

API D'INFÉRENCE · COMPATIBLE OPENAI

L'IA souveraine, hébergée en France.

La puissance d'un grand modèle de langage, d'un moteur d'embeddings et d'un reranker — au même endroit, sous droit français, sans que vos données ne quittent jamais le territoire.

ai.3gks.cloud

Qualité · Réactivité · Accompagnement

Une IA française, de bout en bout

Une API d'intelligence artificielle française, hébergée sur l'infrastructure 3GK, sans aucune dépendance à un fournisseur extra-européen — la chaîne complète (grand modèle de langage, embeddings, reranking) sous droit français, sans que vos données ne quittent jamais le territoire.

EN RÉSUMÉ

HÉBERGEMENT	100 % France, infrastructure GPU 3GK en datacenter français
MODÈLE	Qwen3.6-27B — poids ouverts, multimodal, 201 langues
CONTEXTE	Fenêtre jusqu'à 128 000 tokens
CHAÎNE COMPLÈTE	LLM + Embeddings + Reranking, un seul fournisseur
COMPATIBILITÉ	API compatible OpenAI — migration en changeant une URL
CONFORMITÉ	RGPD by design, aucun log du contenu, droit français
CONTINUITÉ	Repli automatique sur infrastructure souveraine en maintenance
TARIFICATION	Au token prépayé, prix figé à l'appel — ou forfait mensuel sans engagement

La souveraineté, vraiment

La plupart des API d'IA reposent sur des infrastructures soumises à un droit extra-européen. Vos prompts, vos documents, vos données clients y transitent — et y sont parfois conservés. L'approche 3GK est l'inverse, par construction.

- **Hébergement intégralement français.** Les modèles tournent sur du matériel GPU que 3GK possède et opère, colocalisé dans un datacenter français. Vos requêtes ne sortent pas du territoire.
- **Aucune journalisation du contenu.** Le contenu des prompts et des réponses n'est jamais enregistré. Seules les métadonnées techniques (volumétrie, horodatage) sont conservées 6 mois — jamais le texte échangé.
- **Aucun entraînement sur vos données.** Vos données ne servent ni à entraîner ni à affiner aucun modèle. Elles vous appartiennent.
- **Droit français.** Contrat et traitement sous droit français, sans exposition à une législation extraterritoriale.
- **Intégration native à l'écosystème 3GK.** Pour les clients déjà hébergés et infogérés chez 3GK, l'IA s'ajoute à une infrastructure que vous maîtrisez déjà, avec un interlocuteur unique.

SECTEURS SENSIBLES

Une réponse directe aux exigences RGPD et aux contraintes de souveraineté
— **finance, immobilier, santé, secteur public.**

02

Qwen3.6-27B, l'état de l'art à poids ouverts

Un choix simple et assumé : servir **un seul modèle de langage, le meilleur de sa catégorie**, plutôt qu'un catalogue dispersé. Le Qwen3.6-27B, retenu pour trois raisons.

27 Md PARAMÈTRES · DENSE	201 LANGUES	128k TOKENS CONTEXTE	3x TEXTE · IMAGE · VIDÉO
------------------------------------	-----------------------	--------------------------------	------------------------------------

- **De tout premier plan.** Modèle dense de 27 milliards de paramètres, il dépasse sur les grands benchmarks de code des modèles open-source bien plus lourds de la génération précédente. Il excelle en production : compréhension de documents, extraction structurée, recherche, raisonnement multilingue.
- **Accessible et pérenne.** Modèle à poids ouverts (licence Apache 2.0). 3GK détient et opère le modèle sur sa propre infrastructure : personne ne peut le désactiver à distance, le restreindre, ni le retirer du marché.
- **Il évolue.** La famille Qwen progresse rapidement. 3GK évaluera les versions plus performantes publiées en poids ouverts et fera évoluer le service — sans rupture pour vous, l'API restant stable.

POURQUOI C'EST DÉCISIF

En juin 2026, deux modèles propriétaires de premier plan ont été rendus indisponibles en quelques heures, pour l'ensemble de leurs utilisateurs dans le monde, sur directive d'une autorité étatique étrangère. Des clients en production se sont retrouvés sans service, sans préavis ni recours. **Un modèle à poids ouverts servi sur une infrastructure souveraine vous met structurellement à l'abri de ce risque.**

03

Une chaîne RAG souveraine, complète

3GK ne fournit pas seulement un « chat ». L'offre couvre **toute la chaîne technique** pour interroger vos propres documents en langage naturel — un pipeline **RAG** (*Retrieval-Augmented Generation*) — chez un seul fournisseur français.

01 •

Embeddings

Chaque document devient une représentation vectorielle qui capture son sens. Vos contenus deviennent « cherchables » par similarité, au-delà du mot-clé.



02 •

Reranking

Parmi les documents candidats, le reranker replace les plus pertinents en tête. L'étape qui fait passer un résultat « correct » à « précis ».



03 •

LLM [Qwen3.6-27B](#)

À partir des documents pertinents, le modèle rédige la réponse, en extrait l'information structurée, ou produit une synthèse.

EXEMPLE CONCRET

Une société dispose de milliers de documents internes — contrats, comptes-rendus, fiches techniques. Ses équipes posent une question en langage naturel : les **embeddings** identifient les documents proches, le **reranking** sélectionne les plus pertinents, le **LLM** rédige une réponse sourcée — ou en extrait les données utiles (JSON, tableau...). Sans qu'aucun document ne quitte l'infrastructure souveraine.

Vous n'avez pas à assembler trois prestataires aux régimes juridiques différents : **tout au même endroit, sous le même cadre de conformité, une seule facture.**

04

MIGRATION

Une migration sans friction

L'API 3GK est **compatible avec le standard OpenAI**. Si vous utilisez déjà une bibliothèque cliente OpenAI (ou tout outil qui s'y conforme), la bascule se résume à deux changements.

- 1 Changer l'URL de base (`base_url`) pour pointer vers `ai.3gks.cloud`.
- 2 Remplacer votre clé d'API par une clé 3GK (préfixe `sk-3gk-`).

Aucune réécriture de code applicatif. La friction de migration est quasi nulle.

```
# client OpenAI existant
base_url = ".../v1"

# vers 3GK – souverain
base_url =
"https://ai.3gks.cloud/v1"
api_key = "sk-3gk-....."
```

05

CONTINUITÉ

Continuité de service garantie

Aucune infrastructure n'est à l'abri d'une opération de maintenance ou d'un incident matériel. 3GK l'assume par conception : en cas d'indisponibilité de l'infrastructure principale, **le service bascule automatiquement** sur une infrastructure de repli, elle aussi souveraine.

NOMINAL

Infrastructure GPU 3GK

Datacenter français

bascule auto



REPLI

Partenaire français (Scaleway)

Sol français, même conformité

La continuité prime, **sans jamais compromettre la souveraineté** : le repli reste sur le territoire, sous droit français. Vos appels continuent de répondre, de façon transparente.

06

TARIFICATION

Transparente, au plus juste

Deux formules au choix, selon votre usage — **à l'usage** (au token), ou **forfait mensuel** pour les volumes réguliers.

A

À l'usage

AU TOKEN · PRÉPAYÉ · SANS ABONNEMENT

Démarrage & volumes variables

SERVICE	MODÈLE	TARIF
Génération (LLM)	Qwen3.6-27B	0,80 € / M entrée · 1,80 € / M sortie
Embeddings	qwen3-embedding-4b	0,12 € / M entrée
Reranking	qwen3-reranker-4b	2,00 € / M (requête + documents)

Crédit prépayé, prix figé à l'appel. Le prix appliqué est celui en vigueur au moment exact de l'appel : aucune révision rétroactive, traçabilité comptable intégrale. *Tarifs HT. Reranking facturé sur requête + documents, sans tokens de sortie.*

OU

B

Forfait mensuel

VOLUME · USAGE RÉGULIER

500 M tokens / mois

450 € / mois HT Sans engagement, reconductible.

Enveloppe unique — LLM (entrée + sortie) et reranking confondus.

Embeddings gratuits — hors enveloppe, non décomptés.

QoS automatique — capacité répartie équitablement entre clients.

Au-delà de 500 M — alerte 80 % / 100 %, puis tarif au token. Plafond strict activable.

07 SÉCURITÉ & CONFORMITÉ

La conformité, intégrée

RGPD BY DESIGN

Aucune journalisation du contenu, pas d'entraînement sur vos données, hébergement et traitement en France.

RÉTENTION MAÎTRISÉE

Métadonnées techniques conservées 6 mois ; le contenu échangé n'est jamais conservé.

AUTHENTIFICATION RENFORCÉE

Authentification à deux facteurs (TOTP) disponible et activable sur votre compte.

CLÉS D'API DÉDIÉES

Préfixe `sk-3gk-`, révocables à tout moment.

08 LE PARTENAIRE

3GK SOFTWARE, votre partenaire unique

3GK n'est pas qu'un fournisseur d'API. C'est un **hébergeur** et un **infogérant** français, éditeur de logiciels métier. Pour vous, cela signifie :

- **Un interlocuteur unique** pour l'hébergement, l'infogérance et l'IA.
- **Un support humain francophone**, sans barrière de fuseau ni de langue.
- **Une IA qui s'intègre nativement** à l'infrastructure que 3GK opère déjà pour vous.

Là où les API extra-européennes vous laissent seul face à un portail en libre-service, 3GK vous accompagne.

PASSONS À L'ACTION

Reprenez le contrôle de votre IA.

Parlons de votre cas d'usage et obtenez une clé d'API souveraine.
Migration accompagnée, sans réécriture de code.

WEB

ai.3gks.cloud

SOCIÉTÉ

3GK SOFTWARE — Tours, France

Tarifs hors taxes. Conditions détaillées dans les CGU / DPA.