



Bienvenue sur

*data.ina.fr*

*Le site qui donne à voir les données de l'INA*

Dossier de presse  
Octobre 2024

**DATA.INA.FR**

*Inédit*

# UNE NOUVELLE PLATEFORME POUR DÉCRYPTER LA TÉLÉ ET LA RADIO GRÂCE À L'INTELLIGENCE ARTIFICIELLE ET AUX DONNÉES DE L'INA

l'Institut national de l'audiovisuel (INA) dévoile data.ina.fr, **une plateforme inédite qui révolutionne la manière de comprendre les médias audiovisuels français grâce à l'intelligence artificielle.**

Pour la première fois, les données du dépôt légal de l'audiovisuel sont utilisables par tous !



**data.ina.fr permet d'exploiter les données de référence de la télévision et de la radio françaises grâce à des outils d'IA**, dont un algorithme développé par les équipes de recherche de l'INA.

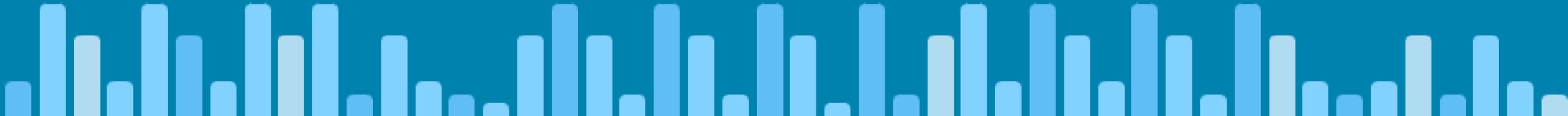
Très accessible grâce à la **datavisualisation**, la plateforme met ainsi en lumière des tendances statistiques majeures des médias et de la société avec une approche à la fois transparente et pédagogique.

# UNE EXPERIENCE ACCESSIBLE POUR TOUS LES PUBLICS

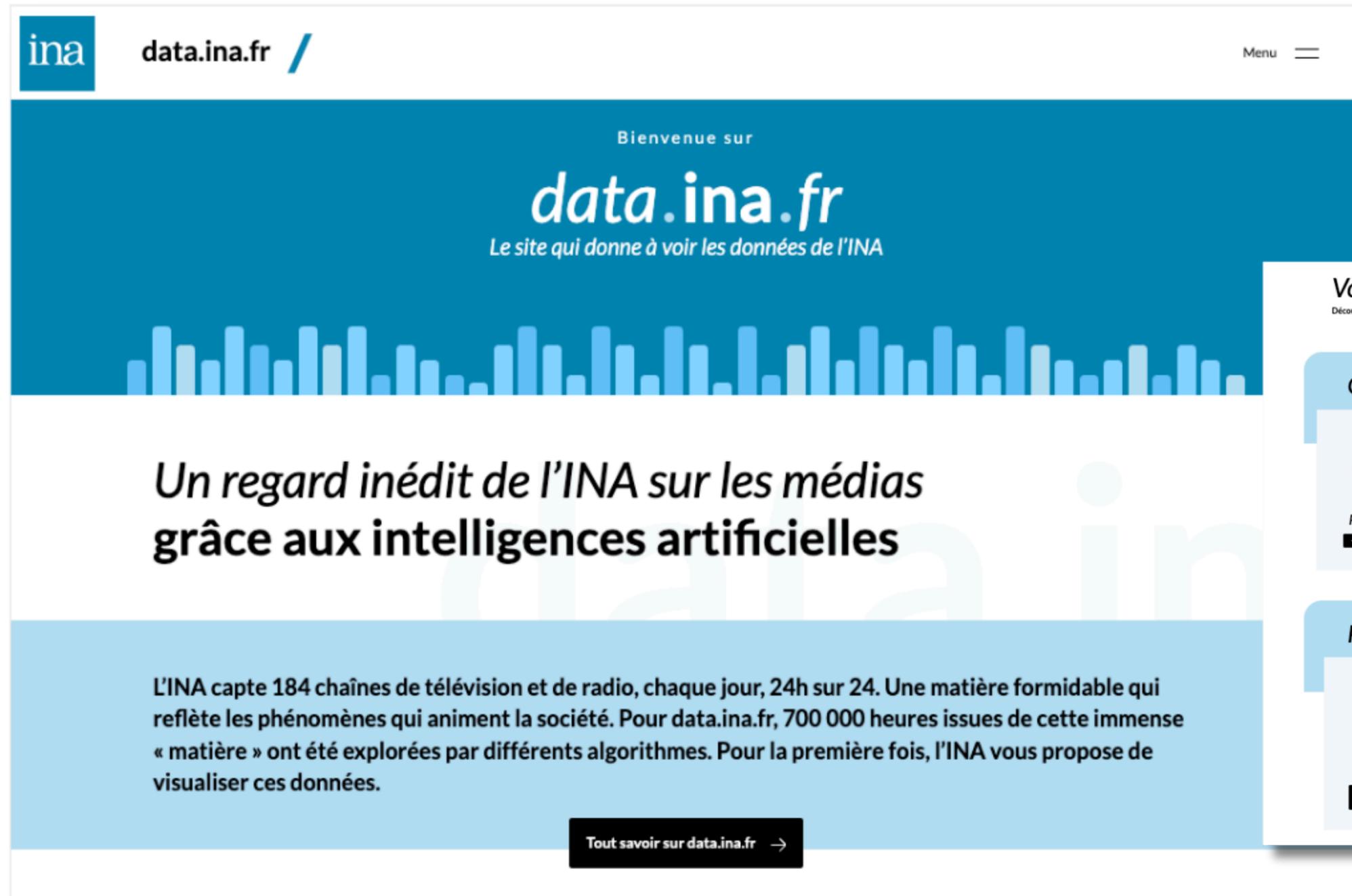
Alimenté par des flux de **20 chaînes de télévision et de radio**, le site permet d'explorer de manière personnalisée **les grandes tendances de présence médiatique**.

A destination du **grand public, journalistes, experts ou chercheurs**, ***data.ina.fr*** permet d'accéder à des informations précieuses à travers **des cartes et graphiques interactifs**.

Par exemple, une étudiante en économie qui souhaite connaître la couverture médiatique du terme "*shrinkflation*" dans les matinales radio, pourra y trouver des éléments de réponses sur une sélection de chaînes qu'elle aura identifiées.



# BIENVENUE SUR DATA.INA.FR !



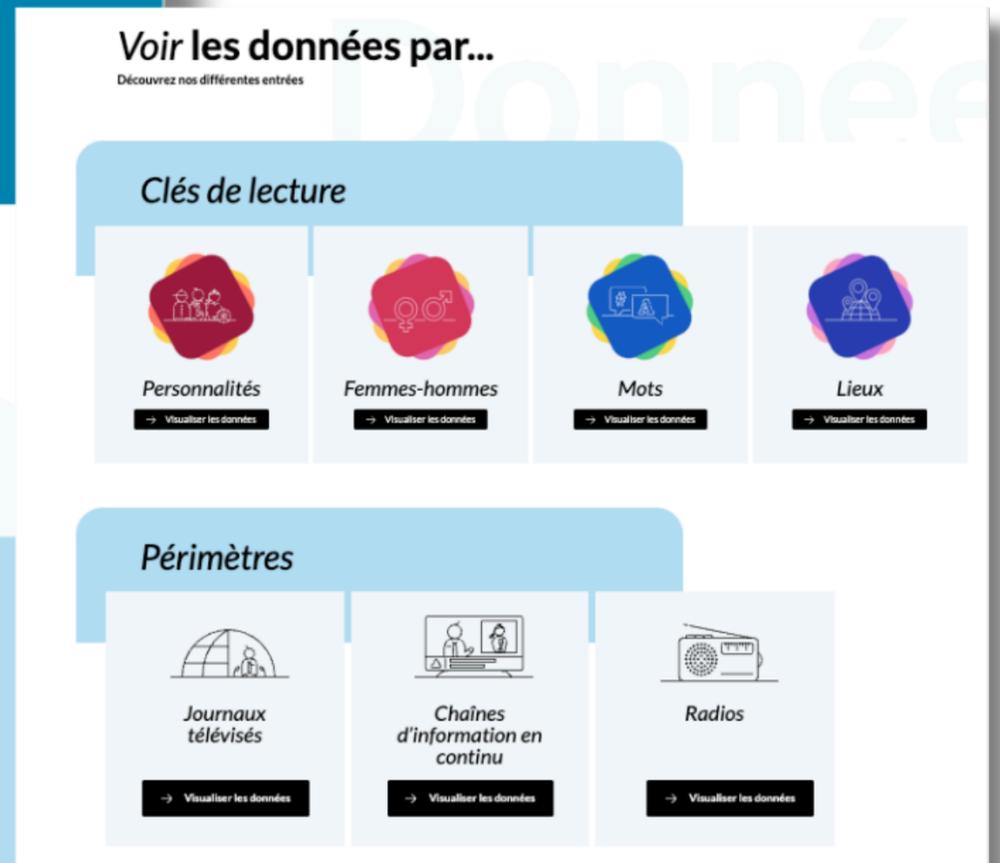
ina data.ina.fr / Menu

Bienvenue sur  
**data.ina.fr**  
Le site qui donne à voir les données de l'INA

*Un regard inédit de l'INA sur les médias  
grâce aux intelligences artificielles*

L'INA capte 184 chaînes de télévision et de radio, chaque jour, 24h sur 24. Une matière formidable qui reflète les phénomènes qui animent la société. Pour data.ina.fr, 700 000 heures issues de cette immense « matière » ont été explorées par différents algorithmes. Pour la première fois, l'INA vous propose de visualiser ces données.

Tout savoir sur data.ina.fr →



**Voir les données par...**  
Découvrez nos différentes entrées

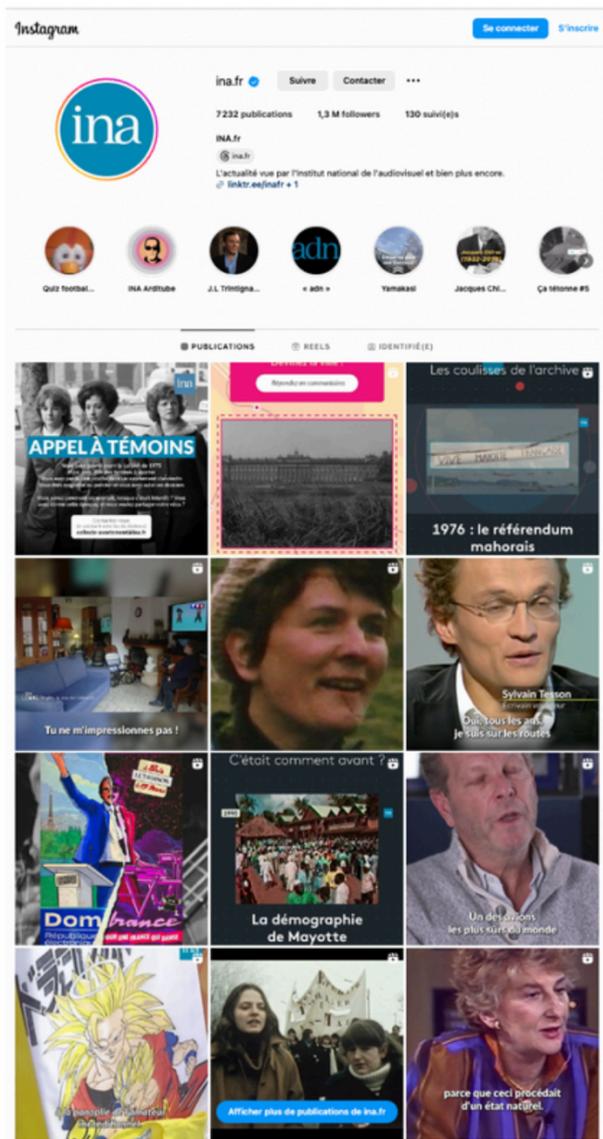
**Clés de lecture**

- Personnalités → Visualiser les données
- Femmes-hommes → Visualiser les données
- Mots → Visualiser les données
- Lieux → Visualiser les données

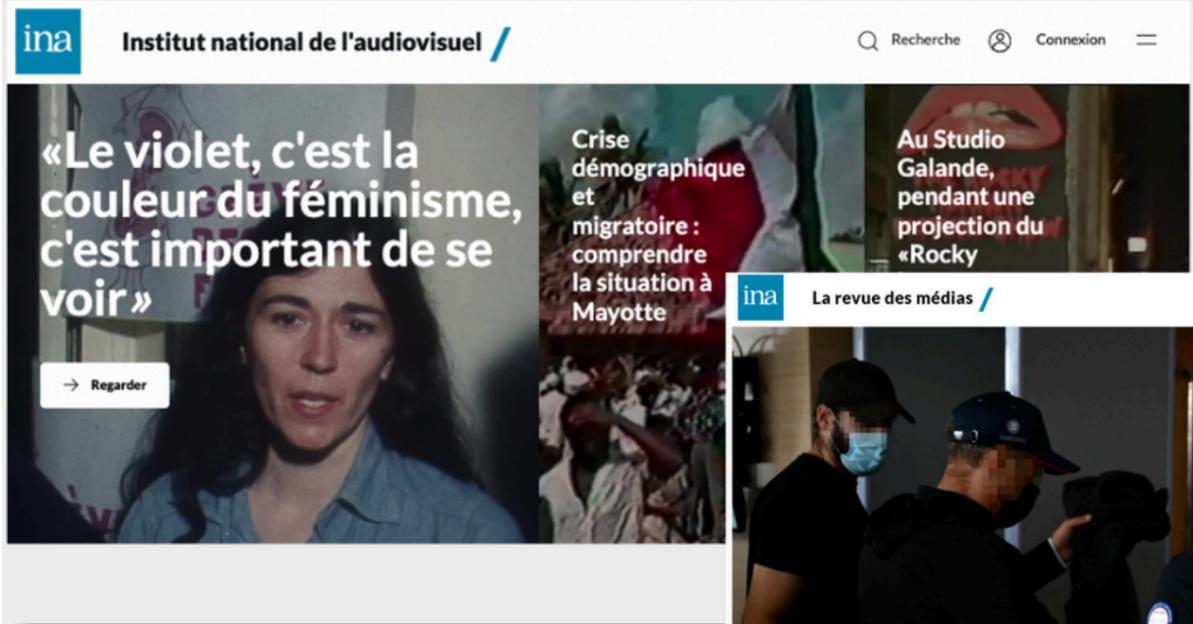
**Périmètres**

- Journaux télévisés → Visualiser les données
- Chaînes d'information en continu → Visualiser les données
- Radios → Visualiser les données

# COMMENT DATA.INA.FR S'INSCRIT DANS LA STRATÉGIE ÉDITORIALE DE L'INA ?



**1,7 milliard**  
de vidéos vues sur  
l'ensemble des  
plateformes\*  
\*par plus de 10 millions  
d'utilisateurs au 31/12/23

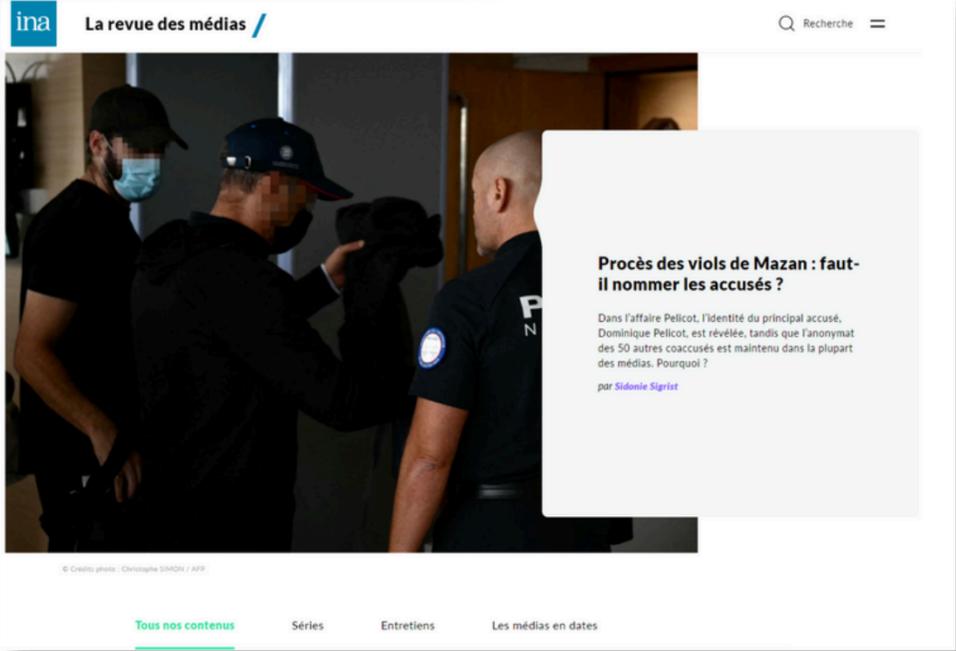


**5,2 M**  
abonnés sur  
nos chaînes  
YouTube

**4 M**  
fans sur  
Facebook

**1,5 M**  
sur  
Instagram

**282 k**  
sur  
Twitter



Source : <https://www.ina.fr/institut-national-audiovisuel/organisation/chiffres-cles>

# UNE EXPERIENCE POUR TOUS LES PUBLICS QUI CHERCHENT À MIEUX COMPRENDRE L'INFORMATION

Que vous soyez amateurs de médias, journalistes, experts, chercheurs ou tout simplement curieux, ce site a été pensé pour le plus grand nombre.

**3**  
*périmètres*

**Périmètres**



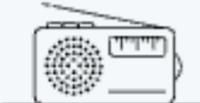
Journaux télévisés

→ Visualiser les données



Chaînes d'information en continu

→ Visualiser les données



Radios

→ Visualiser les données

**4**  
*entrées thématiques*

**Clés de lecture**



Personnalités

→ Visualiser les données



Femmes-hommes

→ Visualiser les données



Mots

→ Visualiser les données



Lieux

→ Visualiser les données

**20**  
*médias*



# UNE EXPERIENCE POUR TOUS LES PUBLICS QUI CHERCHENT À MIEUX COMPRENDRE L'INFORMATION

Des questions **simples et directes** et des réponses entièrement datavisualisées et interactives.

**16** questions grand public

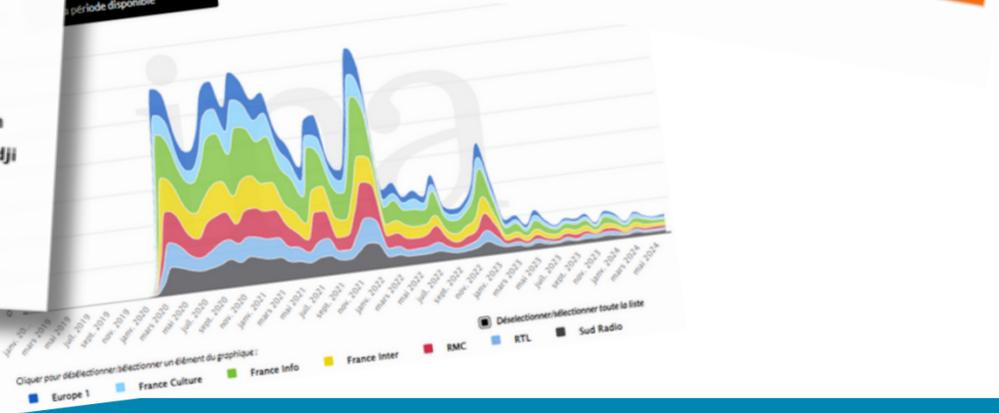
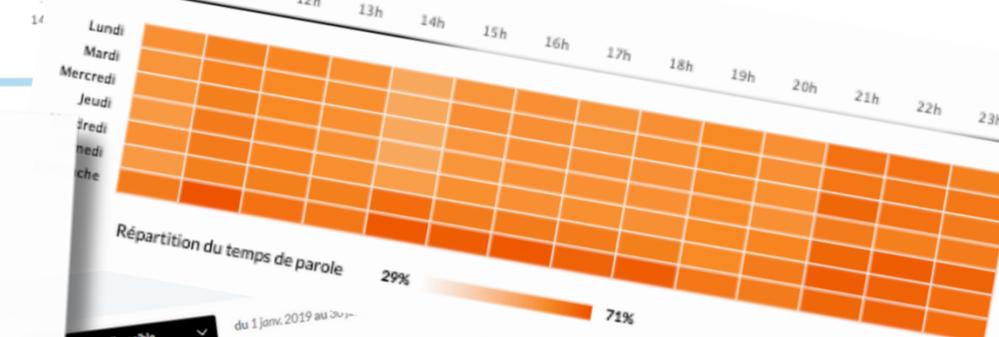
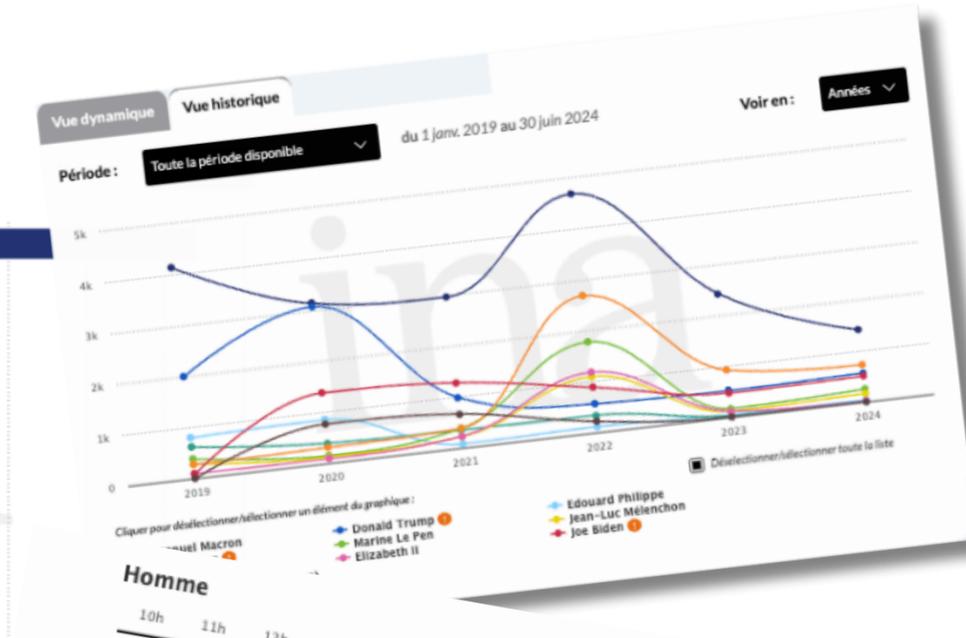
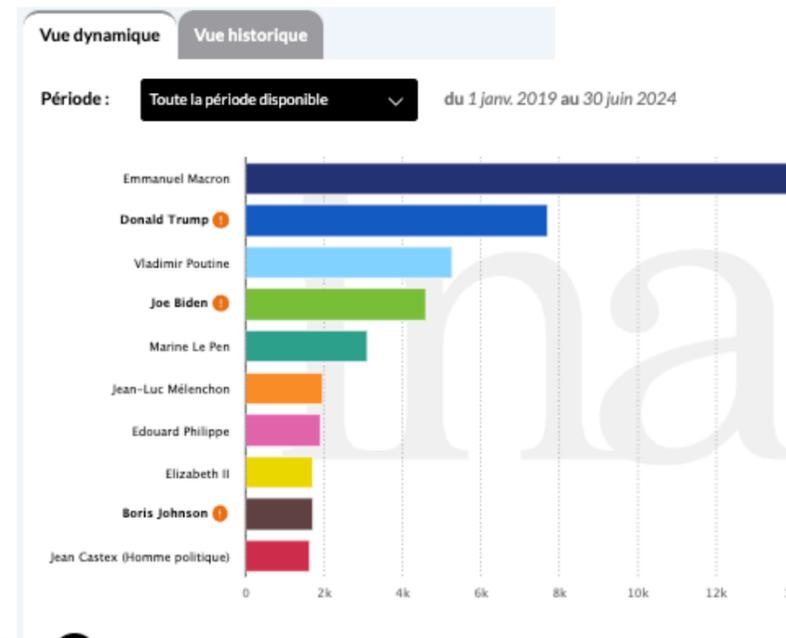
**28** graphiques interactifs

→ Explorez les mots prononcés sur les radios

→ Quelles personnalités mentionne-t-on le plus sur les chaînes d'information en continu ?

→ Comment évolue la répartition du temps de parole entre les femmes et les hommes ?

→ Quels pays mentionne-t-on le plus dans les journaux télévisés ?



# DES RENDEZ-VOUS BIANNUELS

Un lancement avec 5,5 ans d'historique :  
1er janvier 2019 au 30 juin 2024

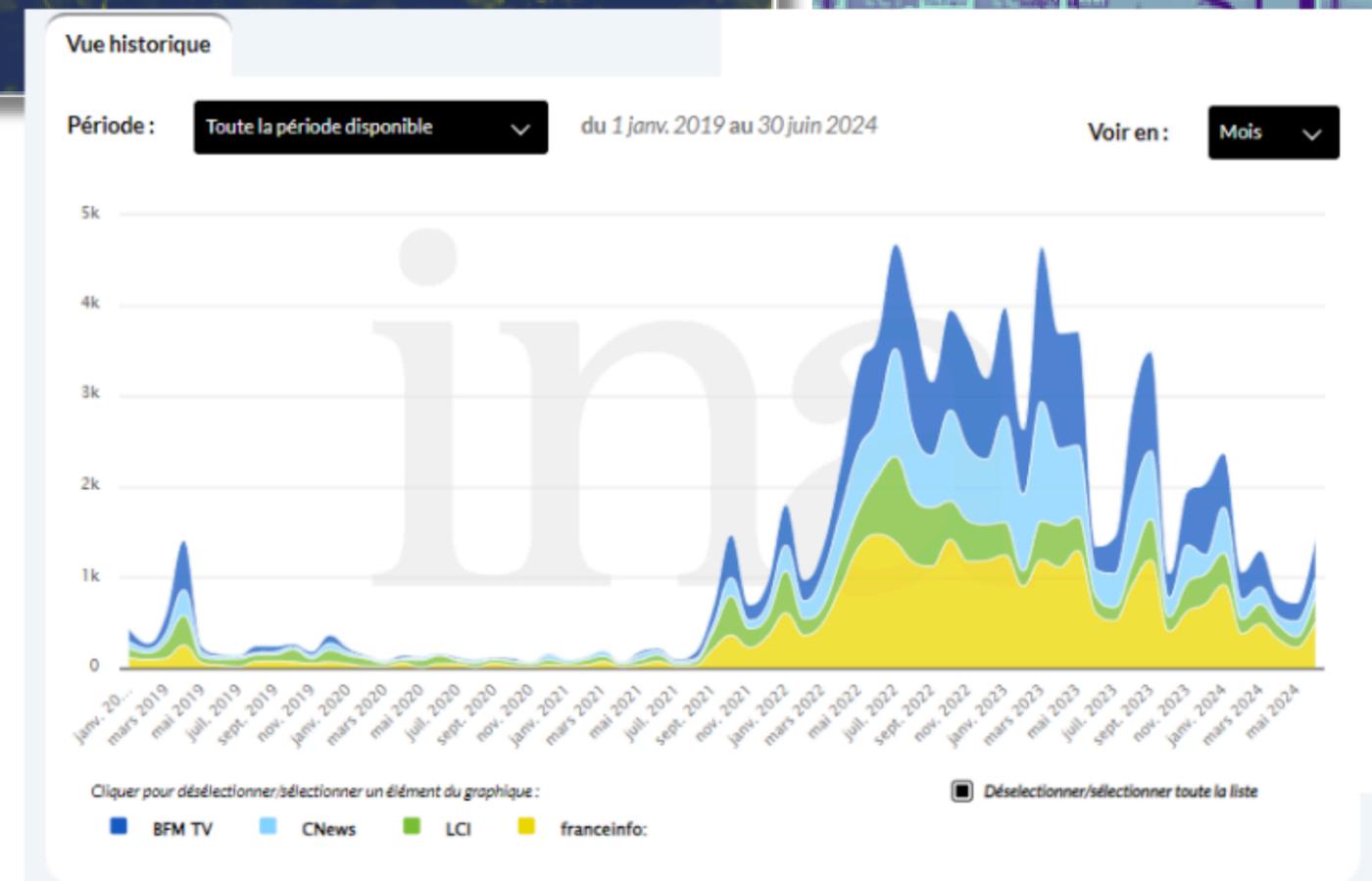
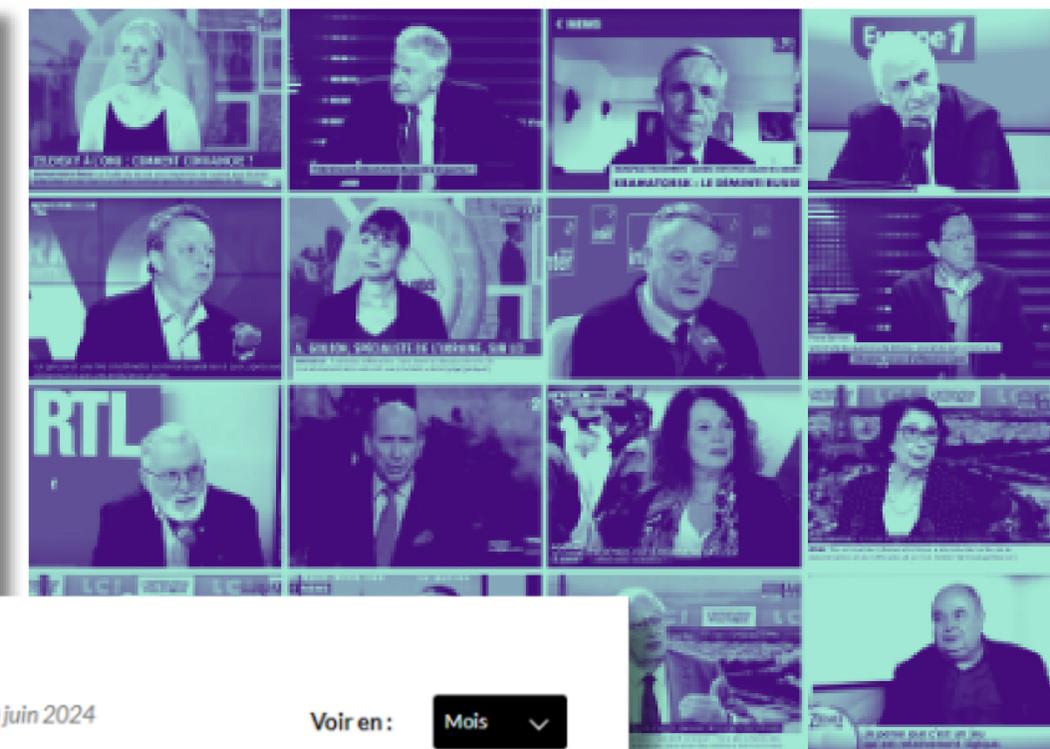
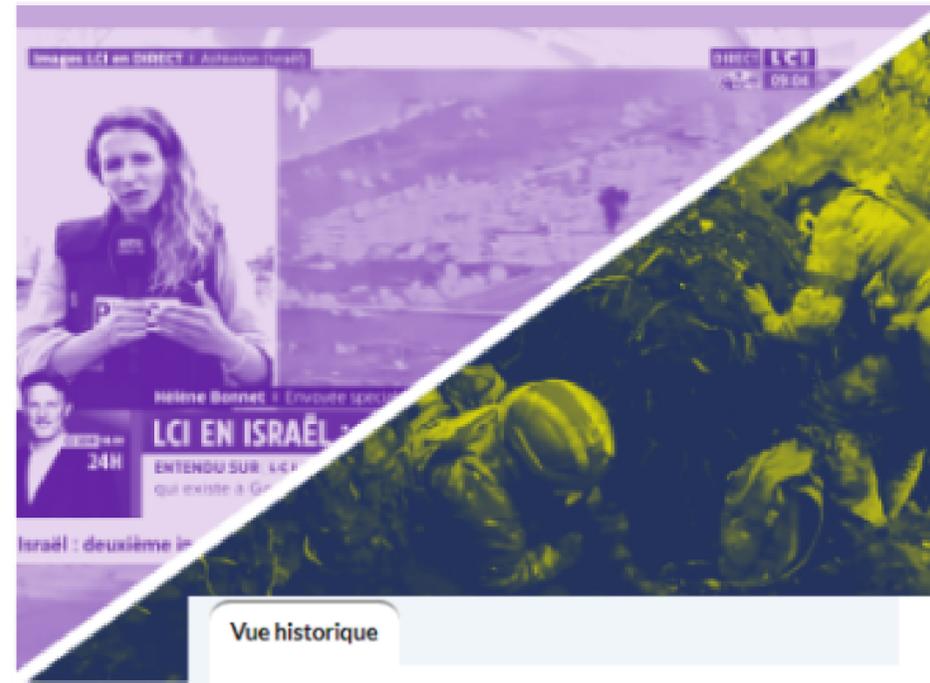
Puis un événement éditorial deux fois par an .  
Objectif = assurer des traitements  
et des contrôles de qualité sur les données  
proposées sur le site.

A chaque mise à jour, le site bénéficiera :

- des données les plus récentes
- d'une profondeur historique toujours plus grande



Une mise à jour  
tous les **6 mois** avec  
de *nouvelles données*

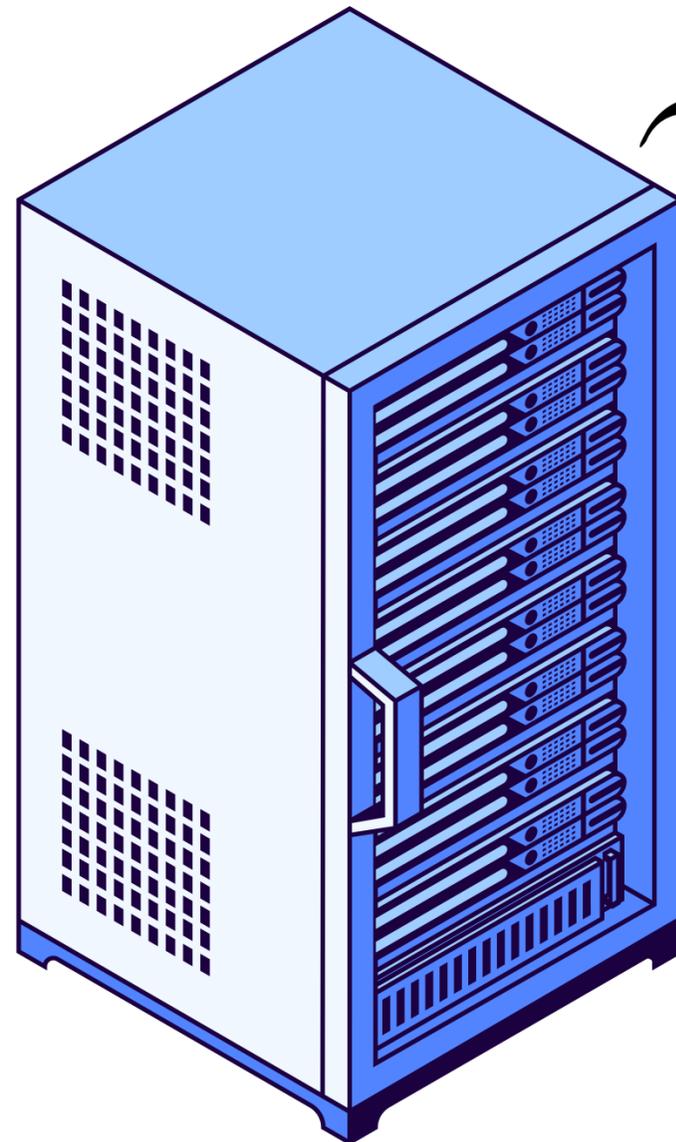


# L'ALLIANCE DE SAVOIR-FAIRE INTERNES

Direction "Data et tech"  
créée en 2021

**27 millions**

d'heures de  
documents TV et radio  
**sont stockés et archivés**  
au 31/12/23



dont une partie a déjà fait l'objet  
de traitements par des outils IA :

- . Transcription
- . Reconnaissance d'entités nommées
- . Classification sonore

**une dizaine**  
*de métiers sollicités*  
(data-analystes, data-scientists,  
data-engineers, architectes d'infra...)



**soit une centaine**  
**de personnes** ayant  
contribué au site

# SOUS LE CAPOT DE DATA.INA

Captation de 184 chaînes

Traitements des médias via des IA

- Transcription **Whisper**
- Extraction d'entités nommées **TextRazor**
- Classification sonore **InaSpeechSegmenter**

Lac de données

Calculs

- Résultats du traitement
- Extraction des entités nommées
  - Transcription
  - Classification sonore

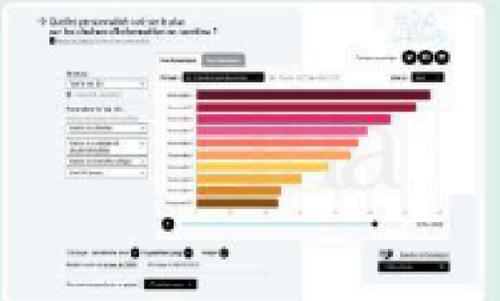
Rapprochement avec le référentiel data.ina.fr

- Personnalités
- Lieux
- Mots et thèmes
- Femmes-hommes

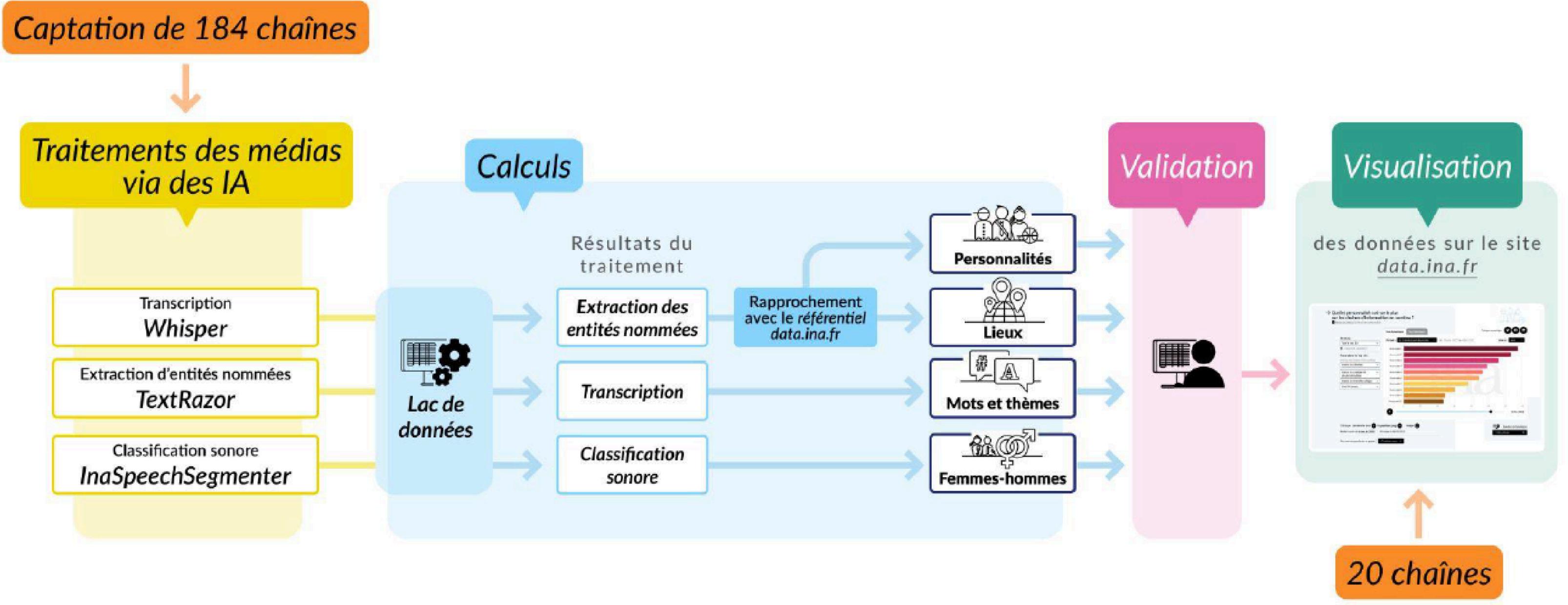
Validation

Visualisation

des données sur le site [data.ina.fr](http://data.ina.fr)



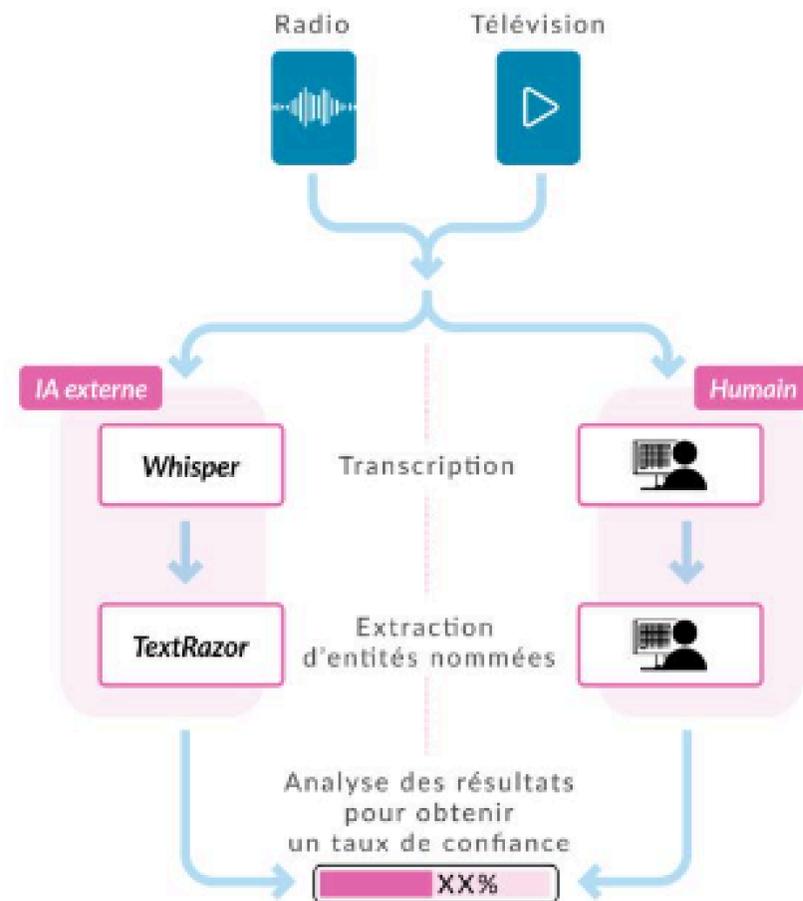
20 chaînes



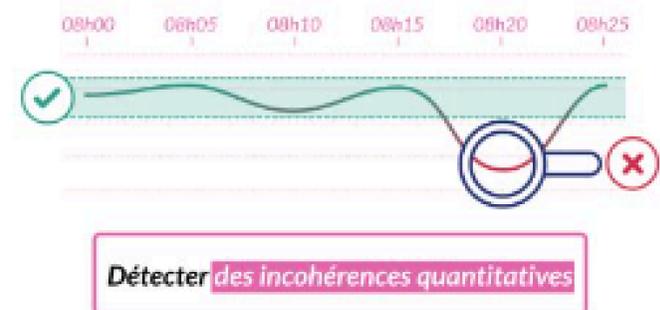
# 3 GRANDES ACTIONS DE CONTRÔLE

## Vérité de terrain sur les IA externes

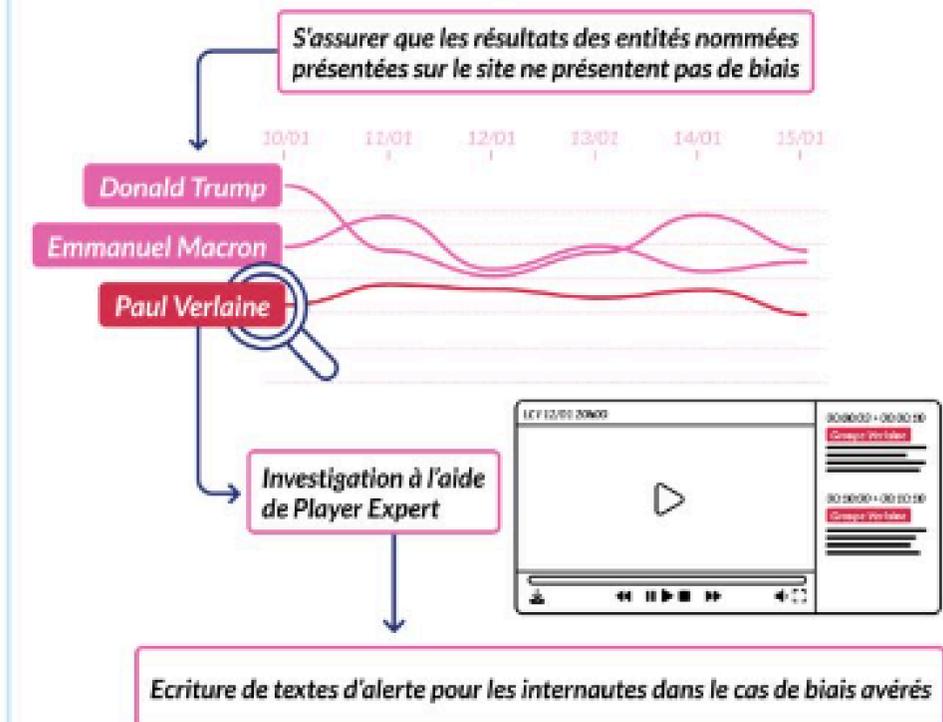
L'objectif est de comparer le traitement IA d'heures de télévision et radio (sélectionnées sur le périmètre data.ina.fr) et le travail humain de transcription et de reconnaissance des entités nommées



## Contrôle des traitements automatiques



## Contrôle de la pertinence

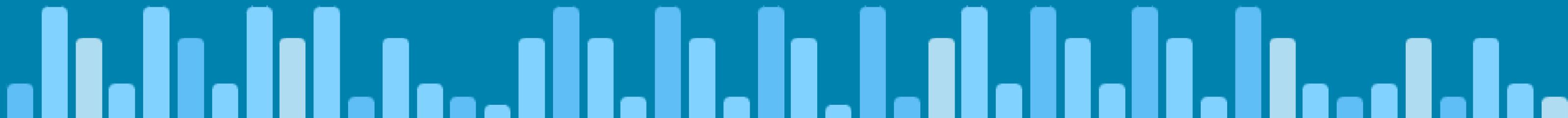


Exemple  
 ● Paul Verlaine !  
 À la suite de notre investigation, les résultats mentionnant Paul Verlaine tendent à indiquer que l'algorithme de reconnaissance des entités nommées confond le poète avec l'entreprise "Groupe Verlaine", présente dans les médias à travers une campagne de publicité. Cela entraîne une sur-représentativité. L'amélioration continue des algorithmes, avec une meilleure...

# Quelques exemples d'utilisation des données

*data.ina.fr*

*Le site qui donne à voir les données de l'INA*



## UN MÉDIA DISTINCTIF QUI INVESTIT L'ANALYSE MÉDIATIQUE SUR LE TEMPS LONG

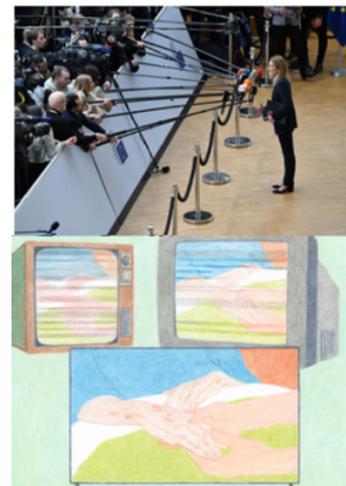
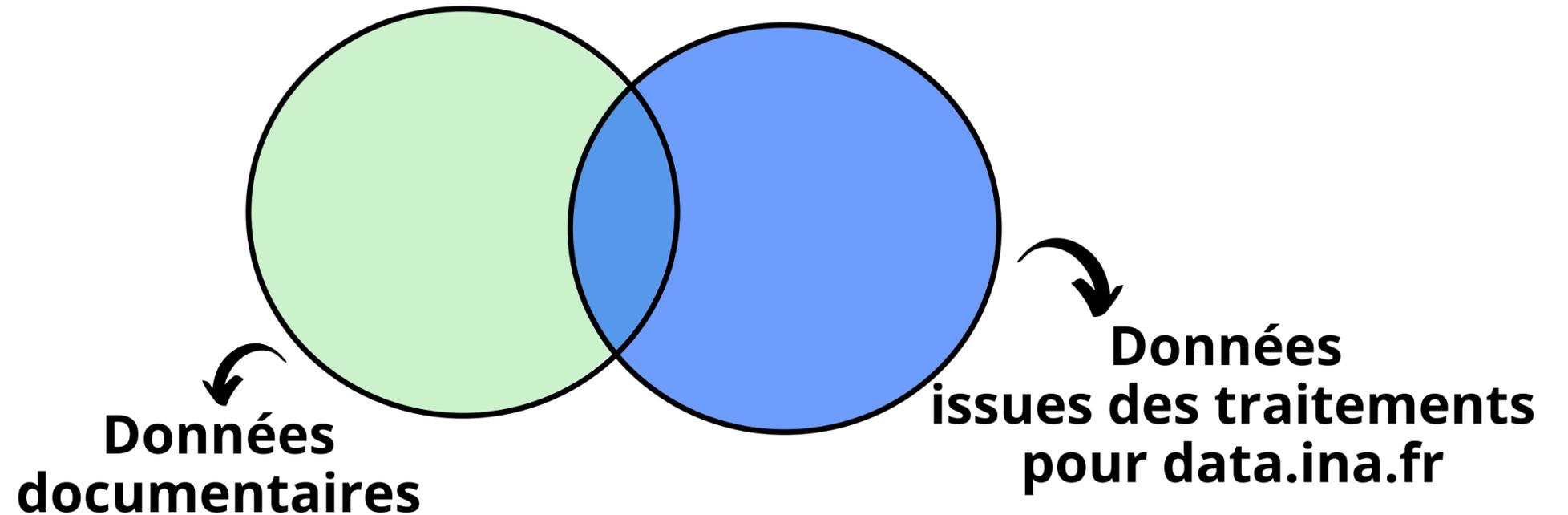
1

Rédactrice en chef  
adjointe en **charge de  
l'éditorialisation des  
données de l'INA**

2

Data-analystes  
sur le site **data.ina.fr**

Près d'une dizaine d'enquêtes  
de datajournalisme parues  
réalisées à partir de :



### Actualité européenne : la grande timidité de nos journaux télévisés

INFOGRAPHIES. Depuis vingt ans, les institutions européennes ont fait l'objet de seulement 3 % de l'offre globale d'information sur les JT du soir. À l'approche du scrutin du 9 juin 2024, *La Revue des médias* vous présente les résultats d'une étude inédite sur la médiatisation des élections européennes.

par [Camille Pettineo](#) - le 23 mai 2024

### La fin de vie : quelle place sur nos écrans depuis trente ans ?

INFOGRAPHIES. La France ouvre un nouveau chapitre législatif sur la fin de vie, à partir de ce 27 mai. Quelle place accordent les médias à ce sujet sensible ? Décryptage en données et en archives de trois temps forts médiatiques.

par [Camille Pettineo](#) - le 27 mai 2024



### Deux ans après l'invasion russe en Ukraine, que retenir de la médiatisation de la guerre ?

INFOGRAPHIES. *La Revue des médias* révèle des données inédites : au cours de la seconde année de conflit, la couverture de la guerre a été divisée presque par quatre dans les JT de 20 heures par rapport à la première année, par presque trois dans les matinales radio et par deux sur les chaînes info.

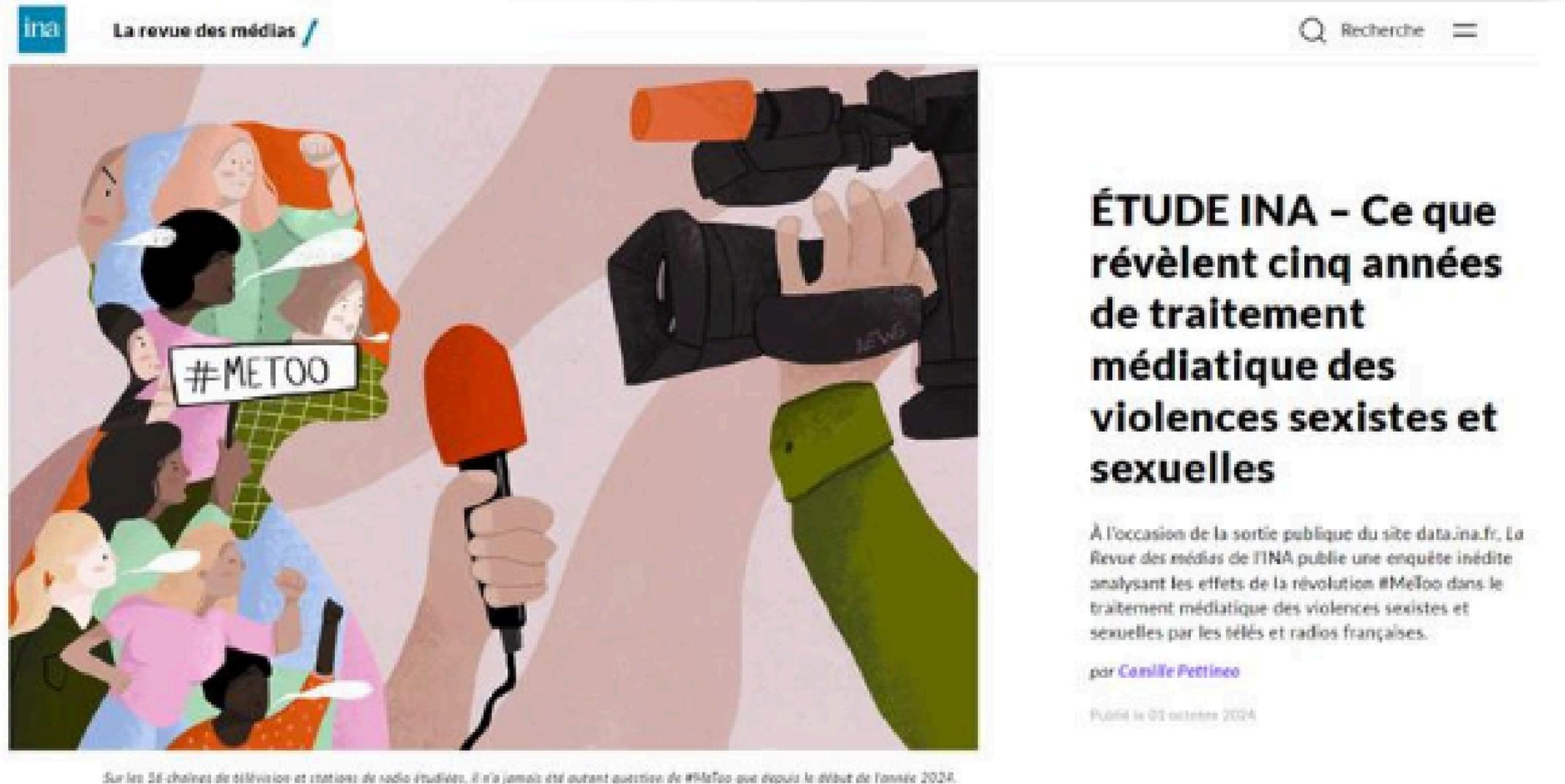
par [Camille Pettineo](#) - le 23 février 2024

### Télévision : pourquoi le séisme au Maroc n'a laissé aucune place aux inondations en Libye

INFOGRAPHIES. Le drame marocain, survenu le 8 septembre 2023, et la tempête qui a frappé la Libye, deux jours plus tard, n'ont pas bénéficié du même traitement médiatique. *La Revue des médias* révèle des données inédites : le séisme a été 17 fois plus traité sur les chaînes d'information en continu et près de sept fois plus dans les journaux télévisés.

par [Camille Pettineo](#) - le 16 novembre 2023

# UN APERÇU DE LA PROCHAINE ENQUÊTE À PARAÎTRE



The image shows a screenshot of a webpage from INA (Institut National de l'Audiovisuel). The page features a header with the INA logo and the text 'La revue des médias /'. A search bar with the word 'Recherche' is visible in the top right corner. The main content area is dominated by a large, colorful illustration on the left. This illustration depicts a group of diverse people, including women and men of various ethnicities, some holding up a sign that reads '#METOO'. In the foreground, a hand holds a red microphone, and another hand holds a professional video camera. The background is a soft, light pinkish-purple. To the right of the illustration, the article title is displayed in a large, bold, black font: 'ÉTUDE INA - Ce que révèlent cinq années de traitement médiatique des violences sexistes et sexuelles'. Below the title, a short paragraph in a smaller font provides context: 'À l'occasion de la sortie publique du site data.ina.fr, La Revue des médias de l'INA publie une enquête inédite analysant les effets de la révolution #MeToo dans le traitement médiatique des violences sexistes et sexuelles par les télévisions et radios françaises.' The author's name, 'par Camille Pettinato', is listed below the paragraph. At the bottom of the article preview, it says 'Publié le 03 octobre 2024'. At the very bottom of the screenshot, there is a small line of text: 'Sur les 34 chaînes de télévision et stations de radio étudiées, il n'a jamais été autant question de #MeToo que depuis le début de l'année 2024.'

<https://larevuedesmedias.ina.fr/node/2404>



Rendez-vous en janvier 2025  
pour de nouveaux produits d'études



[data.ina.fr](https://data.ina.fr) /

### *Contacts Presse*

#### **INA**

Christophe Chuchu  
06 43 86 05 41  
cchuchu@ina.fr

#### **Majorelle**

Guillaume Ancelin  
06 73 55 09 39  
gancelin@majorelle-pr.fr