

VIDEO ANALYZER AI

*LA VISION INTELLIGENTE QUI
TRANSFORME VOS VIDÉOS EN
DÉCISIONS UTILES.*

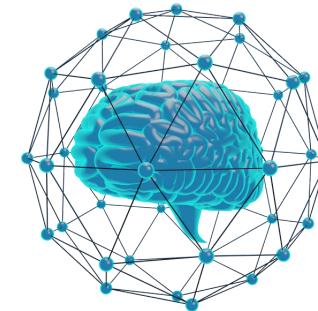
SocialPulse MONASTIR

**L'intelligence émotionnelle
au service de la ville.**

Analyse des émotions, détection
d'événements en temps réel



Tech Nexus



Khaldi Mohammed



Ghassen Belgacem



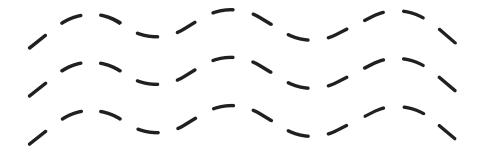
Basdouri Fares



Hmmercha Sarra



Mouissaoui Leila



Plan



1

Problématique

2

Solution envisagée

3

Analyse Concurrentielle

4

Réalisation du Projet

5

Conclusion & Perspectives



Problématique

Le Défi de l'Infobésité Vidéo

- Volume massif : Des heures de vidéos (surveillance, médias, cours) sont générées chaque jour.
- Traitement manuel impossible : Visionner une vidéo prend du temps.
- Notre Objectif : Créer un outil capable de "regarder" la vidéo à notre place et de nous dire de quoi elle parle en quelques secondes.

Solution envisagée

SocialPulse transforme une vidéo brute en métadonnées intelligentes grâce à un pipeline multimodal :

- **Audio** : sens & intention
- **Vision** : contexte & visuel → Une compréhension complète et exploitable du contenu.



SocialPulse
MONASTIR

**L'intelligence émotionnelle
au service de la ville.**

Analyse des émotions, détection
d'événements en temps réel



Analyse des émotions, détection d'événements en temps réel

Solution envisagée

2. Originalité

- **Privacy First** : tout le traitement s'exécute localement, aucune donnée envoyée au Cloud.
- **Arabe optimisé** : gestion native du RTL et des spécificités linguistiques.



L'intelligence émotionnelle
au service de la ville.

Analyse des émotions, détection
d'événements en temps réel



Analyse des émotions, détection d'événements en temps réel

Solution envisagée

3-Valeur Ajoutée :

- Résumer 1h de vidéo en 2 minutes.
- Analyse robuste malgré les langues (Ar/Fr/En).

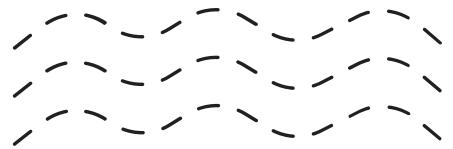


SocialPulse
MONASTIR

**L'intelligence émotionnelle
au service de la ville.**

Analyse des émotions, détection
d'événements en temps réel





Analyse Concurrentielle

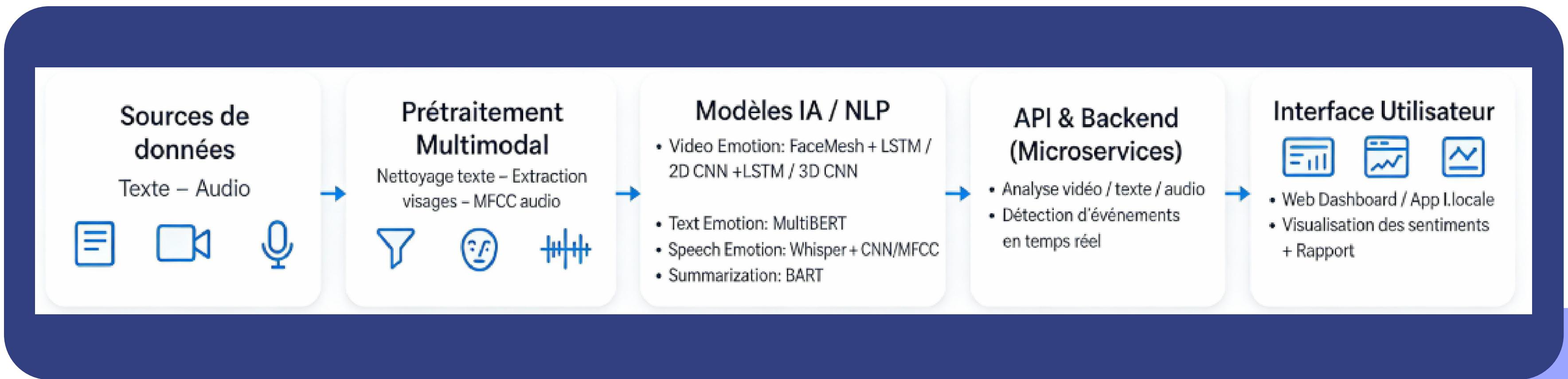


Critères	Solutions Cloud (AWS / Azure)	Notre Solution – SocialPulse Monastir
Confidentialité	✗ Faible — traitement sur serveurs US	✓ Souveraineté totale — traitement 100% local
Coût	✗ Très élevé — facturation par minute	✓ Gratuit & Open Source
Analyse avancée	✓ Analyse sémantique puissante	✓ Analyse enrichie : émotions + sujets + résumé contextuel
Support de l'Arabe	⚠ Standard, peu flexible	✓ Optimisé RTL : adapté au dialecte tunisien
Complexité	✗ Déploiement difficile — experts DevOps nécessaires	✓ Plug & Play : installation directe, interface intuitive
Public cible	Grandes multinationales	✗ Étudiants, chercheurs, médias locaux, PME tunisiennes

Réalisation du Projet



Architecture générale



Comparatif Modèles Vidéo:

FaceMesh + LSTM Approche Géométrique

- 468 points clés 3D
- Précision : 82%
- Latence : Faible
- Ressources : CPU OK

2D CNN + LSTM Spatio-Temporel Hybride

- Features spatiales + temporelles
- Précision : **92.6%**
- Latence : Moyenne

RETIENUE

3D CNN Volumétrique Pur

- Traitement simultané
- Précision : 93.1%
- Latence : Élevée
- Ressources : GPU Heavy



Audio Text Modèles :

Modèle Texte : MultiBERT

- Pré-entraîné 104 langues
- Fine-tuné AR/FR/EN
- 8 classes d'émotion
- Précision : 91.5%

Modèle Audio : Whisper + CNN

- Hybride Dual-Path
- Transcription + analyse tonale
- Robustesse bruit/accents
- Features : MFCC + Mel



Audio Text Modèles :

mBERT : F1-Score par Langue
8 Classes d'Émotion

Langue	F1-Score
Anglais	92%
Français	93%
Arabe	89%

Robustesse exceptionnelle. Légère baisse sur l'Arabe due à la complexité dialectale.

Audio Hybride (Whisper + CNN)
Fusion Multimodale

Approche	Accuracy	Avantage
Whisper seul	77%	Transcription
Hybride	85%	Robustesse

Avantage Hybride

- Robustesse au bruit ambiant
- Analyse tonale + sémantique



Résultats Finales

Métriques par Composant						
Composant	Modèle	Acc.	Préc.	Rap.	F1	
Vidéo	2D CNN+LSTM	92.6%	0.93	0.92	0.92	
Texte	MultiBERT	91.5%	0.91	0.90	0.91	
Audio	Hybrid CNN	85.2%	0.86	0.84	0.84	
Système	Fusion	87.3%	0.88	0.87	0.87	



Liste des services et fonctionnalités

- Module Vidéo (Emotion Recognition) Extraction des landmarks faciaux avec FaceMesh
Reconnaissance d'émotions via notre modèle pré_entraîné .
- Module Texte (NLP Multilingue) : Prétraitement avancé (normalisation arabe, nettoyage, tokenization) , Classification des émotions (8 classes) avec MultiBERT .
- Module Audio (Speech Emotion) : Whisper → transcription → BERT Extraction Fusion des résultats audio + texte pour plus de précision .
- Module Summarization Résumé automatique multilingue via BART Résumé court vidéos .



Technologies utilisées

1

Backend & IA

Python 3.9

TensorFlow / PyTorch

HuggingFace (BERT, mBERT, BART)

MediaPipe (FaceMesh)

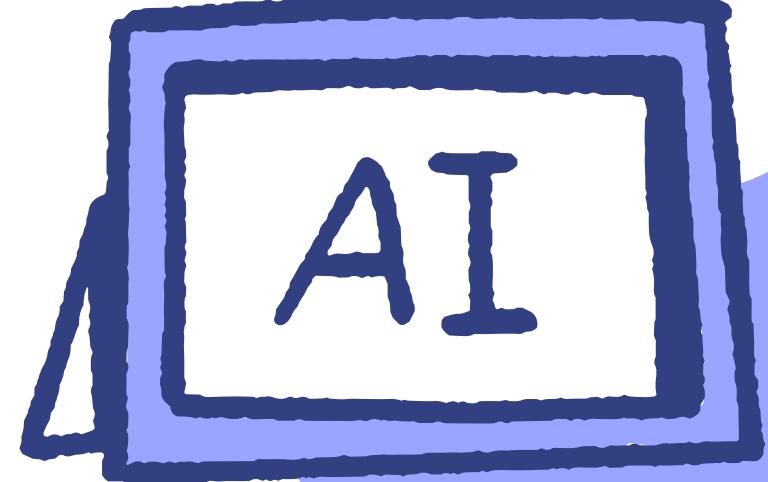
OpenAI Whisper (STT)

2

Frontend

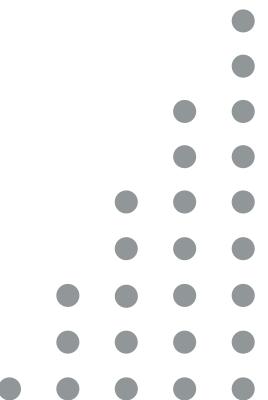
Flask

Dashboard temps réel





Impact Durable



passionnés en Data Science, motivés

Conclusion & Perspectives

En somme, SocialPulse Monastir incarne une vision d'IA souveraine, capable de transformer vos données vidéo en décisions stratégiques tout en garantissant une confidentialité totale. Notre ambition est désormais de perfectionner cette intelligence en maîtrisant les subtilités du dialecte tunisien et en déployant l'analyse en temps réel, pour une ville toujours plus réactive et connectée.





Merci Pour Votre Attention

