avec une excellente précision, en français comme dans de nombreuses autres langues. Idéal pour les réunions, interviews, notes vocales, cours ou sous-titres de vidéos.

1. Ouvrez l'outil de transcription depuis le raccourci **« Transcription »** du bureau.
2. Déposez votre fichier audio ou vidéo (MP3, WAV, M4A, MP4…).
3. Choisissez la langue (ou laissez la détection automatique), puis lancez la transcription.
4. Récupérez le texte en **.txt**, ou en **.srt** pour les sous-titres d'une vidéo.

Grâce à la carte graphique, une heure d'audio se transcrit généralement en quelques minutes. Vous pouvez ensuite coller le texte dans l'assistant IA pour en faire un résumé ou un compte rendu.

### Utiliser les images générées de façon responsable

- Ne créez pas d'images de personnes réelles sans leur accord, et ne faites pas passer une image générée pour une vraie photo.
- Évitez de reproduire des personnages, logos ou œuvres protégés par le droit d'auteur.
- Pour un usage commercial (publicité, produits vendus), vérifiez la licence du modèle d'images utilisé.

## 7 Licences et assistance

Les logiciels installés sont gratuits. Les modèles d'IA sont publiés par leurs éditeurs sous leurs propres licences : la plupart autorisent un usage personnel et professionnel, mais certains limitent l'usage commercial. En cas d'usage professionnel, vérifiez la licence du modèle concerné (indiquée sur sa page de téléchargement).

**Besoin d'aide ?**
Forge PC Elite, monteur de PC gaming, workstations et PC IA à Labège (31).
Rendez-vous sur **forgepcelite.fr**, rubrique contact, en indiquant votre numéro de commande.