

Datajournalisme :

comment l'IA éclaire notre regard sur l'actualité ?

Qui suis-je ?



Camille Pettineo

Rédactrice en chef adjointe en charge de la data
à l'INA & responsable éditoriale de data.ina.fr

- Data-journaliste spécialisée en IA



The Pulitzer Prizes

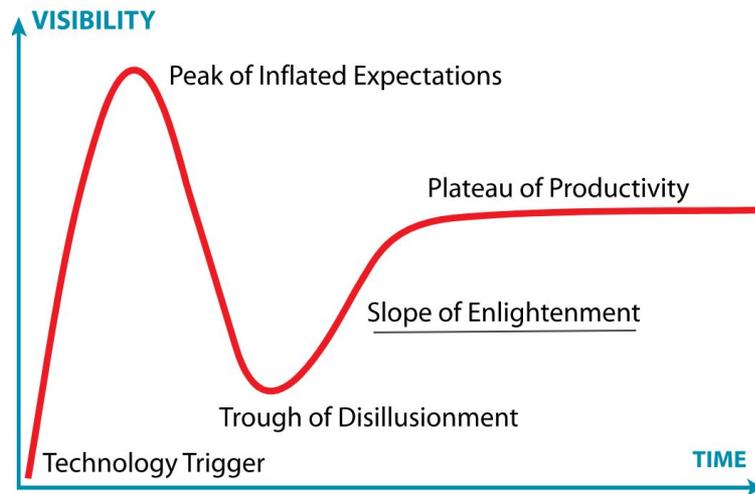
Honoring Excellence in Journalism and the Arts

IA & journalisme : Sur le (long) chemin de l'éclaircissement

L'étude "[Large Language Models, Small Labor Market Effects](#)" (avril 2025)*, menée par des chercheurs de l'Université de Copenhague et de l'Université de Chicago révèle qu'au Danemark:

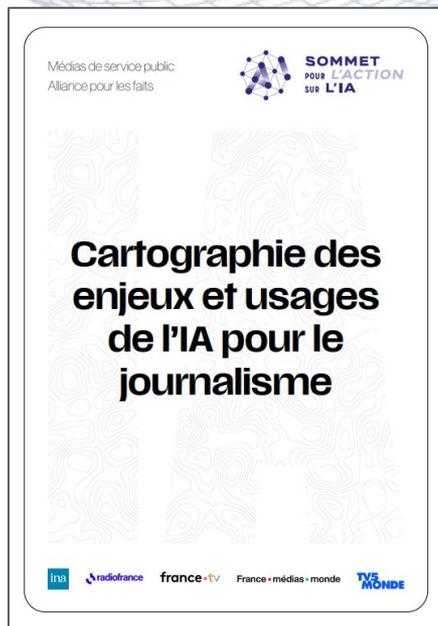
- > l'IA n'a pas eu d'impact sur les revenus ou les heures consacrées au journalisme
- > **67 %** des journalistes interrogés déclarent que les outils d'IA générative permettent de **gagner du temps** mais le gain de temps signalé est inférieur à celui signalé dans d'autres professions.
- > **80 %** déclarent réinvestir ce temps (quelques minutes par jour) dans d'autres tâches
- > **8,4%** déclarent que l'utilisation de ces outils (chatbots..) ont créés de nouvelles tâches imputables à l'IA elle-même : vérification des faits, révision de contenus générés, test de garantie de la conformité à l'éthique journalistique...

***Méthodologie** : menées en deux cycles (fin 2023 et 2024) couvrant chacun 25 000 personnes exerçant plusieurs professions dont le journalisme au Danemark.



"[Hype Cycle de Gartner](#)" | source : Creative Commons - https://commons.wikimedia.org/wiki/File:Gartner_Hype_Cycle.svg#filelinks

Quels sont les territoires de l'IA dans le journalisme ?



https://larevuedesmedias.ina.fr/sites/default/files/2025-02/Alliance_for_facts_IA_journalisme_FR_compressed_0.pdf



Emergence de chartes pour encadrer les usages de l'IA

Parmi les grands principes communs aux chartes sur l'IA dans les médias :

- ▶ Priorité à l'humain
- ▶ Transparence et traçabilité de la production IA
- ▶ Respect de la déontologie et des règles éthiques
- ▶ Prendre part à la gouvernance de l'IA
- ▶ Contrôle éditorial
- ▶ Ne pas nourrir les IA aux dépens des auteurs
- ▶ Personnalisation responsable
- ▶ Formation continue des journalistes
- ▶ Etc..

Tous les médias d'information n'ont pas encore adopté de charte sur l'IA (manque de maturité sur le sujet, travail en cours, rigidité d'un cadre dans un contexte technologique encore instable..).



<https://larevuedesmedias.ina.fr/les-medias-face-lintelligence-artificielle-20-chartes-passees-au-crible>



https://cdjm.org/files/recommandations/recommandation_IA.pdf

Depuis quand Data, IA & journalisme vont de pair ?

Quelques moments qui font dates

-> **Année 1960-70** : structuration d'une **nouvelle méthodologie** empruntant aux canons scientifiques pour faire du journalisme d'investigation avec les chiffres.

Exemple : Philip Meyer, journaliste au "Detroit Free Press", avec une enquête sur l'émeute de Détroit de 1967.

-> **1980-1990** : arrivée des premiers ordinateurs personnels. Aux Etats-Unis des journalistes **s'emparent de l'informatique pour l'intégrer à leur travail**.

Exemple : Bill Debman, journaliste d'investigation au "Atlanta Journal-Constitution", avec son enquête "La couleur de l'argent" en 1988.

-> **2009** : Simon Rogers, alors rédacteur en chef du Guardian, invente le terme de « data-journalisme » dans son livre "Facts are sacred" (2013).

-> **2010-2011** : WikiLeaks publie des dizaines de milliers de documents secrets sur les guerres en Afghanistan et en Irak. Collaboration entre WikiLeaks et cinq rédactions (*Le Monde, Der Spiegel, El Pais et The Guardian, The New York Times*)

-> **Novembre 2022** : arrivée grand public de ChatGPT

A lire sur le sujet

Alice Martin - "[Intelligence artificielle générative dans les rédactions françaises : perspectives, défis et enjeux](#)" (2024)

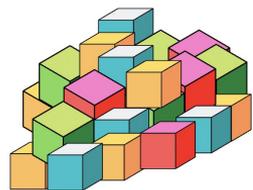
Sylvain Parasio - "Computing the news : datajournalism and the search for objectivity" (2022)



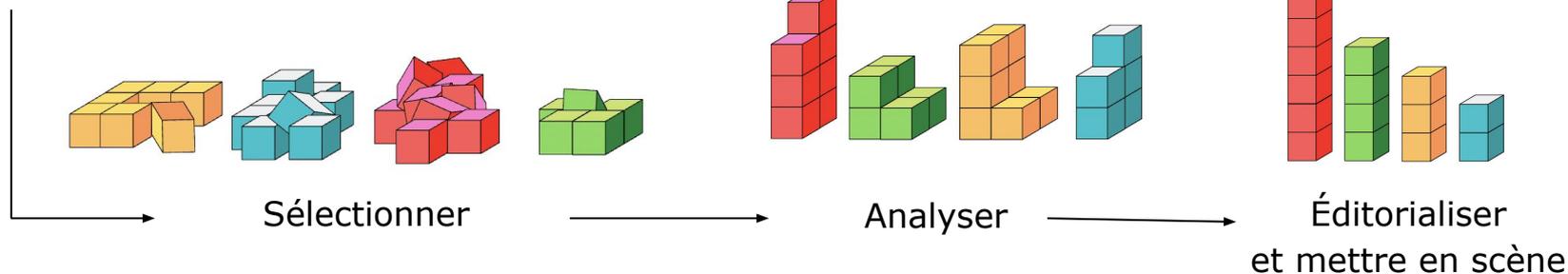
<https://larevuedesmedias.ina.fr/intelligence-artificielle-chatgpt-journalisme-il-foglio-ai>

Data-journalisme : de quoi on parle ?

Discipline de la rigueur au croisement des **compétences journalistiques, statistiques, techniques** (excel, code...), de **design d'information** (datavisualisation, infographie...).



Sourcer
des données



Une expertise au service du journalisme (national, régional et même local)

Quelques exemples de rédactions françaises avec des data-journalistes



LES DÉCODEURS

VENONS-EN AUX FAITS

**ouest
france**

Libération

franceinfo:

L'EXPRESS

Le Télégramme
dataspot

On défriche, on déchiffre, on décrypte



Rue89 Strasbourg

ENQUÊTES ET ACTUALITÉ À STRASBOURG ET EUROMÉTROPOLE

La revue des médias /

•contexte

ina

mind
MEDIA

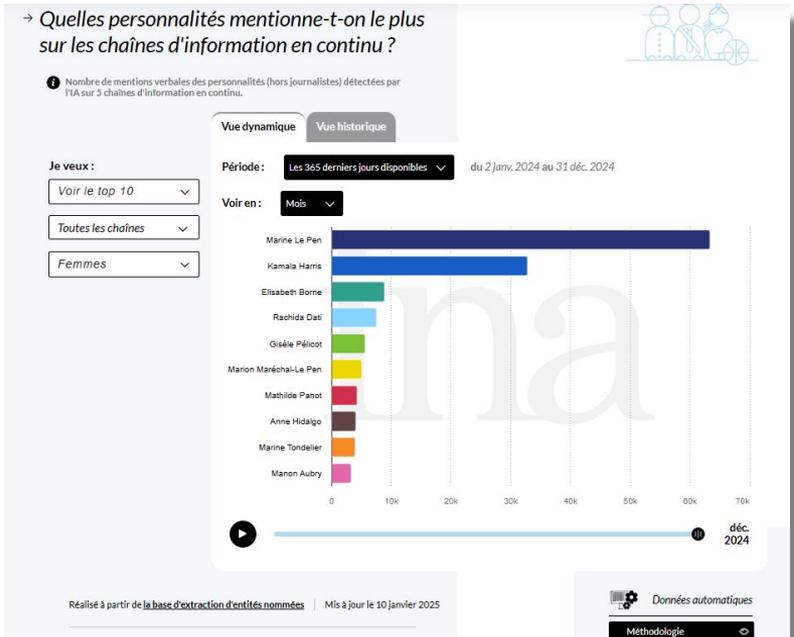
**WE
DO
DATA**

aef
info

IA et datajournalisme : pour faire quoi ?

Quelques exemples d'assistance via des outils IA dans le domaine du data-journalisme

▶ Créer une base de données



<https://data.ina.fr/cles-lecture/personnalites>

▶ Nettoyer & préparer des données (détection d'anomalies...)

▶ Évaluer rapidement le potentiel d'une base de données (points saillants, recommandation d'analyse...)

▶ Faire de la data visualisation

▶ S'aider dans la manipulation de tableurs (segmentation des données, aide avec les formules...)



<https://mailchi.mp/festival-infolocale/outils-numeriques-16530577?e=411081ec0d>

▶ Explorer de nouveaux champs d'enquêtes (volumétrie de données, analyse prédictive, modélisation...)



<https://ici.radio-canada.ca/info/2022/03/etalement-urbain-densite-population-villes-transp-ort-commun-changements-climatiques/>

A l'INA, comment on utilise la data et l'IA pour faire de l'enquête ?

Enquête conduite

avec les données des transcriptions IA

Comment ?

Création de dictionnaire de termes

Recherche sur les mots-clés dans les transcriptions

Interviews d'experts pour contextualiser les données

Périmètre : 16 médias entre le 1er janvier 2019 et le 30 juin 2024

- BFM TV, LCI, CNews et France Info, les matinales radios d'Europe 1, France Culture, France Inter, France Info, RMC, RTL et Sud Radio et les JT de TF1, France 2, France 3, Arte et M6



<https://larevuedesmedias.ina.fr/metoo-violences-sexistes-sexuelles-etude-ina>

Résultats

29 fois plus.

En 2024, le terme "féminicide" est près de 30 fois plus utilisé que le terme "crime passionnel".

Féminicide s'est imposé médiatiquement en 2019.

+17%

C'est la **hausse du traitement médiatique des violences sexistes et sexuelles** sur les chaînes info entre 2019 et 2024.

#Metoo

L'expression a été autant **employée en 2024 qu'en 2018**. On dénombre 2 031 mentions moyennes par an entre 2017 et 2024 sur les 16 médias étudiés.

A l'INA, comment on utilise la data et l'IA pour explorer l'actualité ?

Analyses des données de transcriptions IA & entités nommées

Périmètre :

16 médias sur l'année 2024

BFM TV, LCI, CNews et France Info, les matinales radios d'Europe 1, France Culture, France Inter, France Info, RMC, RTL et Sud Radio et les JT de TF1, France 2, France 3, Arte et M6

Résultats

7 fois plus.

Depuis le lancement de ChatGPT le 30 novembre 2022, les mots « IA » et « intelligence artificielle » sont 7 fois plus prononcés à l'antenne des 16 médias télé et radios présents sur data.ina.fr.

Mayotte

En 2024, rapporté à sa population, Mayotte est **plus mentionné que Paris** sur les 16 médias étudiés. Cela fait suite au cyclone Chido et à ses conséquences.

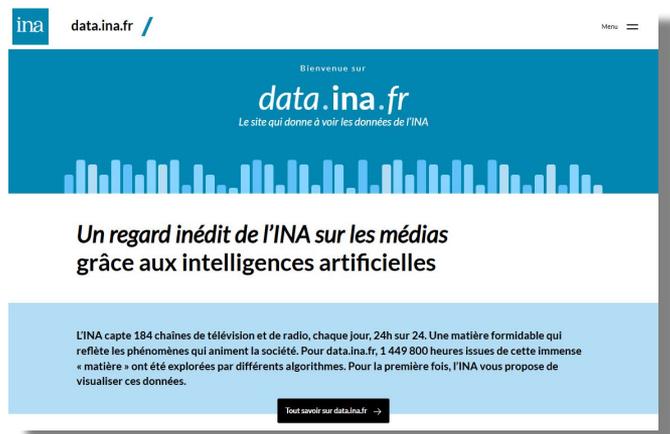
4 fois plus.

L'élection présidentielle américaine a été 4 fois plus médiatisée **que les élections européennes.**



A télécharger par ici: <https://info.ina.fr/barometre-de-linfo-1>

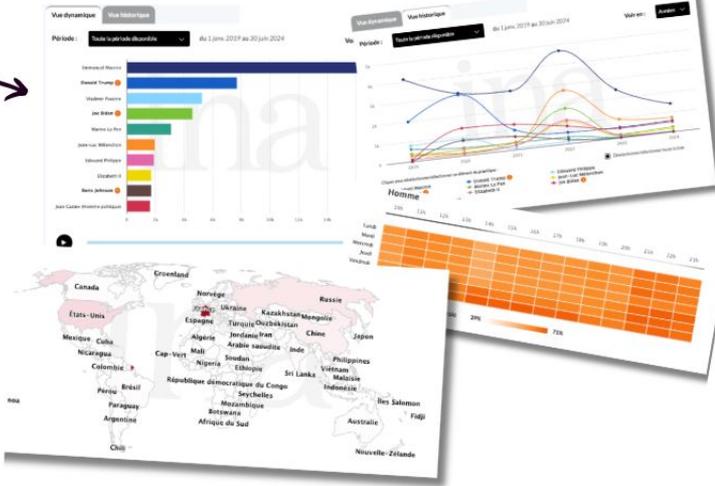
data.ina.fr : dix ans de data audiovisuelles à explorer & une façon inédite de s'acculturer à l'IA



16 questions

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 ?

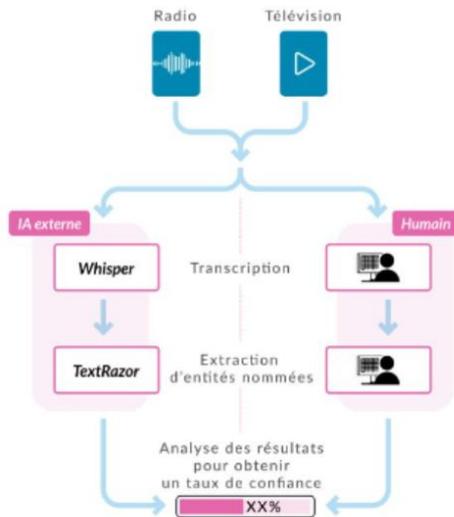


Une plateforme pour tous les publics qui cherchent à mieux comprendre l'information diffusée sur les médias audiovisuels : JT, chaînes d'information en continu, matinales radios.

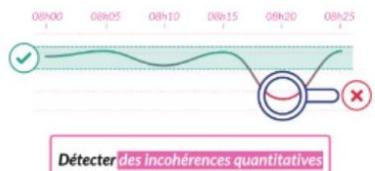
data.ina.fr : les grandes actions de contrôle

Vérité de terrain sur les IA externes

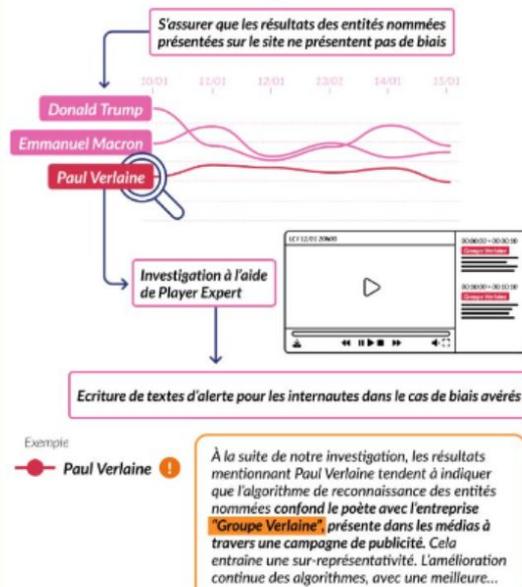
L'objectif est de comparer le traitement IA d'heures de télévision et radio (sélectionnées sur le périmètre data.ina.fr) et le travail humain de transcription et de reconnaissance des entités nommées



Contrôle des traitements automatiques



Contrôle de la pertinence



data.ina.fr : expérimenter des outils en transparence

ina data.ina.fr /

Accueil Les traitements IA

Les traitements IA

La grande particularité du site data.ina.fr est de pouvoir visualiser les résultats chiffrés de l'analyse des sons de chaînes de télévision et de radio françaises. Ces analyses ont été effectuées en partie via des outils d'Intelligence Artificielle.

Avant de vous plonger dans ces graphiques et cartographies, il est important de comprendre les règles méthodologiques établies pour ce site, les sources et les traitements mobilisés. Cette page est faite pour vous aider à lire ces résultats en parfaite connaissance de cause.

Nous sommes à votre disposition [si vous avez des questions et/ou des remarques](#).

Analyses automatisées du son

- La transcription
- La reconnaissance des entités nommées

Explorez la méthodologie complète

→ Découvrir

L'analyse automatique du son + Lire

La transcription + Lire

La reconnaissance des entités nommées + Lire

La méthodologie complète

Personnalités

Quelles personnalités mentionnées en tout premier dans les journaux télévisés ?

Quelles personnalités mentionnées en tout premier sur les chaînes d'information en continu ?

Quelles personnalités mentionnées en tout premier à la radio ?

Femmes-hommes

Quelle est la répartition du temps de parole entre les hommes et les femmes sur les chaînes de radio et de télévision ?

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

La foire aux questions

Pourquoi TMA n'a-t-il choisi de proposer ce site ?

À qui est destiné le site data.ina.fr ?

Quel est l'objectif du site data.ina.fr ?

Comment fonctionne ce site ?

À quelle fréquence le site est-il mis à jour ?

Comment est assurée la qualité des données sur ce site ?

Quelles sont les chaînes de télévision et radio analysées sur ce site ? Pourquoi le web est-il si absent ?

Pourquoi les journalistes sont-ils exclus des résultats du site ?

Comment un nom de personne ou de pays sont-ils détectés dans une vidéo ?

Chaque vidéo qui est captée est composée à la fois d'images et d'une bande son. C'est cette dernière qui nous intéresse comme matériau de départ et que l'on va donc isoler en vue de son analyse par deux outils d'Intelligence Artificielle, chacun spécialisé dans une tâche précise.

Prenons l'ouverture d'un journal télévisé et voyons comment l'on peut, grâce à l'IA, détecter les noms de personnes et de lieux énoncés par le journaliste.

1 De la vidéo au texte : le "speech-to-text"

Il existe de nombreux outils IA permettant de faire de la transcription. Sur data.ina.fr, nous avons choisi Whisper dont le fonctionnement a été décrit dans des papiers scientifiques.

Voici comment l'algorithme procède :

Notre vidéo : l'ouverture d'un journal télévisé

Le son de la vidéo est isolé

L'audio est converti en un spectrogramme qui peut être visualisé sous la forme d'une image

Vidéo (images + bande son)

Le spectrogramme est envoyé dans le modèle "transformer", qui est une architecture de réseau de neurones de type encodeur-décodeur.

Il permet d'encoder le signal issu du spectrogramme et de le décoder sous la forme du texte associé.

Son architecture lui permet de :

- Se concentrer sur les moments d'activité vocales
- Transcrire la parole en français
- Horodater la phrase

Algorithme

Dans notre cas, l'algorithme a produit le résultat suivant sous la forme de séquences de langue, de tâche, de mots, de temps (aussi appelées "tokens" multitâches).

tokens de... langue tâche temps token de mots prochaine prédiction

FR transcription 0,0 Madame monsieur bonjour ...

FR transcription 0,0 Madame monsieur bonjour dans ...

Après avoir prédit l'ensemble des séquences de mots de l'extrait vidéo, le résultat final est le suivant :

FR transcription 0,0 Madame monsieur bonjour dans les titres de cette édition Emmanuel Macron se rend à Paris pour annoncer sa feuille de route

2 De la transcription aux entités nommées : repérer les personnalités et les pays

Nous utilisons TextRazor, un outil pour la reconnaissance d'entités nommées (REN). Voici comment des outils similaires fonctionnent en utilisant une combinaison de techniques d'analyses linguistiques et de modèles d'apprentissage automatique pour identifier et catégoriser les entités du texte.

L'outil REN part du résultat de l'outil de transcription fourni sous la forme de séquence de "tokens" et il analyse les mots.

Madame monsieur bonjour dans les titres de cette édition Emmanuel Macron se rend à Paris pour annoncer sa feuille de route

nom ponctuation déterminant

Madame monsieur bonjour dans les titres de cette édition Emmanuel Macron se rend à Paris pour annoncer sa feuille de route

préposition verbe nom propre pronom réfléchi

Pour identifier le rôle grammatical de chaque mot dans la phrase, l'algorithme va procéder à l'analyse morphosyntaxique des "tokens".

Pour aller plus loin



<https://www.ina.fr/actualites-ina/data-donnees-visualisation-intelligence-artificielle-dataina>

Prédire et analyser la présence médiatique grâce à l'IA



<https://mediarama.io/predire-et-analyser-la-presence-mediatique-grace-a-lia/>



Des questions ?