

Librechat

- [Procédure de mise en production sur Librechat](#)
- [Mettre à jour Librechat](#)
- [Mettre à jour les modèles dans Librechat](#)

Procédure de mise en production sur Librechat

1. Phase de test sur PreLibrechat.

Tester les modifications qui seront mises sur Librechat (SP26) au préalable.

- Modifications sur le fichier de configuration de prelibrechat (librechat.yaml)
- Modifications dans la stack (docker-compose) de prelibrechat sur SD11

2. Présenter les changements (liste des documents), pour validation, au responsable des études.

La mise en production ne peut pas commencer sans l'accord du responsable des études.

3. Définir la date pour la MEP et s'assurer que plusieurs personnes compétentes sont disponibles à la date définie.

Procéder à la mise en production en matiné idéalement.

4. Prévenir les usagers de Librechat que l'application sera indisponible

5. Se préparer à un retour en arrière en cas de problème durant la mise en production

- Faire une copie des fichiers, avant modification, pour pouvoir facilement rétablir la version précédente

(- Éventuellement faire un snapshot)

6. Procéder aux modifications nécessaires et tester à la fin si tout est fonctionnel

En cas de soucis, il faut s'assurer que l'application soit fonctionnelle :

- tenter de résoudre le problème
- contacter les autres personnes qui pourraient aider à résoudre le problème
- revenir en arrière en restituant la version précédente (voir le point numéro 5)

Mettre à jour Librechat

Librechat repose sur une stack Docker.

Le changelog se trouve sur <https://github.com/danny-avila/LibreChat/releases>

Attention : ne prendre que les versions finales. Eviter les beta ou release candidate (rc).

1. Aller sur portainer et choisir le serveur SD11 (pour la préproduction) ou SP26 (pour la production)

2. Lister les stacks et choisir librechat

<input type="checkbox"/> Name ↓↑ Filter ▾	Type ↓↑	Control	Created ↓↑	Updated ↓↑	Ownership ↓↑
<input type="checkbox"/> dify	Compose	Limited ⚠	2024-11-29 16:33:05	-	🔗 administrators
<input type="checkbox"/> ia-app	Compose	Total	2024-11-29 16:04:42 by admin	2024-12-02 15:08:33 by admin	🔗 administrators
<input type="checkbox"/> immich	Compose	Total	2025-01-06 08:38:58 by admin	2025-02-18 09:49:39 by admin	🔗 administrators
<input type="checkbox"/> librechat	Compose	Total	2024-12-02 07:52:34 by admin	2025-01-14 09:22:17 by admin	🔗 administrators
<input type="checkbox"/> nodered-ia	Compose	Total	2024-11-29 17:31:46 by admin	2024-12-02 09:17:39 by admin	🔗 administrators

Items per page 100 ▾

3. Cliquer sur l'onglet Editor et modifier le numéro de version

☰ Stack

✎ Editor

This stack will be deployed using `docker compose`.

You can get more information about Compose file format in the [official documentation](#).

🔗 Define or paste the content of your docker compose file here

```
1
2
3 # Do not edit this file directly. Use a 'docker-compose.override.yaml' file if you can.
4 # Refer to `docker-compose.override.yaml.example` for some sample configurations.
5
6 services:
7   api:
8     container_name: LibreChat
9     ports:
10      - 3080:3080
11     depends_on:
12      - mongodb
13      - rag_api
14     image: ghcr.io/danny-avila/librechat:v0.7.6
15     restart: always
16     user: root
17     labels:
18      - traefik.http.routers.librechat.rule=Host(`librechat.vincennes.fr`)
```

4. Cliquer sur déployer

5. Vérifier le bon fonctionnement et corriger les éventuelles erreurs

Les erreurs sont essentiellement dues à des évolutions du fichier de configuration ou des autres services précisés dans le docker-compose.

Mettre à jour les modèles dans Librechat

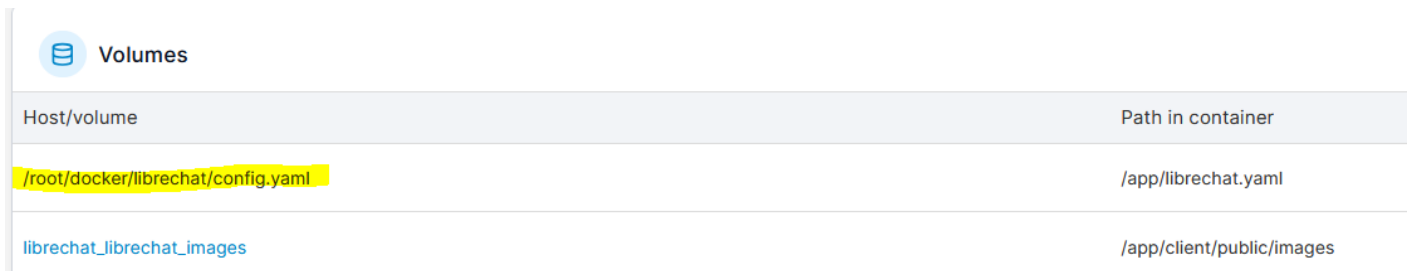
La configuration de Librechat se fait à deux niveaux :

- Dans le fichier de configuration de Librechat
- Dans la stack (docker-compose) de Librechat

Fichier de configuration de Librechat

1. Identifier la localisation du fichier

Aller sur le container Librechat puis sur la partie Volumes



The screenshot shows the Docker interface for the 'librechat' container. Under the 'Volumes' section, there is a table with two columns: 'Host/volume' and 'Path in container'. The first row shows the host path '/root/docker/librechat/config.yaml' highlighted in yellow, which maps to the container path '/app/librechat.yaml'. The second row shows the 'librechat_librechat_images' volume mapped to '/app/client/public/images'.

Host/volume	Path in container
/root/docker/librechat/config.yaml	/app/librechat.yaml
librechat_librechat_images	/app/client/public/images

Le fichier peut être édité

- soit dans le container (via /app/librechat/yaml)
- soit en SSH sur le serveur (via /root/docker/librechat/config.yaml dans le cas présent)

2. Editer le fichier suivant la méthode choisie

Ci-dessous un exemple de configuration pour les modèles Ollama :

```
- name: "Ollama"
```

```
apiKey: "ollama"
# use 'host.docker.internal' instead of localhost if running LibreChat in a docker
container
baseURL: "http://sp26.vincennes.fr:11434/v1/"
models:
  default: [
    "llama3.2-vision:11b", "llama3.1:8b"
  ]
# fetching list of models is supported but the `name` field must start
# with `ollama` (case-insensitive), as it does in this example.
fetch: false
titleConvo: true
titleModel: "current_model"
summarize: false
summaryModel: "current_model"
forcePrompt: false
modelDisplayLabel: "Ollama"
```

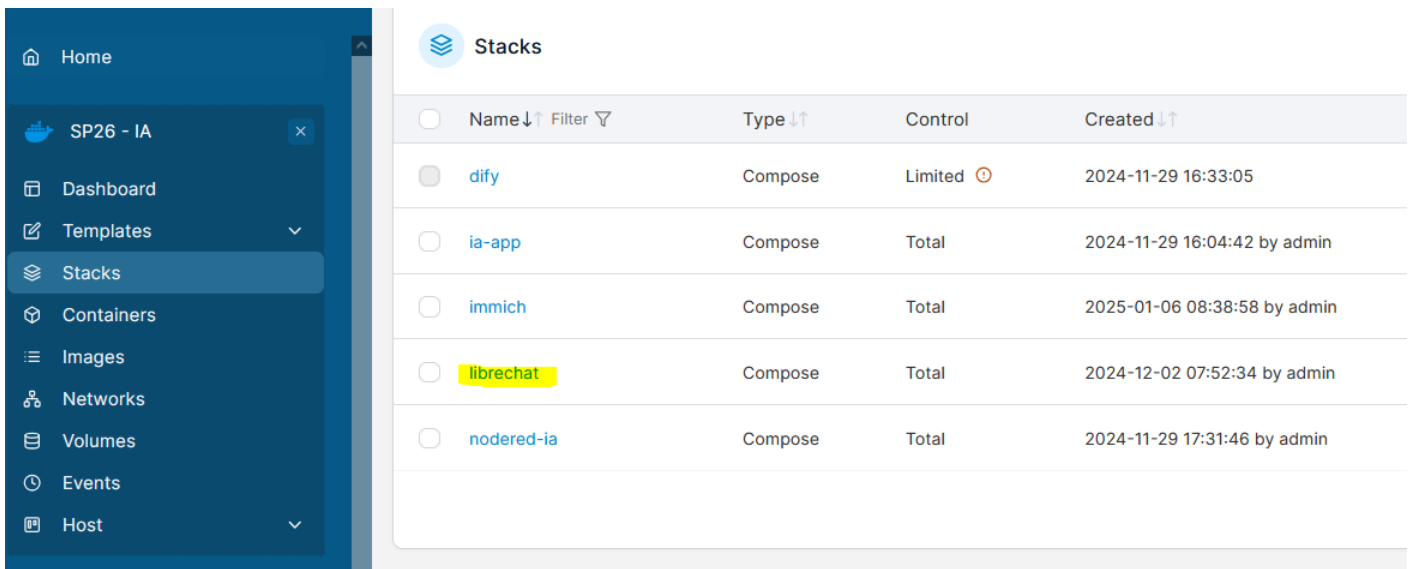
La documentation de Librechat précise la signification des différents champs :

https://www.librechat.ai/docs/configuration/librechat_yaml/object_structure/custom_endpoint

Le numéro de version du fichier config doit être changé

Fichier docker-compose de Librechat

1. Dans stacks, sélectionner Librechat



2. Cliquer sur Editor

3. Modifier la variable d'environnement ENDPOINTS

```
33     - DEBUG_LOGGING=true
34     - DEBUG_CONSOLE=true
35     - ENDPOINTS=ollama
36
37     - PROXY=
```

Elle doit correspondre au champ "name" de config.yaml

4. Embeddings

Dans le cas où le modèle d'embedding du RAG doit être modifié, les options se trouvent sur la partie "rag_api".

```
165     rag_api:
166         container_name: rag_api
167         image: ghcr.io/danny-avila/librechat-rag-api-dev:latest
168         environment:
169             - DB_HOST=vectordb
170             - RAG_PORT=${RAG_PORT:-8000}
171             - EMBEDDINGS_PROVIDER=ollama
172             - OLLAMA_BASE_URL=http://sp26.vincennes.fr:11434/
173             - EMBEDDINGS_MODEL=bge-m3
174         restart: always
175         depends_on:
176             - vectordb
```

Les explications pour les différents champs sur dans la documentation Librechat :

https://www.librechat.ai/docs/configuration/rag_api

5. Cliquer sur Deploy pour mettre à jour Librechat

Tests

Vérifier le bon fonctionnement.

Les logs du container Librechat indiqueront les problèmes éventuels.