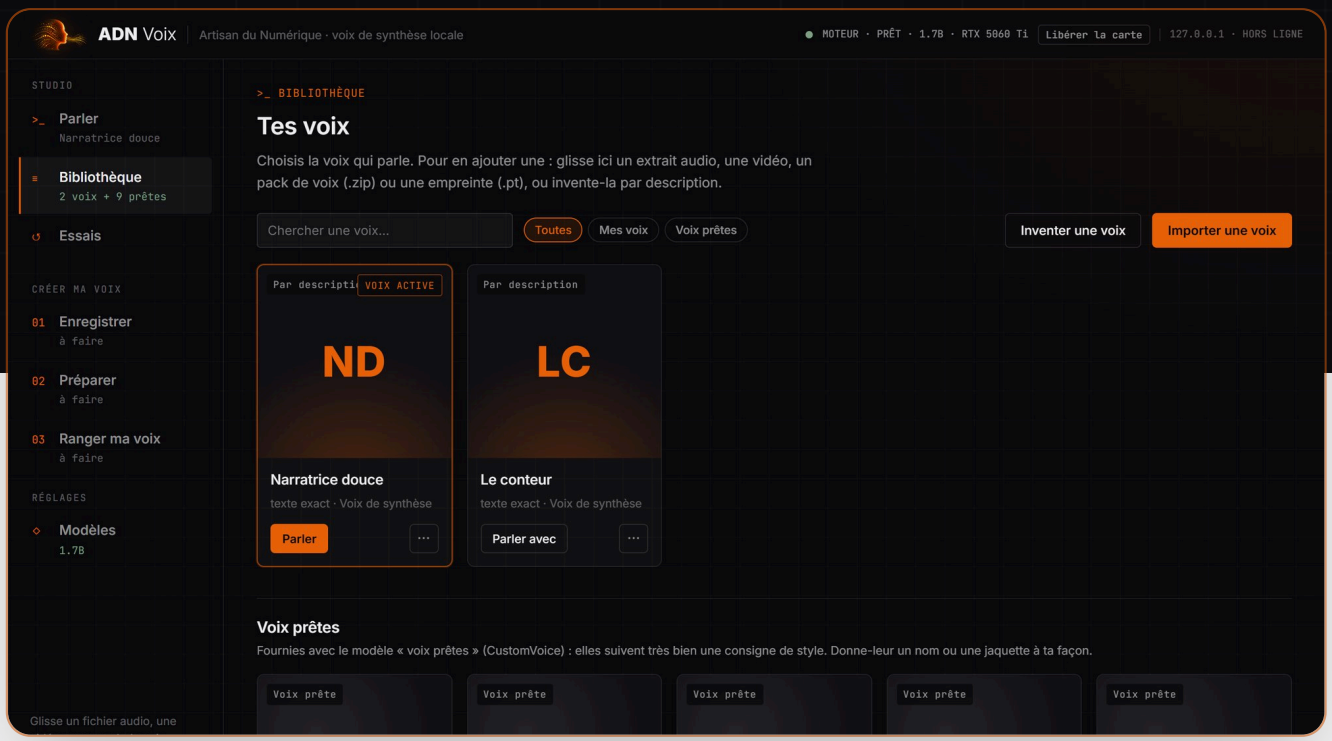
LOGICIEL LOCAL · VOIX DE SYNTHÈSE

ADN Voix

Cloner sa propre voix et la faire parler, entièrement sur son ordinateur : un studio de voix de synthèse sans abonnement, sans envoi en ligne, et qui prête sa voix au jumeau numérique.



CRÉÉ LE
20 octobre 2026, au soir

VERSION
1.1.3 · 21 octobre 2026

NATURE
appli locale + logiciel Windows à installer

MOTEUR
Qwen3-TTS (licence Apache 2.0), carte NVIDIA

RÉSEAU
ce PC seulement · pare-feu interne

RÔLE
la voix du jumeau numérique (ADN Jumeau)

DANS CE DOSSIER

01 En bref
02 L'idée de départ
03 À quoi il sert
04 Les tenants et les aboutissants
05 Les dates
06 Ce qu'il contient
07 La façon de procéder
08 Le rendu final
09 Où il en est

01

En bref

ADN Voix est un **studio de voix de synthèse qui tourne sur la carte graphique du PC**. On y enregistre une fois sa voix au micro ; ensuite, on écrit un texte, aussi long qu'on veut, et le logiciel le dit avec cette voix.

Il range les voix dans une bibliothèque : la sienne, des voix importées depuis un extrait audio ou vidéo, des voix inventées à partir d'une simple description (« homme âgé, voix grave de conteur »), et neuf voix prêtes à l'emploi. Une consigne de style (« chaleureux », « posé », « comme un conteur ») change la façon de parler sans changer le timbre.

Rien ne sort du PC : ni la référence enregistrée, ni les voix, ni les textes. ADN Voix sert aussi de **voix au jumeau numérique** de Thibault, qui lui confie la lecture de ses réponses.

2 min 13
de voix calculées en 79 s (lecture par lots)

10 à 13 s
avant le premier son d'une longue réponse

0,985
ressemblance mesurée de la voix clonée avec une consigne de style

114 Mo
l'installeur ; le moteur s'installe au premier lancement

02

L'idée de départ

Le 20 octobre 2026, Thibault ouvre le chantier du **jumeau numérique** : un dossier qui apprend qui il est, pour un jour répondre à sa place. Le soir même, il se pose la question de la voix : avec des enregistrements et leurs transcriptions, pourrait-on recréer un programme qui parle comme lui ?

> *Télécharge carrément le bon modèle de clonage de voix et on préparera quelque chose que j'enregistrerai spécialement pour ça avec le SM58.*
>
> en substance — 20 octobre 2026, 20 h 49

L'objectif : **entendre sa voix clonée, en local**, pour alimenter le jumeau ; et préparer le terrain pour qu'il n'y ait plus qu'à fournir l'enregistrement de référence. Le lendemain, une autre envie s'ajoute : en faire un logiciel à **donner à un ami**.

>_ Le choix du modèle

Le modèle retenu est **Qwen3-TTS 1.7B Base** : il parle français, clone une voix à partir de quelques secondes, fait mieux quand on lui donne le texte exact de la référence, et sa licence (Apache 2.0) permet un usage commercial. Ont été écartés Voxtral, XTTS, F5-TTS et Fish Speech, dont les licences interdisent l'usage commercial ; Chatterbox (licence MIT) reste l'alternative à comparer si le français déçoit.

03

À quoi il sert

- **Créer sa voix, une seule fois** : enregistrer un paragraphe au micro avec un prompteur et un vumètre, laisser l'appli contrôler le niveau (saturation, volume, durée), couper les silences et égaliser le son, puis ranger la voix dans la bibliothèque.
- **Faire parler une voix** sur un texte sans limite de longueur, avec, si on veut, une façon de parler. Chaque essai est gardé (texte, voix, consigne) et se reprend en un clic.
- **Gérer une bibliothèque de voix** : importer un extrait (audio, vidéo, pack .zip), inventer une voix par description, utiliser les voix prêtes, renommer, donner une jaquette, exporter.
- **Choisir le bon modèle pour chaque usage** (cloner, voix prêtes, voix par description), téléchargé une fois puis chargé tout seul au besoin ; « Libérer la carte » rend la carte graphique sans fermer l'appli.
- **Prêter sa voix au jumeau** : ADN Jumeau envoie ses réponses à ADN Voix et les fait lire avec la voix de Thibault, morceau après morceau.
- **Travailler autrement** si on préfère : trois scripts en console, ou des nœuds maison dans ComfyUI.

04

Les tenants et les aboutissants

>_ Une voix clonée est une clé

Une voix clonée peut servir à une fraude (un faux appel d'un proche, par exemple). Le projet en tire des règles qui ne se discutent pas :

- la référence, les voix et les fichiers produits **restent sur le PC** : jamais en ligne, jamais par messagerie ;
- chaque fichier produit part avec un **avertissement « voix de synthèse »**, et toute diffusion se signale (c'est aussi une obligation du règlement européen sur l'IA) ;
- **jamais la voix d'un tiers** sans son accord écrit : l'appli le rappelle au moment d'enregistrer ;
- un **mot de code familial** est conseillé pour déjouer un faux appel d'urgence.

>_ Les choix techniques qui en découlent

- **100 % local.** L'appli n'écoute que `127.0.0.1` ; un pare-feu interne lui interdit toute connexion vers l'extérieur ; un site ouvert dans un navigateur ne peut ni lire une voix ni allumer le micro.
- **Un format de voix universel** : l'extrait audio et le texte exact qu'il contient. L'empreinte propre à chaque modèle s'en recalcule en 2 secondes, donc changer de modèle ne casse rien, et d'autres logiciels de clonage savent lire ces packs.

- **La consigne de style, mesurée.** Le modèle de clonage l'ignore ; une voix clonée avec consigne passe donc par le modèle « voix prêtes » en lui donnant l'empreinte de la voix. Résultat mesuré : le style suit (débit 40 % plus lent pour « lent », plus fort et plus aigu pour « enthousiaste ») et la voix reste reconnaissable (0,985 de ressemblance, contre 0,992 sans consigne ; deux voix différentes sont autour de 0,94).
- **Une carte graphique partagée.** Un seul modèle en mémoire à la fois ; ADN Voix et ADN Jumeau se passent la carte à tour de rôle sans qu'on ait à fermer quoi que ce soit.

> LIMITES
>
> Une carte NVIDIA est conseillée (8 Go pour le modèle 1.7B) ; sans elle, le processeur met plusieurs minutes par phrase. Le premier lancement du logiciel installé télécharge son moteur (environ 3,5 Go). L'installeur n'est pas signé : Windows affiche un avertissement la première fois.

05

Les dates

20/10 · 20 h 45 — La question, puis le choix du modèle

Comparaison des modèles et de leurs licences, installation du moteur (Python, PyTorch pour la RTX 5060 Ti, Qwen3-TTS) et de trois scripts : enregistrer, préparer, cloner.

20/10 · 21 h 30 — Branché dans ComfyUI

Nœuds maison (texte long, enregistrement, transcription), trois graphes prêts, fiche PDF de quatre pages.

20/10 · soir — L'appli « ADN Voix »

Une fenêtre en quatre étapes : enregistrer avec prompteur et vumètre, préparer, créer la voix, parler.

21/10 · 0 h 30 — La vraie référence

Enregistrement au micro SM58 sur la table Yamaha AG08 : la voix clonée de Thibault entre dans l'appli.

21/10 · nuit — Version 1.1

Bibliothèque, voix importées et inventées, consignes de style, un modèle par usage, logo.

21/10 · journée — Le logiciel à donner

« ADN Voix.exe » et son installeur, testés de bout en bout puis installés ; garde contre l'emballement, « Libérer la carte ».

21/10 · soir — Versions 1.1.1 à 1.1.3

Lecture par lots (trois fois plus rapide), lecture progressive, puis « tapis roulant » : la voix lit un morceau pendant qu'elle calcule les suivants.

06

Ce qu'il contient

PARTIE	RÔLE
`appli/serveur.py + api/`	le serveur local (FastAPI) et ses 7 groupes de routes
`appli/metier/`	23 modules : moteur, génération, découpage du texte, lots, tapis roulant, bibliothèque, garde réseau…
`appli/web/`	l'interface : 20 modules JavaScript, 5 feuilles de style, polices de la charte
`logiciel/`	le logiciel à installer (Electron + installeur) ; installe son moteur au premier lancement
`bibliotheque/`	une voix = un dossier : fiche, extrait, texte exact, jaquette, empreintes par modèle
`1- 2- 3-*.bat + _outils/`	le parcours en console : enregistrer, préparer, cloner
`ComfyUI-ADN-Voix`	nœuds maison pour ComfyUI, trois graphes, fiche d'utilisation
`_moteur/`	Python, PyTorch et les modèles (hors sauvegarde) ; le supprimer désinstalle tout

88
fichiers de code

6 500
lignes de code environ

5
modèles officiels pris en charge

9
voix prêtes fournies avec le modèle

07

La façon de procéder

- **Préparer le terrain avant d'avoir la voix.** Tout a été construit et testé sur une voix d'essai tirée d'une vidéo ; la vraie référence n'a eu qu'à prendre sa place.
- **Mesurer plutôt que deviner** : durée, énergie, hauteur et ressemblance pour juger une consigne ; secondes de calcul par seconde de voix pour la vitesse ; durée des blancs pour la lecture continue.
- **Corriger depuis l'usage.** Une phrase contenant « 20 versions », un tiret long et des guillemets faisait marmonner le modèle pendant 164 s : chaque morceau a désormais un plafond proportionnel à son texte, et il est refait s'il le dépasse. « La carte ne tourne qu'à 33 % » : lecture par lots. « Qu'elle lise le 1 pendant qu'elle calcule le 2 » : tapis roulant.
- **Vérifier la machine avant d'accuser le code** : des lenteurs pendant les tests venaient d'un jeu qui occupait 71 % de la carte graphique.
- **Tester l'installeur comme un inconnu** : installation dans des dossiers isolés, téléchargement d'un modèle puis annulation, petite carte, fermeture ; puis installation réelle sur le poste de Thibault.

08

Le rendu final

Captures faites le 25/10/2026 sur l'appli réelle, lancée à part avec deux **voix de démonstration inventées** (« Narratrice douce », « Le conteur ») : la voix de Thibault n'apparaît pas.

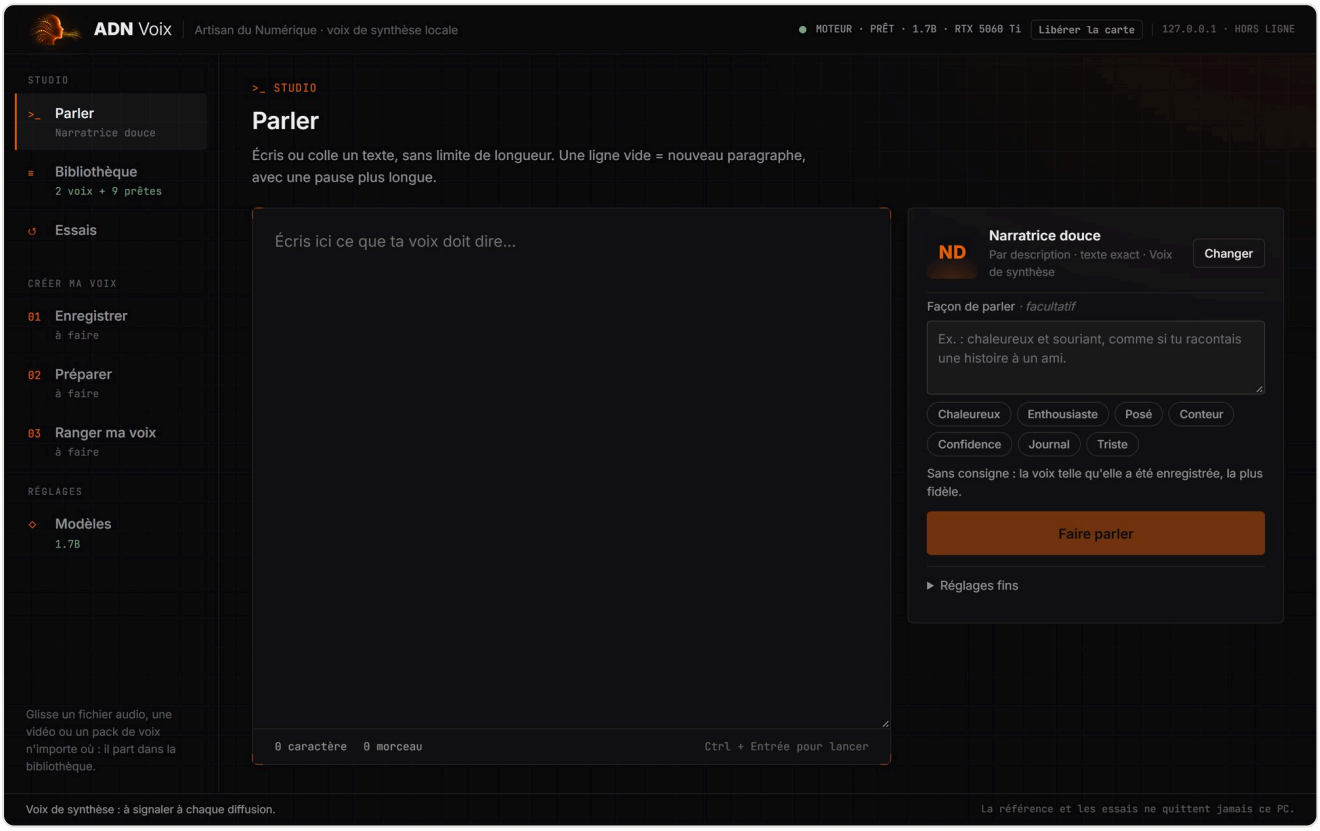


fig. 1 — Parler : le texte à dire, la voix choisie, une façon de parler en un clic ; le témoin du moteur en haut à droite.

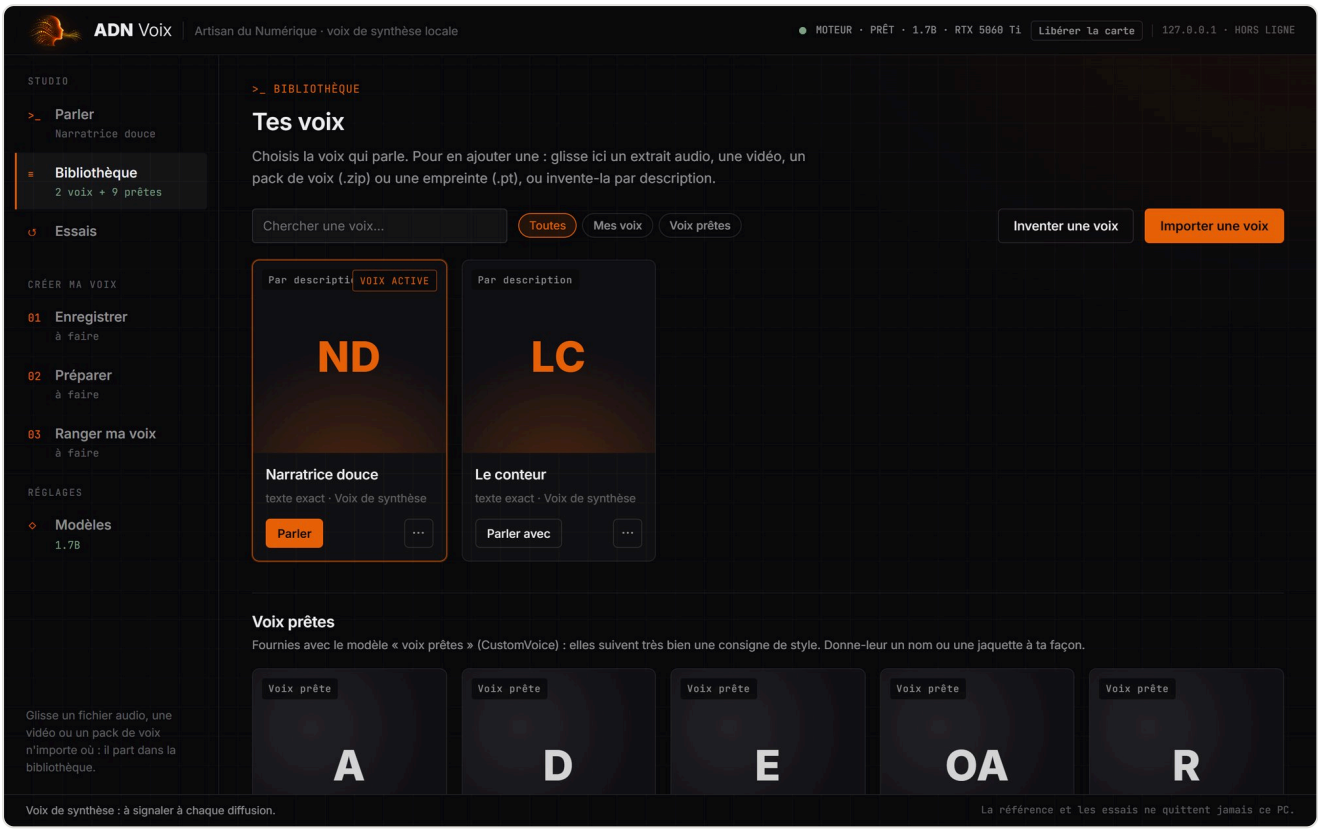


fig. 2 — La bibliothèque : ses voix, les voix inventées, les neuf voix prêtes ; importer ou inventer une voix.

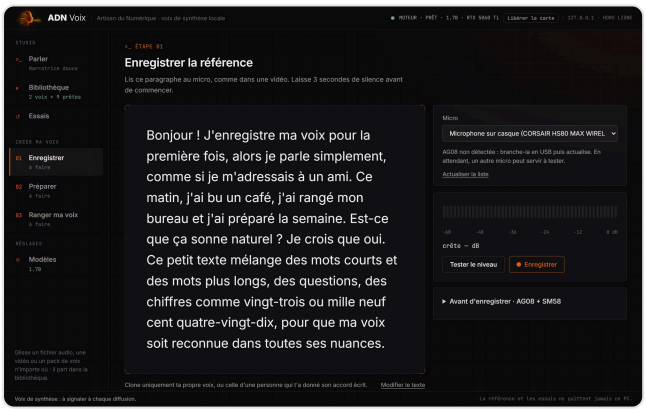


fig. 3 — Étape 01 : enregistrer la référence avec prompteur et vumètre.

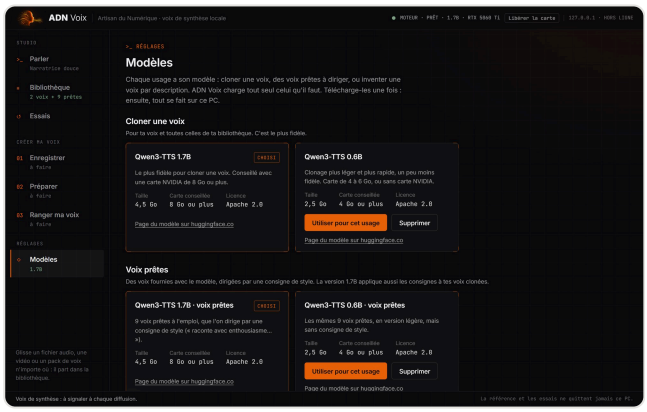


fig. 4 — Les modèles : un par usage, taille, carte conseillée, licence.

CE QUE PRODUIT ADN VOIX

Des fichiers audio WAV 24 bits, rangés avec leur avertissement « voix de synthèse », des voix exportables en packs, et une lecture en continu pour ADN Jumeau : premier son en 10 à 13 secondes, puis aucun blanc entre les morceaux.

09

Où il en est

La version **1.1.3** est installée sur le poste de Thibault et sert de voix au jumeau ; l'installeur est prêt à être donné. Pistes ouvertes :

- comparer avec Chatterbox (licence MIT) si le rendu du français doit encore progresser ;
- signer le logiciel pour supprimer l'avertissement de Windows.

Sources : fichiers du projet, notes de développement, captures faites le 25/10/2026 avec des voix de démonstration

>_ Artisan du Numérique