



Magic LEMP

# Présentation des technologies

22/03/2023

# Nos activités

CONSEIL

RECHERCHE

FORMATION

L'IA sur mesure et sur site

# L'équipe

- Les fondateurs



**Raphaël-David Lasserri**

**Président Fondateur**

**Docteur en physique nucléaire**

*Ancien élève de l'Ecole Normale Supérieure de Paris-Saclay*



**Marie Chupeau**

**Co-fondatrice, Responsable projet R&D**

**Docteur en physique théorique**

*Ancienne élève de l'Ecole Normale Supérieure de Paris-Saclay  
Agrégée de physique*

**Thomas Epalle**

**Co-fondateur, Chercheur**

**Docteur en combustion**

*Ancien élève de l'Ecole Normale Supérieure de Paris-Saclay  
Ancien élève de l'Ecole CentraleSupélec*



**Antonin Penon**

**Co-fondateur, Directeur technique**

**Data-scientist**

*Ancien élève de l'Ecole Normale Supérieure de Paris-Saclay*



L'équipe Magic LEMP, transdisciplinaire et spécialisée en intelligence artificielle de pointe, compte une quinzaine de profils hautement qualifiés et adaptables (40% de docteurs).

# Ils nous ont fait confiance

INSERTION AU SEIN DE L'ÉCOSYSTÈME FRANÇAIS ET LABELS DE QUALITÉ



NOS PARTENAIRES



Pluralisme



# Nos principaux domaines d'application



SANTE



DEFENSE



POLITIQUES  
PUBLIQUES

Bâtir une IA de confiance souveraine  
pour les applications régaliennes et sensibles



# Domaines d'expertise

# Expertise complète

Une expertise complète :

- Formulation et analyse de la problématique
- Sélection des données pertinentes
- Benchmark de différents modèles d'IA et choix en fonction des contraintes
- R&D de pointe en IA
- Intégration technologique et humaine

Nous investissons du temps de R&D, hors contrats et marchés, pour rester à l'état de l'art de la recherche.

# Domaines d'expertise

Nous avons développé une expertise humaine de pointe sur les technologies suivantes :

- Transcription de la parole (*speech to text*)
- Synthèse vocale (*text to speech*)
- Analyse vidéo
- Analyse de texte (*natural language processing*)
- Grands modèles de langue experts
- Analyse des réseaux sociaux



# Transcription de la parole

## 3 MODÈLES PRÊTS AU DÉPLOIEMENT

- **Modèle rapide :**
  - ⚡ 12 fois plus rapide que le temps réel
  - 🗄️ utile pour les gros volumes
- **Modèle haute précision :**
  - ⚡ 6 fois plus rapide que le temps réel
  - 📌 grande précision sur la ponctuation, les noms propres et la grammaire
- **Modèle phonétique :**
  - ⚡ 12 fois plus rapide que le temps réel
  - 🌐 retranscription en alphabet phonétique international
  - 🔍 utile pour la détection robuste des mots-clés quels que soient la langue ou l'accent

## FONCTIONNALITÉS COMMUNES

- Ré-entraînement sur vos données pour optimiser les performances (*fine-tuning*)
- Transcription multilingue (français, anglais, espagnol, italien, arabe, russe, etc.)
- Segmentation de la transcription par intervenant et identification des intervenants
- Possibilité d'extension du vocabulaire

## POSSIBILITÉS D'INTÉGRATION

- Déploiement on premise en environnement déconnecté/protégé
- Déploiement sur du matériel embarqué
- Accès API en mode SaaS (géré par Magic LEMP et hébergé en France)

# Transcription de la parole

Ne nous croyez pas sur parole, testez notre technologie sur [pluralisme.fr](https://pluralisme.fr)

- analyse des interventions politiques (identification des personnalités, transcription des prises de parole, analyse thématique du discours)
- traitement en temps quasi-réel du flux audiovisuel français \*

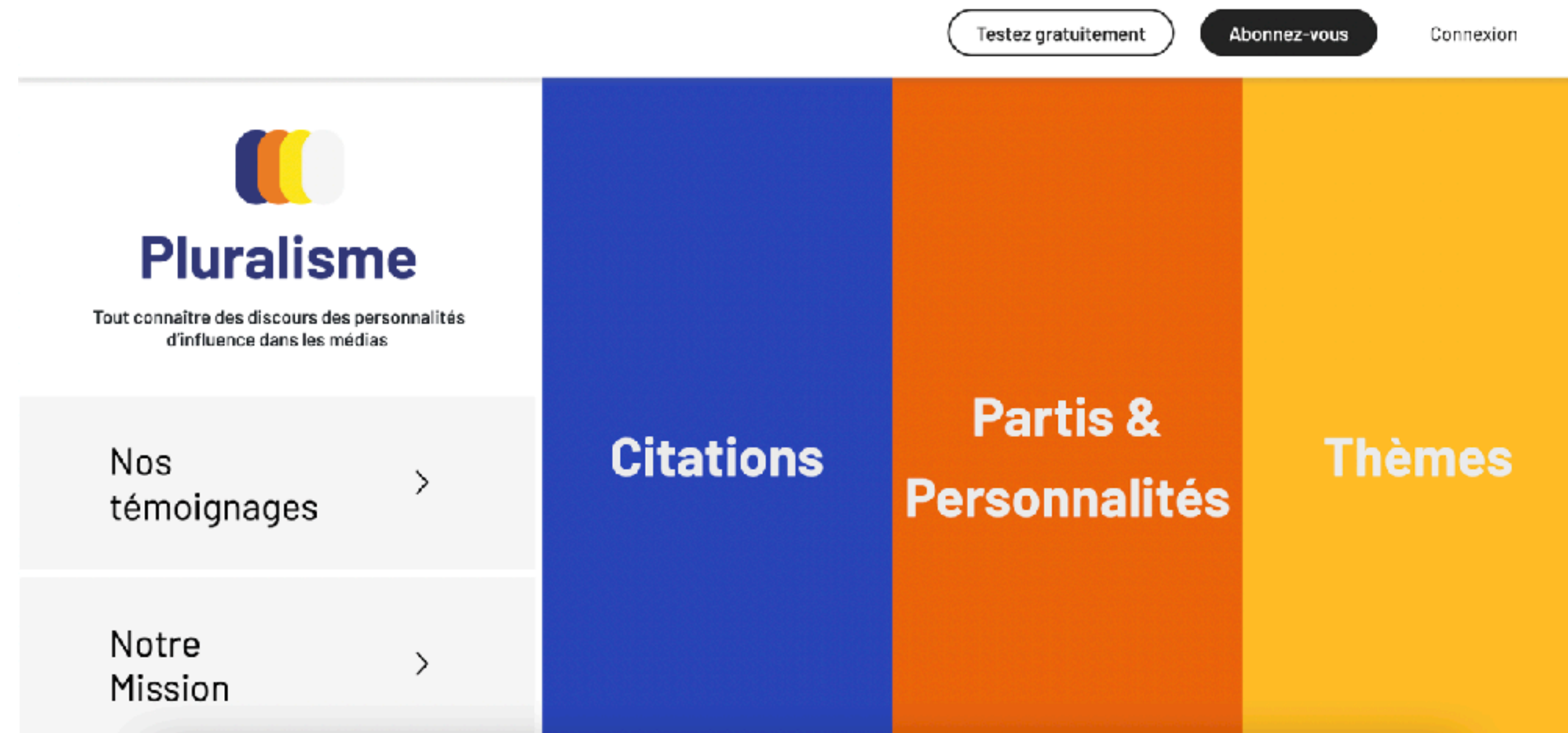


Pluralisme dans la presse :

<https://bit.ly/3no3xwo>

<https://bit.ly/3nvOBwu>

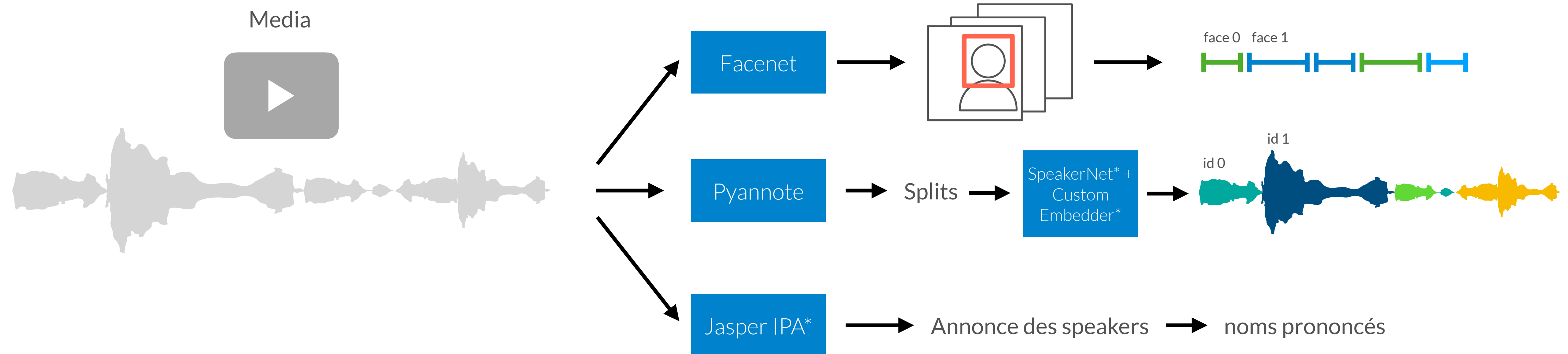
<https://bit.ly/40PSIlp>



\* Volume de données de 11 To/Jour sur une infra Bare métal de 6 Gpu et un système maître esclave

# Transcription de la parole : Pipeline de traitement

## IDENTIFICATION DES INTERVENANTS

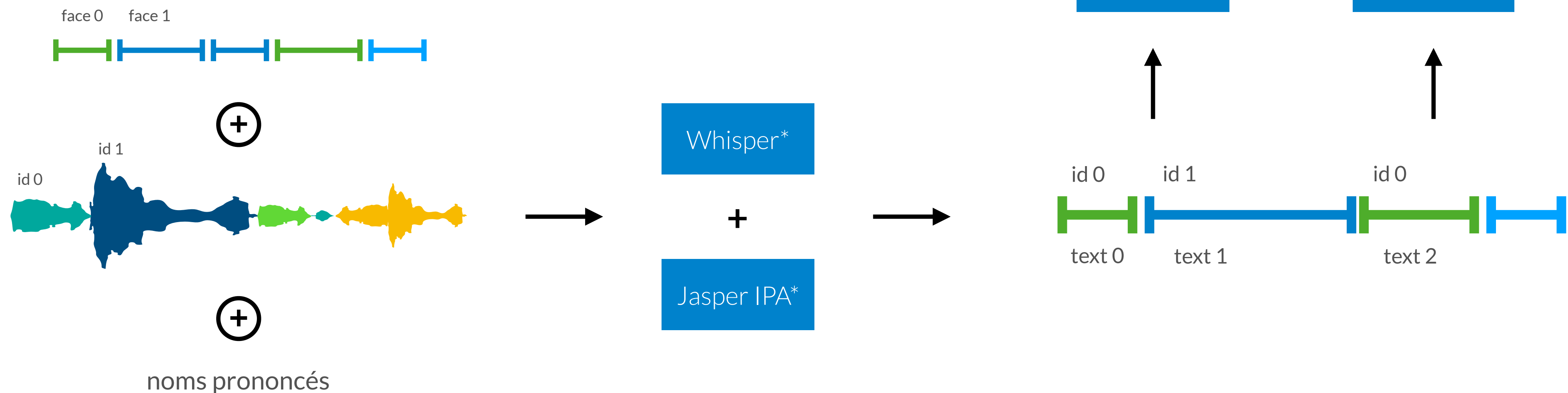


Jasper IPA\* : entraîné à retranscrire l'alphabet phonétique international (IPA) à partir d'un échantillon audio. Entraînement réalisé sur le dataset Librispeech converti en interne en IPA. WER : 13.7 (il est à noter que comme la représentation phonétique n'est pas unique, ici seule la métrique la plus pénalisante a été retenue).

SpeakerNet\* + Customer Embedder\* : entraîné sur des extraits audio des personnalités politiques françaises.

# Transcription de la parole : Pipeline de traitement

## TRANSCRIPTION ET ANALYSE TEXTUELLE



**GPT-NeoX\*** : entraîné sur des jeux de données wikipedia, d'article de presse généraliste et de discours publics (dataset viepublique) en français.

**Camembert\*** : entraîné sur un jeu de données d'articles de presse généraliste en français.  
AUC: 0.90

**Whisper\*** : fine tune sur des jeux de données LibriSpeech et CommonVoices en français. WER 7.6

# Synthèse vocale (*text to speech*)

## FONCTIONNALITÉS

- Génération d'un extrait audio à partir d'un texte
- Génération d'une voix originale à partir de références spécifiées
- Possibilité de spécifier les caractéristiques de la voix désirée (accent, tonalité, émotion,...)

## MODALITÉS D'UTILISATION

- Déploiement on premise d'une voix personnalisée
- Accès API en mode SaaS (géré par Magic LEMP et hébergé en France)

# Analyse automatique de textes (*natural language processing*)

## FONCTIONNALITÉS

- Résumé automatique multilingue\*
- Traduction multilingue
- Extraction automatique d'informations\*\* : lieux, dates, noms propres, etc.
- Recherche documentaire par sujet traité
- Segmentation automatique de document\*\*\*
- Entraînement et fine-tuning de grands modèles de langue standard (LLaMA, GPT, GPT-NeoX, BLOOM)

\* benchmark sur Blue score (peu pertinent cependant dans le cadre de l'application particulière considérée)

\*\* F1 score de 0.91 sur un jeu de données d'arrêts de justice (appel, administratif, cassation, prud'homme, ...) en français.

\*\*\* Précision de 0.85 sur un jeu de données d'arrêts de justice (appel, administratif, cassation, prud'homme, ...) en français.

## MODALITÉS D'UTILISATION

- Déploiement on premise
- Accès API en mode SaaS (géré par Magic LEMP et hébergé en France)

# Grands modèles de langue experts

## POINTS FORTS

- Maîtrise des technologies (Transformers, Memory Transformers, etc.)
- Entraînement sur architecture souveraine (bare metal ou Cloud souverain)
- Capacité à produire une version experte sur un domaine précis

### APPLICATION JURIDIQUE

- R&D en cours avec notre partenaire éditeur juridique [Lexbase](#)
- Fonctionnalités avancées :
  - ▶ Résumé de jurisprudences
  - ▶ Requêtes en langage naturel sur base de données juridique
  - ▶ Réponse à une requête avec citation des sources sur un corpus expert
  - ▶ Caractérisation juridique de faits
  - ▶ Suivi des évolutions législatives

### APPLICATION POLITIQUE

- R&D en cours avec notre filiale [Pluralisme](#)
- Fonctionnalités avancées :
  - ▶ Résumé des interventions politiques
  - ▶ Requêtes complexes en langage naturel
  - ▶ Quantification de la solidité argumentative d'une intervention
  - ▶ Suivi des évolutions de prises de positions politiques

# Analyse video

## FONCTIONNALITÉS

- Décompte de personnes / foules
- Analyse de la proximité de personnes
- Détection d'entrée dans une zone interdite
- Détection de comportement suspect (en cours)
- Floutage automatique des visages
- Description de contenu vidéo et image
- Fine-tuning possible sur des données spécifiques

## MODALITÉS D'UTILISATION

- On premise ou accès API en mode SaaS (géré par Magic LEMP et hébergé en France)
- Optimisé pour l'embarqué
- On-the-fly, sans conservation des données d'entrée



# Suivi des réseaux sociaux

## NOTRE APPROCHE

- Crawling des réseaux sociaux :  
Twitter, Snapchat, Instagram (en cours), TikTok
- Analyse de la propagation d'informations  
(thèse CIFRE en cours)

## FONCTIONNALITÉS

- Système d'alerte sur mots-clés
- Analyse du contenu multimédia :
  - description d'une image ou d'une vidéo
  - transcription des messages audio et vidéo
- Traduction automatique

## MODALITÉS D'UTILISATION

- On premise
- Accès API en mode SaaS (géré par Magic LEMP et hébergé en France)

# Conclusion

## NOS VALEURS

- Expertise de pointe à l'état de l'art
- Trois verticales à enjeux sociétaux : politique, défense, santé
- Construction de systèmes d'IA dignes de confiance : fiabilité, explicabilité, quantification des incertitudes



## MERCI DE VOTRE ATTENTION

Contact :

Raphaël-David Lasserri, président de Magic LEMP

[lasseri@magic-lemp.com](mailto:lasseri@magic-lemp.com)

06 16 26 08 42

[magic-lemp.com](http://magic-lemp.com)