

# PROJET STATION TV



## CAHIER DES CHARGES

SUIVI DES MODIFICATIONS DU DOCUMENT			
Version	Date	Validation	Commentaire
1.0	08/10/2025	Frédéric Chauvin	Rédaction initiale

REDACTEUR : Dorian BRISSON	VALIDATEUR : Frédéric CHAUVIN
CLIENT : Polytech Tours	ENCADRANT : Mathieu DELALANDRE

## OBJET DU DOCUMENT

Le présent document a pour objectif de définir le cahier des charges fonctionnel et technique du projet « Transcription audio haute performance sur Station TV », mené dans le cadre du module Projet de Recherche et d'Innovation (PRI) de 5<sup>e</sup> année à Polytech Tours (filiale ISIE). Ce projet s'inscrit dans la continuité du programme de recherche Station TV, initié au LIFAT (Laboratoire d'Informatique Fondamentale et Appliquée de Tours). Il vise à optimiser la chaîne de transcription automatique audio vers texte (Speech-To-Text, STT) basée sur l'outil Whisper (OpenAI) afin de permettre une transcription rapide et scalable des flux télévisuels issus de la TNT.

Le projet comprend une phase de mise à niveau matérielle (installation de 256 Go de mémoire ECC sur la station Dell 5820), une phase de benchmark et d'optimisation logicielle, ainsi qu'une phase d'industrialisation du pipeline dans l'infrastructure Station TV.

## FONCTIONNEMENT D SIR 

Le syst me d velopp  doit permettre la transcription automatique de flux audio complexes issus des cha nes de t l vision captur e par la Station TV.

Le fonctionnement attendu suit les  tapes suivantes :

1. Initialisation du syst me :
  - Lancement des scripts de traitement et chargement des mod les Whisper.
  - V rification automatique de la configuration mat rielle (RAM, CPU, stockage).
2. Pr paration des donn es audio :
  - Lecture des fichiers audio captur s (format MP3/256 kbps).
  - D coupage automatique en lots (batches) selon un protocole temporel d fini.
3. Transcription :
  - S lection du mod le Whisper (small, medium, large).
  - Ex cution parall le multi-process (jusqu'  36 threads sur CPU Xeon W-2295).
  - Production des transcriptions textuelles au format .txt ou .csv.
4.  valuation et m triques QoS :
  - Mesure du temps de traitement, du d bit (throughput), de l'utilisation CPU/RAM.
  - Comparaison de la qualit  linguistique (taux d'erreur WER, stabilit  des segments).
5. Analyse et visualisation :
  - G n ration automatique de rapports et graphiques de performance.
  - Export des r sultats et int gration dans le pipeline Station TV.
6. Maintenance et  volutivit  :
  - Documentation compl te (installation, scripts, r sultats).

## L'EXISTANT

### Infrastructure matérielle :

- Machine Dell Precision 5820 – Xeon W-2295 (18 cœurs / 36 threads).
- 64 Go DDR4 ECC installés avant mise à niveau.
- Station complémentaire Dell PowerEdge T640 (80 threads).
- Stockage total : 38 To (RAID interne) + 190 To externe.

### Infrastructure logicielle :

- OS : Windows Server 2022 / Ubuntu 22.04.
- Librairies : Whisper (OpenAI), Torch, FFmpeg, NumPy, Pandas.
- Scripts existants : BasicTestWhisper.py, RunBatchWhisper.py, ComputeQoS.py.

### Bases de données audio :

- Hello World : 588 heures audio issues de 7 chaînes TNT.
- Bases synthétiques de test : baseMod, baseBatch, baseMtr.

### Constat :

- Saturation mémoire sur les modèles medium et large.
- Trashing CPU lors d'exécutions parallèles intensives.
- Absence de métriques QoS centralisées.
- RAM insuffisante pour les transcriptions longues à haute qualité.

## CAHIER DES CHARGES

### Objectifs principaux

- Permettre l'installation et la validation de 256 Go de RAM ECC sur la Dell 5820.
- Garantir la compatibilité BIOS, la stabilité et la détection quad-channel.
- Réaliser des tests de performance Whisper sur les modèles *small*, *medium* et *large*.
- Mettre en place un système d'évaluation QoS automatisé.
- Industrialiser le pipeline de transcription audio dans l'écosystème Station TV.

### Fonctionnel

#### Matériel :

- Installation supervisée par la DSI Polytech.
- Validation complète post-installation (stress test RAM/CPU).

#### Logiciel :

- Script principal capable de transcrire des fichiers de plusieurs heures.
- Gestion des fichiers batch/run.
- Sauvegarde automatique des résultats et logs.

#### Performance :

- Objectif : traitement de 588h d'audio en < 12h (x50 temps réel).
- Débit minimal requis : 5× pour *small*, 1× pour *medium*, 0.5× pour *large*.
- Mémoire utilisée :  $\leq 240$  Go (sans swapping).

#### Qualité de transcription :

- Word Error Rate (WER)  $\leq 10$  % sur segments clairs.
- Absence de coupures ou erreurs de segmentation.