



# Transcription audio haute performance sur Station TV

PRI 5A - Formation ISIE  
Polytech Tours - 2025-2026

Chef de projet : Dorian BRISSON  
Client : M. DELALANDRE

# Sommaire



- Contexte et existant
- Problématique et objectifs
- Gestion de projet
- Choix techniques
- Conception
- Protocole expérimental
- Résultats
- Conclusion et perspectives



# Contexte et existant

## Contexte : Station TV

- Projet Station TV (LIFAT)
- Flux TNT
- Plateforme de calcul parallèle
- Services : analytics TV, fact-checking, détection publicité..



- Transcription audio



# Matériel

- Dell Precision 5820
- CPU : Intel Xeon W-2295 (18 cœurs physiques, 36 threads)
- RAM : 64 Go
- OS : Windows





# Problématique et objectifs

# Problématique



La problématique est de transcrire de grandes quantités d'audio de la TNT, de qualité variable, tout en garantissant un bon compromis entre qualité linguistique, performance et stabilité sur des exécutions longues.

## Existant : travaux précédents



- Un premier script multi-processus a été développé, permettant de lancer des transcriptions en parallèle sur plusieurs fichiers audio.
- Une étude préalable sur les solutions de speech-to-text a été menée.
- Whisper a été retenu comme solution de transcription, notamment car il entraîné sur un nombre colossal de données en multilingue.



## Limites de l'existant et constats

---

- La solution actuelle fonctionne mais il n'y a pas de reprise automatique en cas d'erreur, ni de métriques générées à la fin d'un traitement. Un crash ou une interruption implique de relancer manuellement les transcriptions.
- Lors des tests avec des modèles Whisper de plus haute qualité, une limitation mémoire a été identifiée
- Rechargement du modèle à chaque fichier, ce qui entraîne une perte de performances importante

## Objectifs du projet



-  volution mat rielle de la station TV, en augmentant la m moire vive de 64 Go   256 Go.
- Mettre en place des m triques automatis es afin de mesurer les performances (le d bit de transcription, l'utilisation des ressources etc...)
-  valuer les mod les Whisper de plus haute qualit , notamment les mod les medium et large
- Mise en place d'une pipeline de transcription en mode run/batch, capable de traiter automatiquement de grands volumes de fichiers audio.



# Gestion de projet

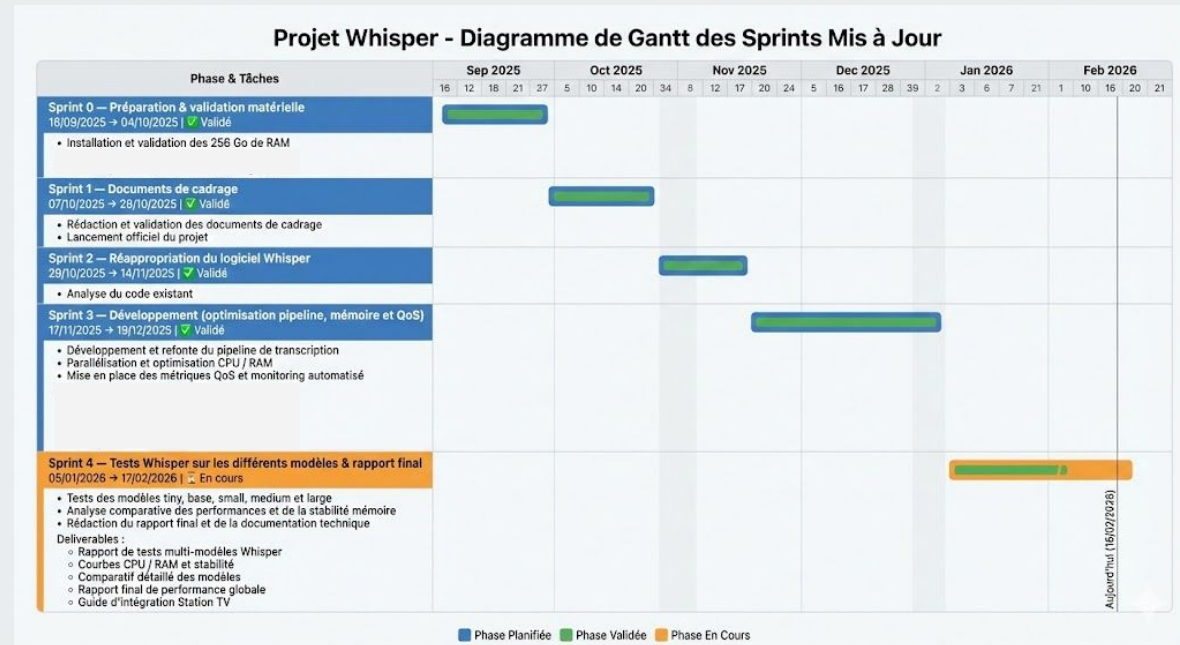
# Organisation



| Rôle               | Nom                | Organisation   |
|--------------------|--------------------|----------------|
| Chef de projet     | Dorian BRISSON     | Polytech Tours |
| Client / Encadrant | Mathieu DELALANDRE | Polytech Tours |

# Méthodologie

- Méthodologie agile
- Travail par sprints



**PRI\_5A\_Station\_TV**
Public

Watch 0
 Fork 0
 Star 0

main
 1 Branch
 0 Tags

Add file
 Code

dorianDB
 Ajout de la gestion de projet et du rapport de transcription
 3211604 · 1 minute ago
 6 Commits

|                                  |   |               |
|----------------------------------|---|---------------|
| 00 - Documentation/PRI-2025-2026 | Diapo vide  | 3 months ago  |
| 01 - Cahier des charges/CDC      | Structure projet + Doc  | 3 months ago  |
| 02 - Gestion de projet           | Ajout de la gestion de projet et du rapport de transcription    | 1 minute ago  |
| 03 - Rapports et Diapos          | Ajout de la gestion de projet et du rapport de transcription    | 1 minute ago  |
| 04 - Technique                   | Mise à jour du projet : nettoyage du cache, ajout des nouvea... | 3 minutes ago |
| .gitattributes                   | Initial commit  | 3 months ago  |
| .gitignore                       | Structure projet + Doc  | 3 months ago  |

README

## Add a README

Help people interested in this repository understand your project.

Add a README

### About

Projet PRI de 5A sur la transcription audio sur station TV, Polytech Tours

Activity  
 0 stars  
 0 watching  
 0 forks

### Releases

No releases published

[Create a new release](#)

### Packages

No packages published

[Publish your first package](#)

### Languages

Python 99.4%
 Batchfile 0.6%

### Suggested workflows

Based on your tech stack

**Python Package using Anaconda**

Configure

Create and test a Python package on multiple Python versions using Anaconda for package management.

**Python package**

Configure

Create and test a Python package on multiple Python versions.

**Pylint**

Configure

Lint a Python application with pylint.

[More workflows](#)
[Dismiss suggestions](#)

# Communication





# Choix techniques



# Outils de transcription

| Outil / Variante                             | WER (FR)<br>approximatif | Vitesse<br>CPU    | RAM nécessaire typique<br>(quantité) |
|--|--------------------------|-------------------|--------------------------------------|
| Whisper (full)                               | ~8-15 %                  | Moyen             | ~2-5 GB+                             |
| <u>Vosk</u> (modèles FR)                     | ~10-30 %                 | Très rapide       | 50-500 MB modèles<br>légers          |
| <u>wav2vec2-fr</u> (base)                    | ~12-20 %                 | Moyen             | ~1-2 GB                              |
| Kaldi avec modèle FR                         | ~15 %+                   | Très rapide       | ~100-600 MB (dépend<br>modèle)       |
| <u>Coqui STT</u> /<br><u>DeepSpeech</u> (FR) | ~15 %+                   | Moyen à<br>rapide | ~500 MB-1 GB                         |

Source : northflank.com

## Whisper : principe général

Whisper est un modèle de speech-to-text, c'est-à-dire qu'il permet de convertir automatiquement un signal audio en texte. Il est entraîné sur de nombreuses langues, ce qui lui permet de détecter automatiquement la langue et de transcrire des contenus multilingues. Le modèle est également reconnu pour sa robustesse face au bruit, aux accents et aux conditions audio dégradées, fréquentes dans les flux télévisuels.



## Whisper : modèles et compromis



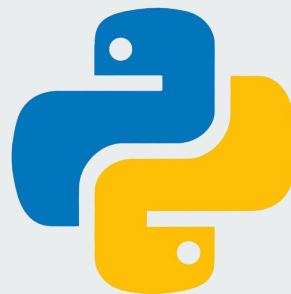
- Plusieurs tailles de modèle : tiny, base, small, medium et large
- Qualité croissante
- Coût RAM croissant

## Technologie retenue

- Whisper



- Python



## Défis système liés à Whisper



- Les modèles Whisper, en particulier les modèles medium et large, consomment une quantité importante de mémoire vive.
- Le temps de calcul par fichier est significatif, en particulier pour les modèles de grande taille. Sans optimisation ou parallélisation, le débit de transcription devient insuffisant pour des volumes importants de données audio.
- Pour atteindre des débits compatibles avec les besoins de la station TV, la parallélisation des traitements est indispensable.



# Conception

# RAM

TICKET n°100769 : Installation RAM haute performance - station TV

ACTIONS ▼

Statut : Archivé

Catégorie : DSI - Acquérir du matériel informatique

Propriétaire : Mathieu Delalandre (delalandre)

Gestionnaire : Sebastien Beaufils (sbeaufils)

Historique Propriétés Fichiers attachés Invités

13/10/25 16:01 : le ticket a expiré.

29/09/25 15:13 : Sebastien Beaufils (sbeaufils) a clos le ticket.

RAM installée, test -> OK

29/09/25 15:13 : Sebastien Beaufils (sbeaufils) a modifié le temps passé.

24/09/25 17:32 : Dorian Brisson (21801799t) a ajouté un commentaire.

Je serai présent. Je vais prévenir mon enseignant par précaution.

Bien cordialement,

24/09/25 17:29 : Mathieu Delalandre (delalandre) a ajouté un commentaire.

Oui c'est ok, merci Dorian de vous rendre disponible et de prévenir au besoin votre enseignant de 14h (si retard en cours).

Bien Cordialement,

24/09/25 09:53 : Sebastien Beaufils (sbeaufils) a ajouté un commentaire.

Lundi 13h30 ?

24/09/25 09:36 : Mathieu Delalandre (delalandre) a ajouté un commentaire.

Plus lundi, de préférence début d'après-midi ou à partir de 9h30 le matin.  
Devons nous prévoir de déplacer la machine, ou on pourra faire cela directement sur la station.

Je précise qui y a deux jeux de barette, soit 4x 64 Go.

Merci

24/09/25 09:13 : Sebastien Beaufils (sbeaufils) a pris le ticket en charge (précédemment libre).

Vendredi ou lundi prochain ?

23/09/25 15:30 : Mathieu Delalandre (delalandre) a ajouté Dorian Brisson (21801799t) à la liste des invités.

23/09/25 15:28 : Mathieu Delalandre (delalandre) a ajouté Sebastien Beaufils (sbeaufils) à la liste des invités.

23/09/25 15:28 : Mathieu Delalandre (delalandre) a créé le ticket.

## Panier – 1 Produit(s)

Vendu(s) par : Conrad Electronic



Kingston KTD-PE432/64G Module  
mémoire pour PC DDR4 64 GB 1 x 64 GB  
ECC 3200 MHz DIMM 288 broches CL22  
KTD-PE432/64G

Livraison : 04.07.2025 -  
08.07.2025

29...

Code produit : 2823381

— pièces  
2 +



652,50 €  
Avec 10% de remise client  
Prix unitaire 326,25 €

Prix HT 652,50 €

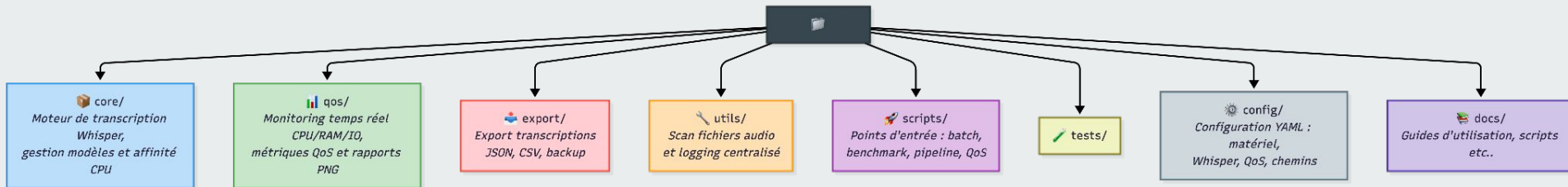
Frais d'expédition 8,99 €

TVA 132,30 €

Prix total 793,79 €  
TVA incluse.

# Architecture globale du système

- Approche modulaire
- Chaque module est indépendant
- Maintenabilité du code





# Fichier de configuration



## Hardware — Matériel cible

- Nombre de cœurs physiques
- Nombre de threads logiques
- Nombre maximal de processus Whisper simultanés
- Quantité totale de mémoire vive
- Seuil maximal d'utilisation de la RAM

## Moteur de transcription

- Modèle Whisper utilisé
- Langue de transcription
- Type de périphérique d'exécution (CPU / GPU)
- Affinité CPU par processus
- Formats de sortie générés (TXT, SRT)
- Horodatage au niveau des mots

# Fichier de configuration



## QoS — Supervision & monitoring

- Intervalle de mesure
- Mesures d'utilisation CPU
- Mesures d'utilisation mémoire
- Mesures d'entrées / sorties disque
- Seuil d'alerte CPU (avertissement)
- Seuil d'alerte CPU (critique)
- Seuil d'alerte RAM (avertissement)
- Seuil d'alerte RAM (critique)
- Suivi de la consommation énergétique
- Estimation du coût énergétique

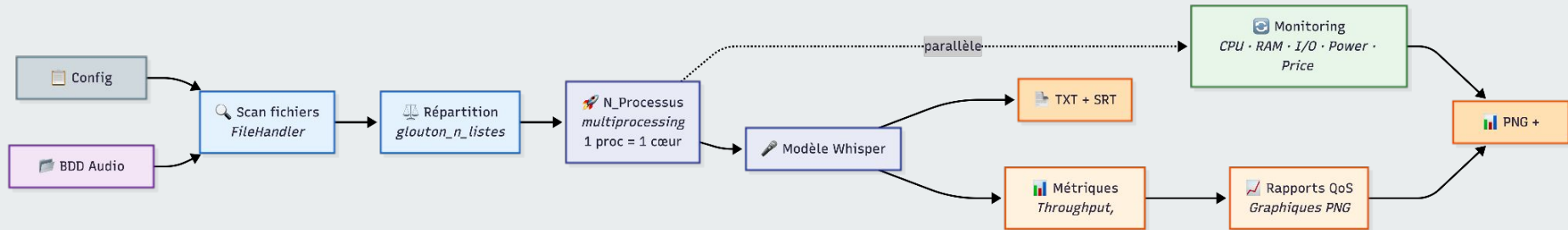
## Export — Sortie des résultats

- Encodage des fichiers de sortie
- Format des métadonnées
- Politique de sauvegarde automatique
- Fréquence des sauvegardes
- Convention de nommage / préfixe d'indexation

## Paths — Chemins d'accès

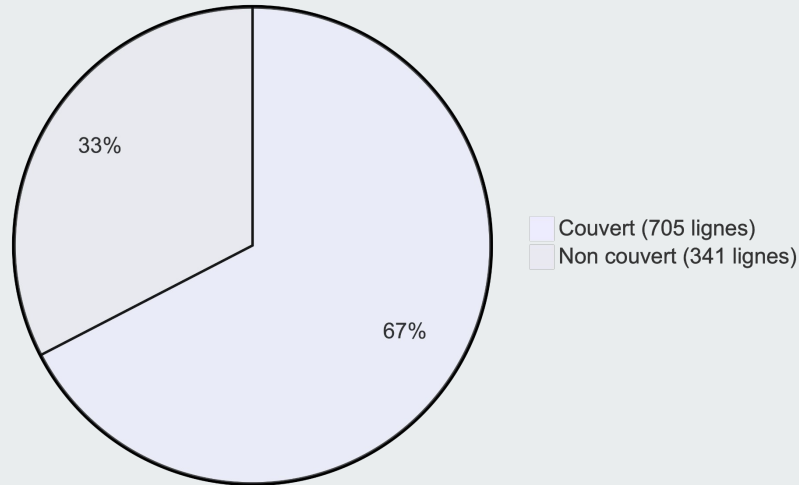
- Répertoire d'entrée des fichiers audio
- Répertoire de sortie des transcriptions
- Répertoire des rapports
- Répertoire des journaux (logs)
- Répertoire des métriques et trackers
- Répertoire des sauvegardes

# Pipeline



# Tests unitaires

Couverture de code



## Métriques



- Taux d'utilisation CPU en %
- Charge mémoire en GB
- Consommation en kWh
- Prix
- Throughput
- Temps moyen pour chaque processus
- I/O
- WER

## Apports techniques majeurs

---

- Amélioration des performances globales de la transcription, en travaillant sur la gestion mémoire, la parallélisation et le chargement des modèles.
- Intégration de métriques automatisées, à la fois sur les performances et sur la qualité des transcriptions.
- Mise en place d'une pipeline de transcription industrialisée.
  - Fichier de configuration
  - Système de sauvegarde et de retries
  - Lancement avec un .bat



# Protocole expérimental

## Bases de test utilisées



- BaseMod 5 fichiers : 0, 240, 480, 720, 960, 1200 en secondes
- BaseBatch : 32 dossiers qui contiennent jusqu'à 24 fichiers d'1h01 d'audio.
- Base HelloWorld (cible)



# Benchmarks

- Benchmark  $10 \times 5 \times$  modèles
- tiny  $\rightarrow$  large
- Mesures throughput

40 à 50% de gain

|      | tiny   | base   | small   | medium  | large |
|------|--------|--------|---------|---------|-------|
| 0    | 0      | 0      | 0       | 0       |       |
| 240  | 55,35  | 97,04  | 270,48  | 778,76  |       |
| 480  | 97,55  | 176,23 | 545,26  | 1435,22 |       |
| 720  | 183,9  | 317,53 | 752,3   | 2181,12 |       |
| 960  | 264,14 | 378,22 | 1311,25 | 2861,25 |       |
| 1200 | 301,51 | 478,76 | 1420,1  | 3630,68 |       |
| Th   | 3,980  | 2,506  | 0,845   | 0,331   |       |
| 1/Th | 0,251  | 0,399  | 1,183   | 3,026   |       |



|      | tiny   | base   | small   | medium  | large   |
|------|--------|--------|---------|---------|---------|
| 0    | 0      | 0      | 0       | 0       | 0       |
| 240  | 39,66  | 65,22  | 176,39  | 526,06  | 1136,98 |
| 480  | 76,69  | 155,73 | 405,5   | 1038,24 | 1822,16 |
| 720  | 127,74 | 198,57 | 626,42  | 1503,48 | 2694,27 |
| 960  | 151,22 | 270,31 | 836,38  | 2120,96 | 4038,48 |
| 1200 | 192,57 | 341,51 | 1021,97 | 2528,38 | 4852,41 |
| Th   | 6,232  | 3,514  | 1,174   | 0,475   | 0,247   |
| 1/Th | 0,160  | 0,285  | 0,852   | 2,107   | 4,044   |

# Benchmarks

- BaseBatch avec k=30, modèle tiny (24 fichiers d'1h01 par process = 728)

| Model           | tiny | base | small         | medium | large |
|-----------------|------|------|---------------|--------|-------|
| CPU rate        |      |      | $\simeq 90\%$ |        |       |
| Memory (GB)     | 22.5 | 27.5 | 48            | 115    | 192   |
| Throughput (Th) | 39.8 | 30.1 | 6.5           | 4.5    | 2     |

## RAPPORT QoS - STATION TV - TRANSCRIPTION AUDIO

### RÉSUMÉ DE LA SESSION

Durée de la session: 13.16 heures  
 Nombre total de fichiers: 720  
 Fichiers réussis: 720  
 Fichiers échoués: 0  
 Taux de réussite: 100%

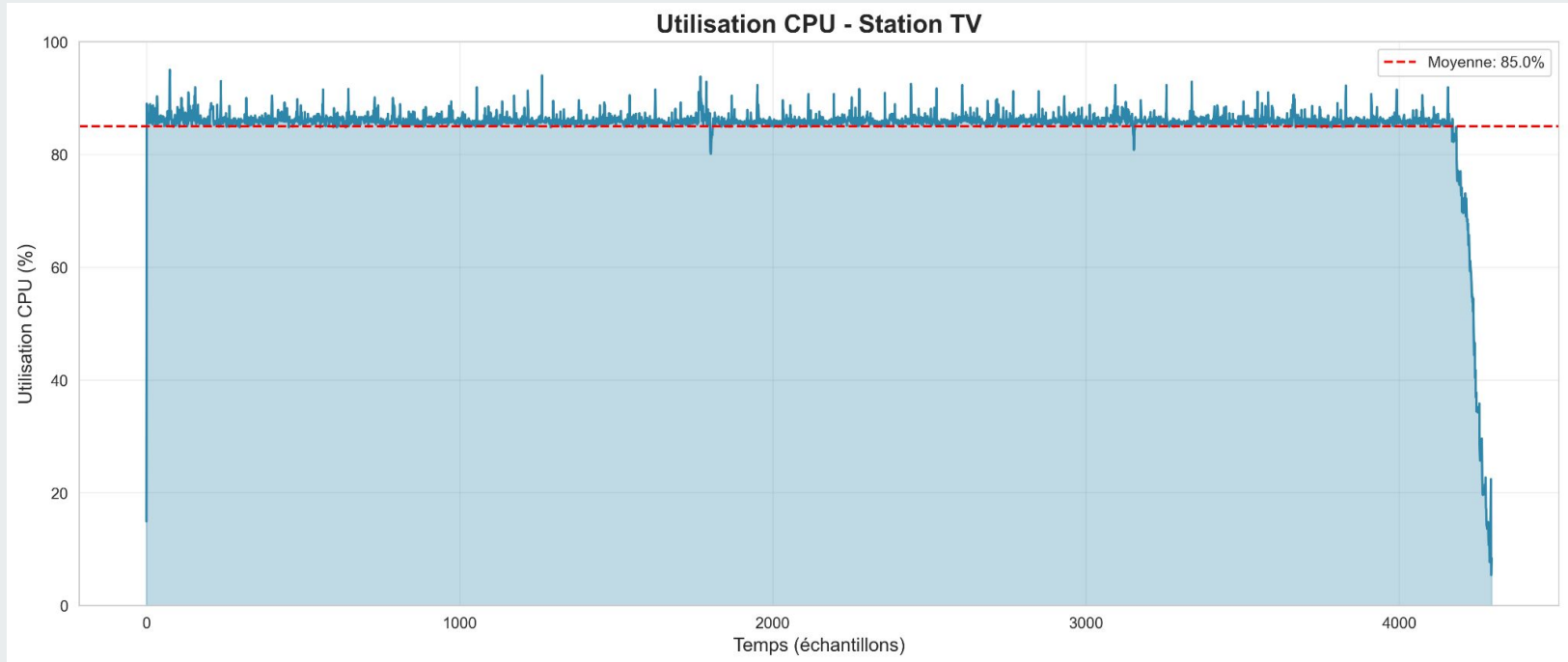
### PERFORMANCE

Durée audio totale traitée: 728 heures  
 Temps de traitement total: 13.16 heures  
 Throughput (Th): 55.32  
 Temps moyen par fichier: 1947.44 secondes

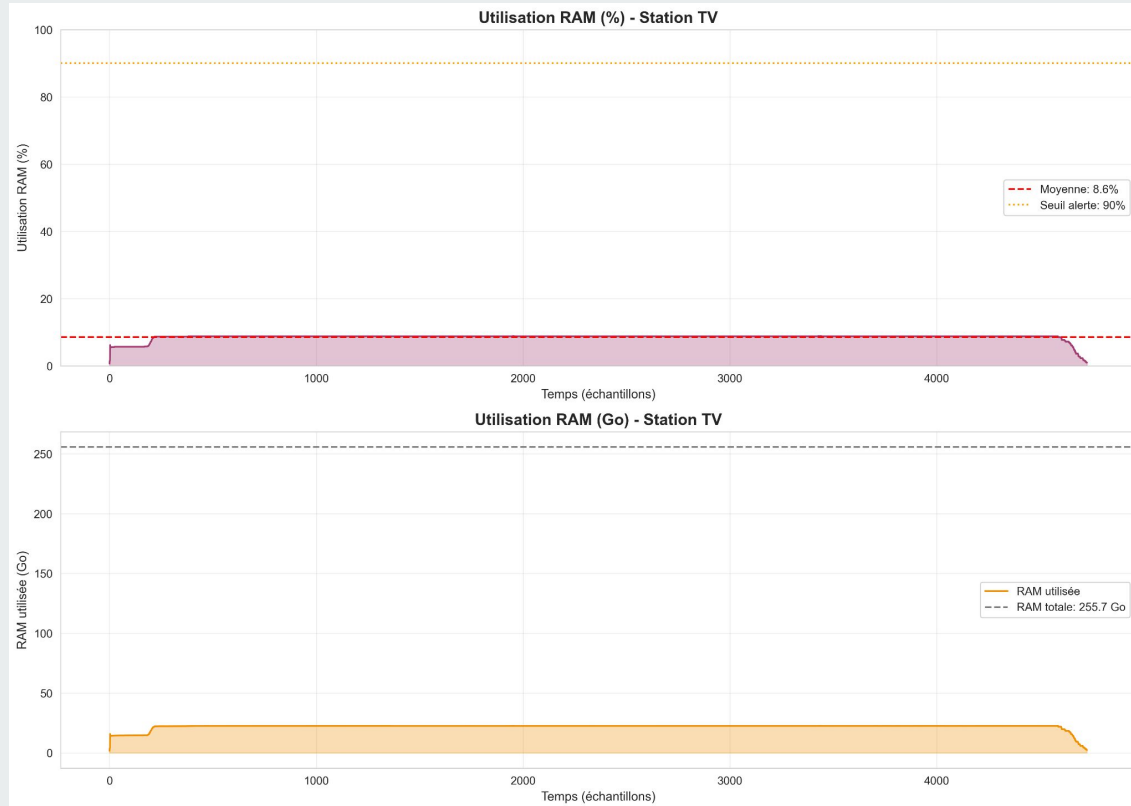
### OBJECTIFS QoS

✓ Taux de réussite 100%  $\geq 99\%$   
 ✓ Traitement terminé avec succès

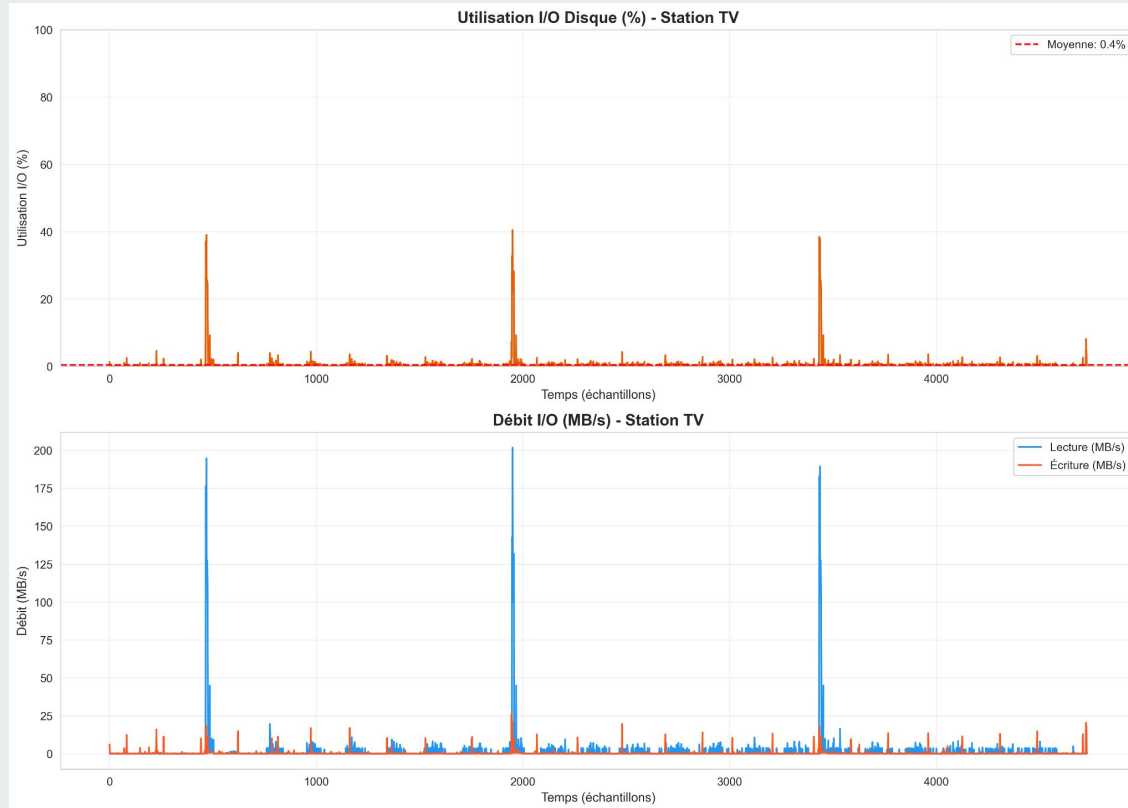
# Benchmarks



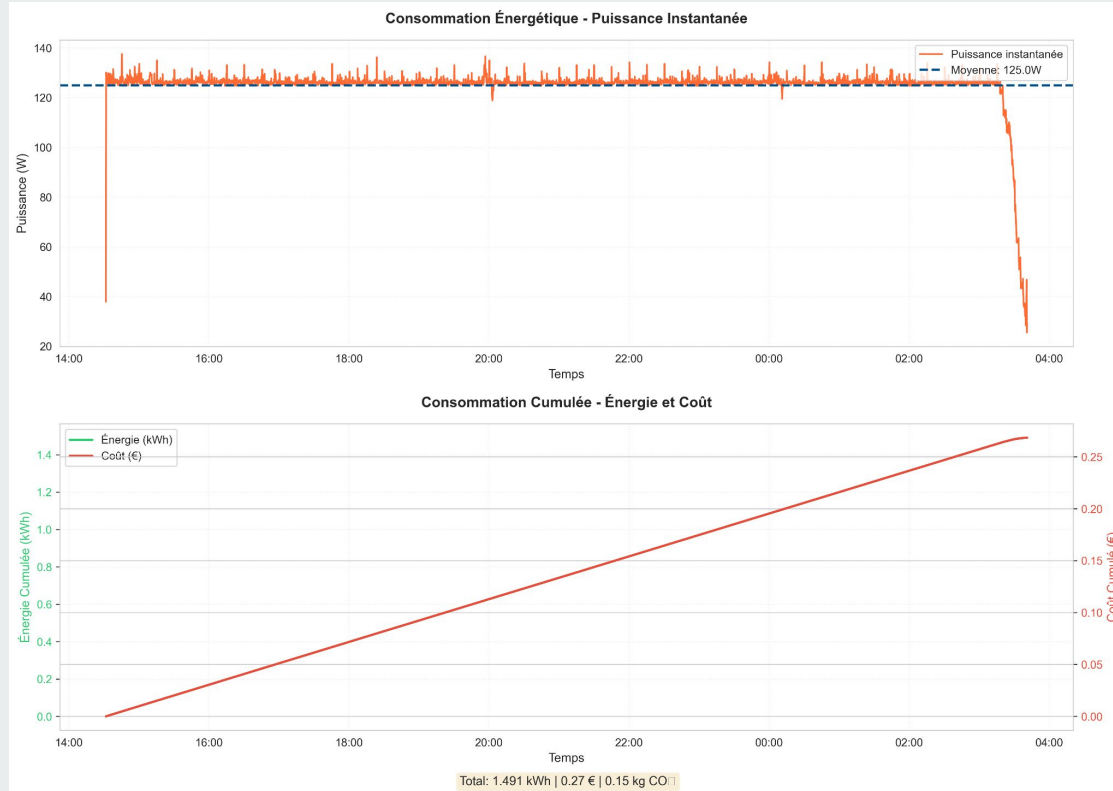
# Benchmarks



# Benchmarks



# Benchmarks



# Benchmarks

- BaseBatch avec  $k=30$ , modèle base

| Model           | tiny           | base | small | medium | large |
|-----------------|----------------|------|-------|--------|-------|
| CPU rate        | $\simeq 90 \%$ |      |       |        |       |
| Memory (GB)     | 22.5           | 27.5 | 48    | 115    | 192   |
| Throughput (Th) | 39.8           | 30.1 | 6.5   | 4.5    | 2     |

# Benchmarks

- BaseBatch avec  $k=30$ , modèle small

| Model           | tiny           | base | small | medium | large |
|-----------------|----------------|------|-------|--------|-------|
| CPU rate        | $\simeq 90 \%$ |      |       |        |       |
| Memory (GB)     | 22.5           | 27.5 | 48    | 115    | 192   |
| Throughput (Th) | 39.8           | 30.1 | 6.5   | 4.5    | 2     |





# Résultats

## Résultats



- Amélioration du throughput global de transcription.
- Th convaincant sur BaseMod
- La pipeline batch a été validée sur les modèles tiny, base et small, dans des conditions représentatives d'un usage réel.



# Conclusion et perspectives

## Conclusion



- Mise en place d'une pipeline de transcription exploitable, robuste et automatisée.
- Les tests sur les modèles tiny, base et small ont été validés : reste à faire sur medium et large
- La pipeline permet désormais de lancer des campagnes de transcription à grande échelle sur des bases de données réelles, comme la base HelloWorld.

## Perspectives

### SYSTRAN/faster- whisper

Faster Whisper transcription with CTranslate2



47

Contributors



9k

Used by



227

Discussions



21k

Stars



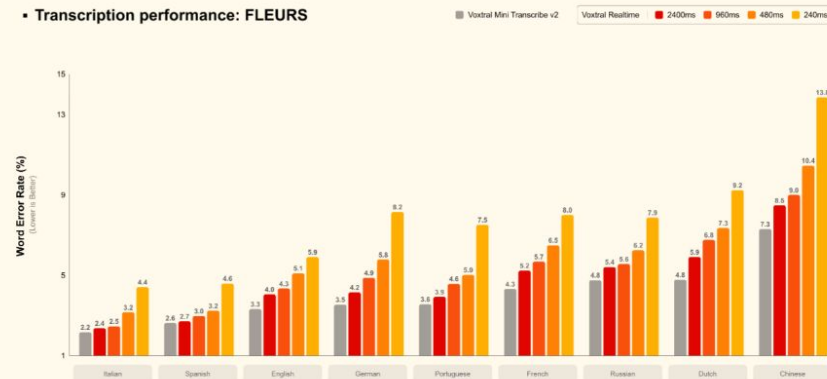
2k

Forks



faster-whisper

#### Transcription performance: FLEURS



Voxtral Transcribe V2



# Questions