

Chart Helm vLLM

[HTTPS://GITHUB.COM/VLLM-PROJECT/VLLM/PULL/9199](https://github.com/vllm-project/vllm/pull/9199)

Autres liens utiles:

Contact LinkedIn :

<https://www.linkedin.com/in/maxime-fournioux-18621925/>

Adresse mail :

Maxime.fournioux@francetravail.fr

DESCRIPTIF PRODUIT

Ce produit a pour objectif de fournir un exemple de chart helm pour déployer vLLM sur un cluster Kubernetes.

ROADMAP PRODUIT

Une évolution anticipée est d' améliorer la distribution des modèles LLMs pendant le déploiement sur k8s en utilisant des images OCIs