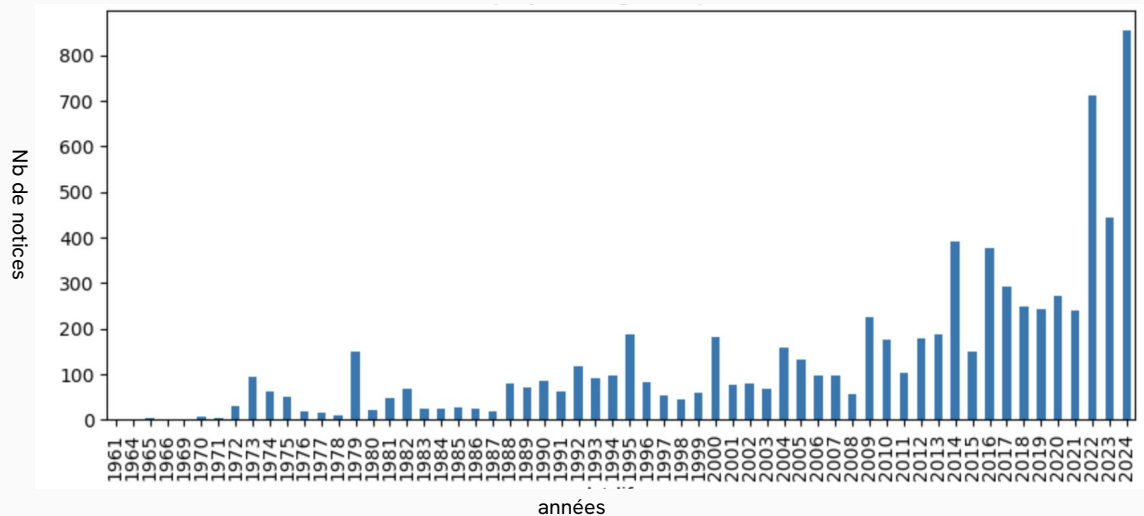


# L'avortement à la télévision française (1960-2024)

---

- **Données :**
  - **Notices documentaires de sujet de journaux télévisés et magazines télévisés**
    - récupérées si matching d'au moins un des mots suivant :  
“avortement”, “ivg” et “ interruption volontaire de grossesse”
    - infos : date, chaîne, intervenants, titre de collection, etc.
  - **Transcriptions** (majoritairement via Whisper)
- **Points étudiés :**
  - Thèmes principaux au cours du temps et par chaînes
  - Proximité des discours télévisés avec différents types d'embeddings
  - Extraction d'extraits pertinents

## Visualisation du corpus



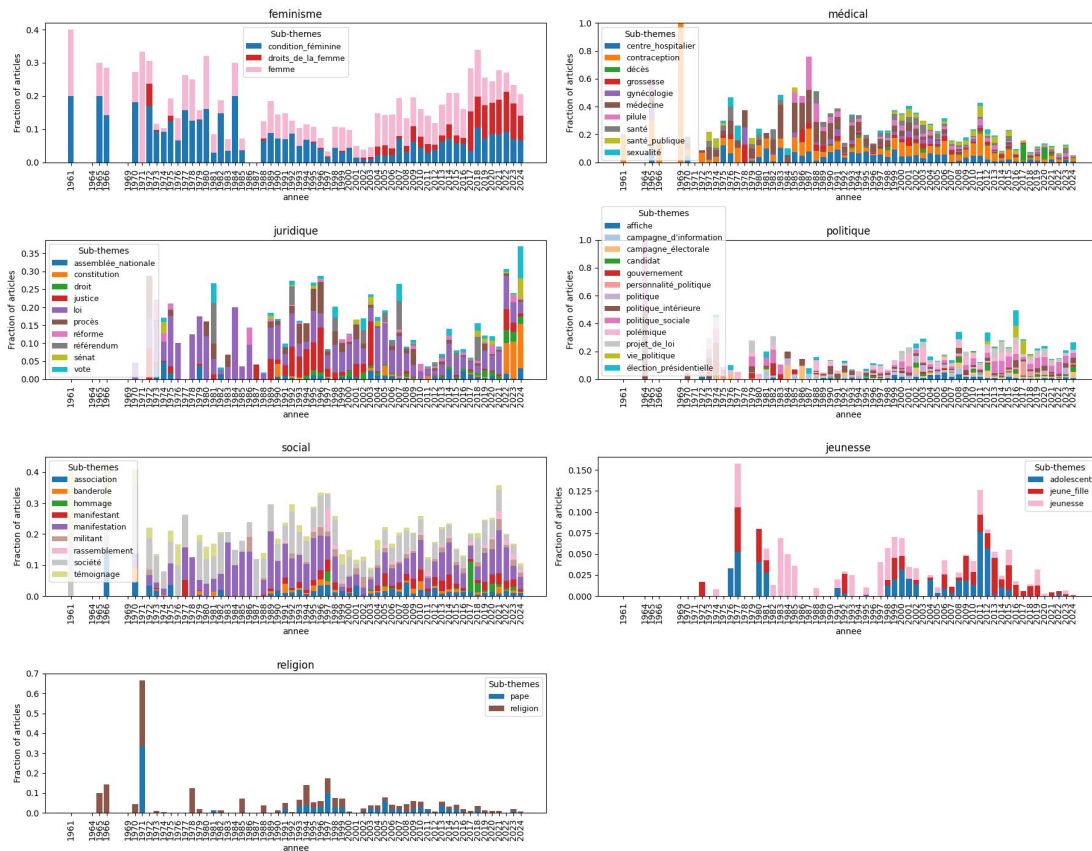
Nombre de sujet de journaux télévisés et magazines par année discutant de l'avortement

→ l'augmentation correspond à l'augmentation générale des contenus diffusés à la télévision et la multiplication du nombre de chaînes télévisées dans le temps

## 1) Identification des thèmes principaux

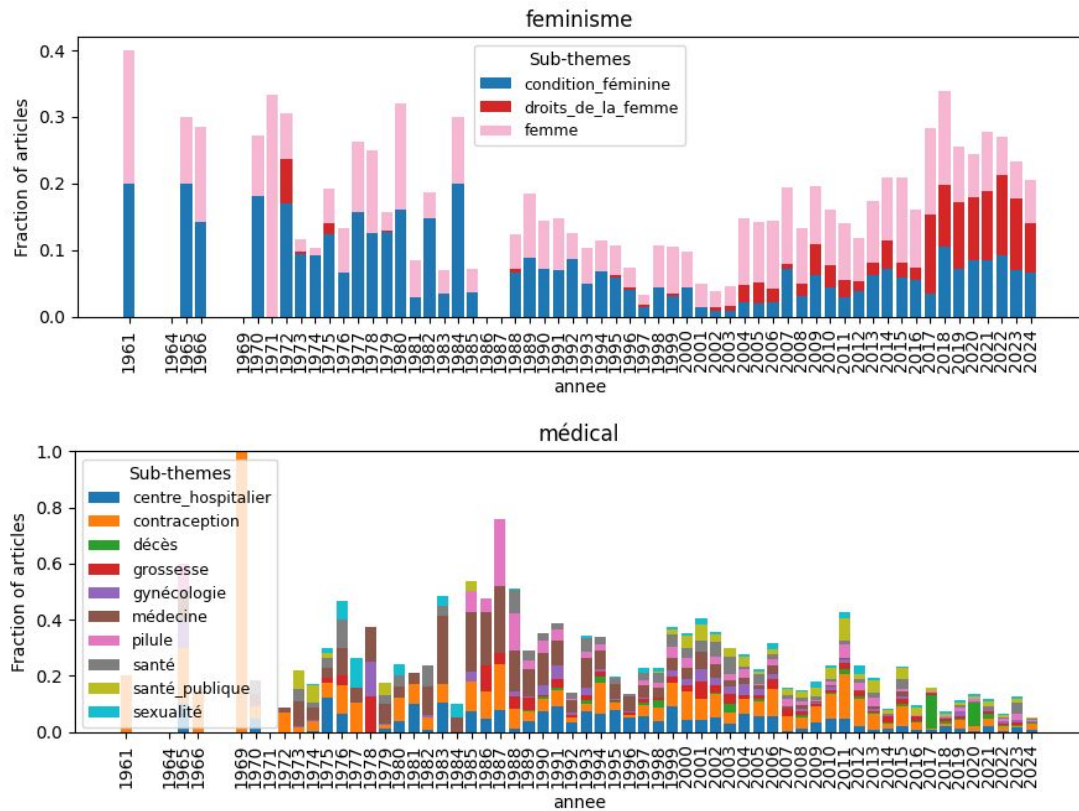
### Méthode :

- Identification de thèmes dans les documents traités  
  
→ construction de grandes catégories à partir des thèmes les plus communs dans la colonne "thes", représentation de la fréquence d'apparition des thèmes pour chaque année proportionnellement au nombre de documents publiés et au nombre de thèmes traités par an.
- Même tâche relativement à une sélection de chaînes de diffusion.



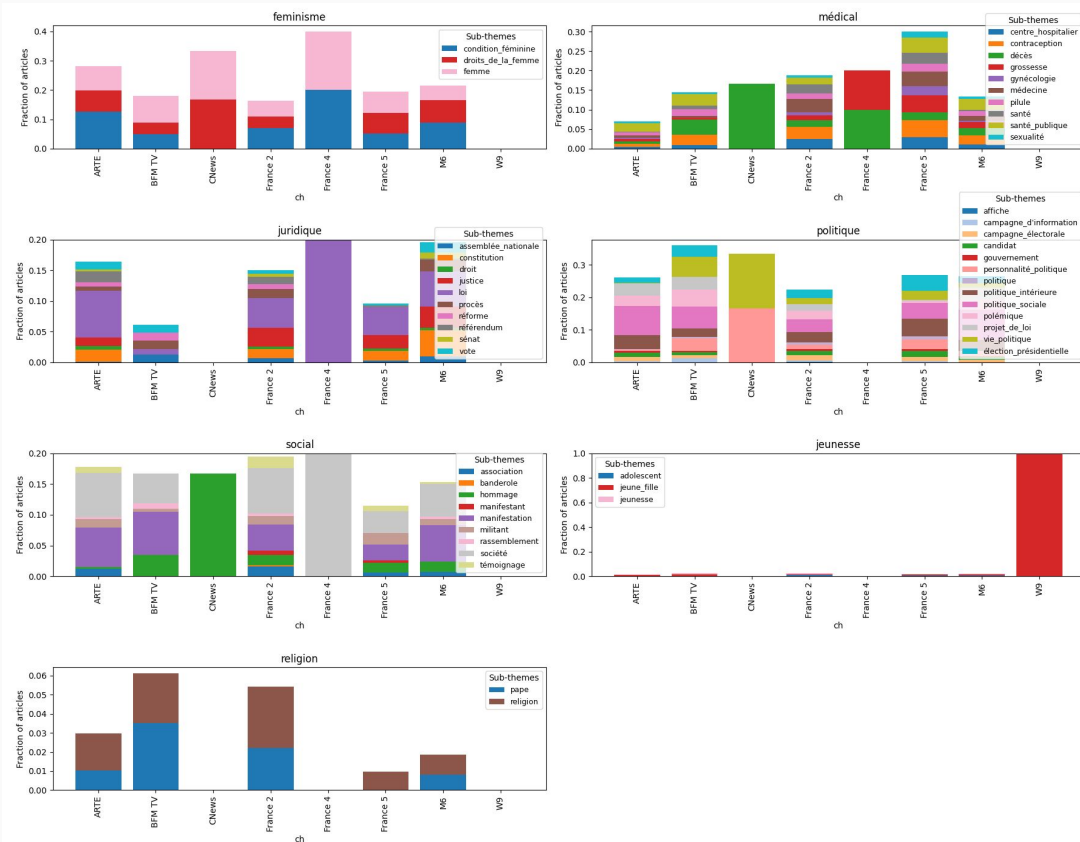
- Apparition du topic des droits de la femme dans les années 2000 puis augmentation de sa pertinence
- Baisse des thèmes médicaux dans le traitement de l'IVG depuis les années 2000.
- Pic de topics juridiques dans les années 2022 à 2024. Creux dans les années 2008 à 2021.
- Pic politique en 2013, vraisemblablement à l'occasion de la loi Taubira.

Thèmes principaux selon les années



- Apparition du topic des droits de la femme dans les années 2000 puis augmentation de sa pertinence
- Baisse des thèmes médicaux dans le traitement de l'IVG depuis les années 2000.
- Pic de topics juridiques dans les années 2022 à 2024. Creux dans les années 2008 à 2021.
- Pic politique en 2013, vraisemblablement à l'occasion de la loi Taubira.

Thèmes principaux selon les années



→ Certaines chaînes traitent de thèmes peu divers au sein des catégories : CNews ne parle que d'hommage pour le social et de décès pour le médical. Grande diversité sur d'autres

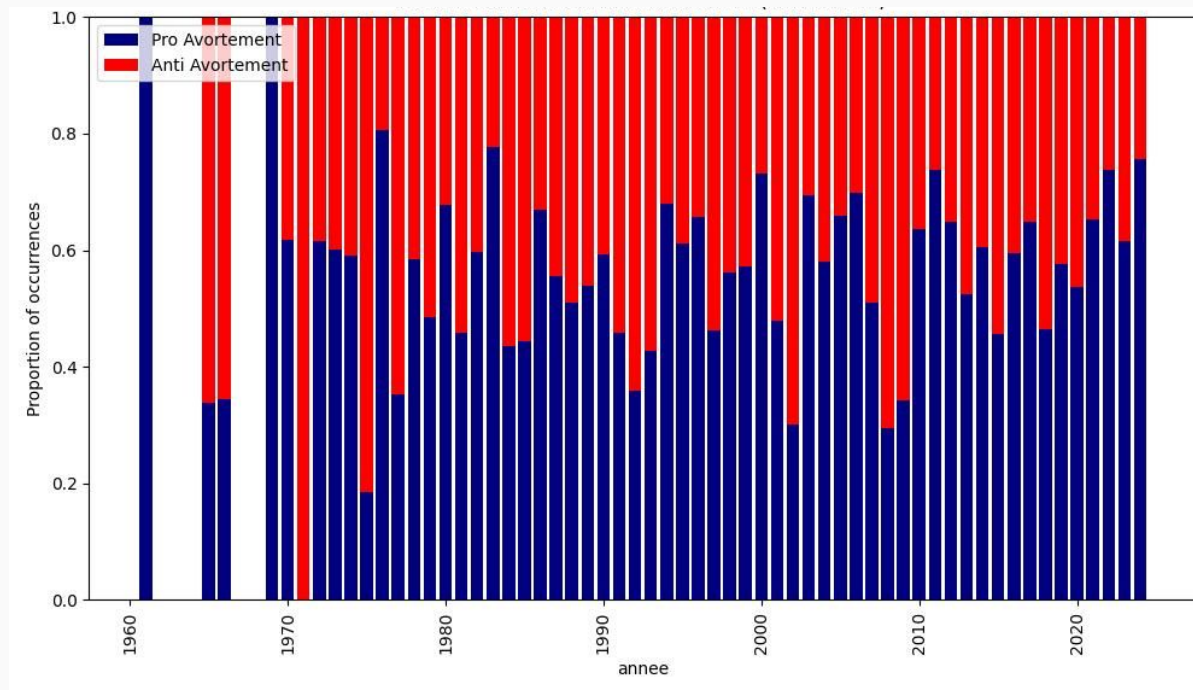
Thèmes principaux par chaînes

## 2) Identification de la tendance pro/anti avortement

### Méthode :

- Construction de listes de mots connotant un discours pro ou anti avortement (voir annexe). Calcul de la fréquence de textes pro ou anti avortement pour chaque année.
- De même pour chaque chaîne.
- Listes utilisées :
  - **pro\_avortement** =  
[ "choix", "liberte", "autonomie", "sante", "reproductive", "legal", "securisee", "femme", "egalite", "contraception", "droits", "protection", "planification", "maternite", "decision", "accès", "soutien", "prevention", "controle", "securite"]
  - **anti\_avortement** =  
[ "vie", "foetus", "enfant", "naissance", "conception", "sacre", "bebe", "provie", "interdiction", "lutte", "defense", "moralite", "famille", "responsabilite", "respect", "opposition", "antiivg", "innocent", "naturel", "precieux", "meurtre", "assassinat", "suppression", "elimination", "destruction", "homicide"]



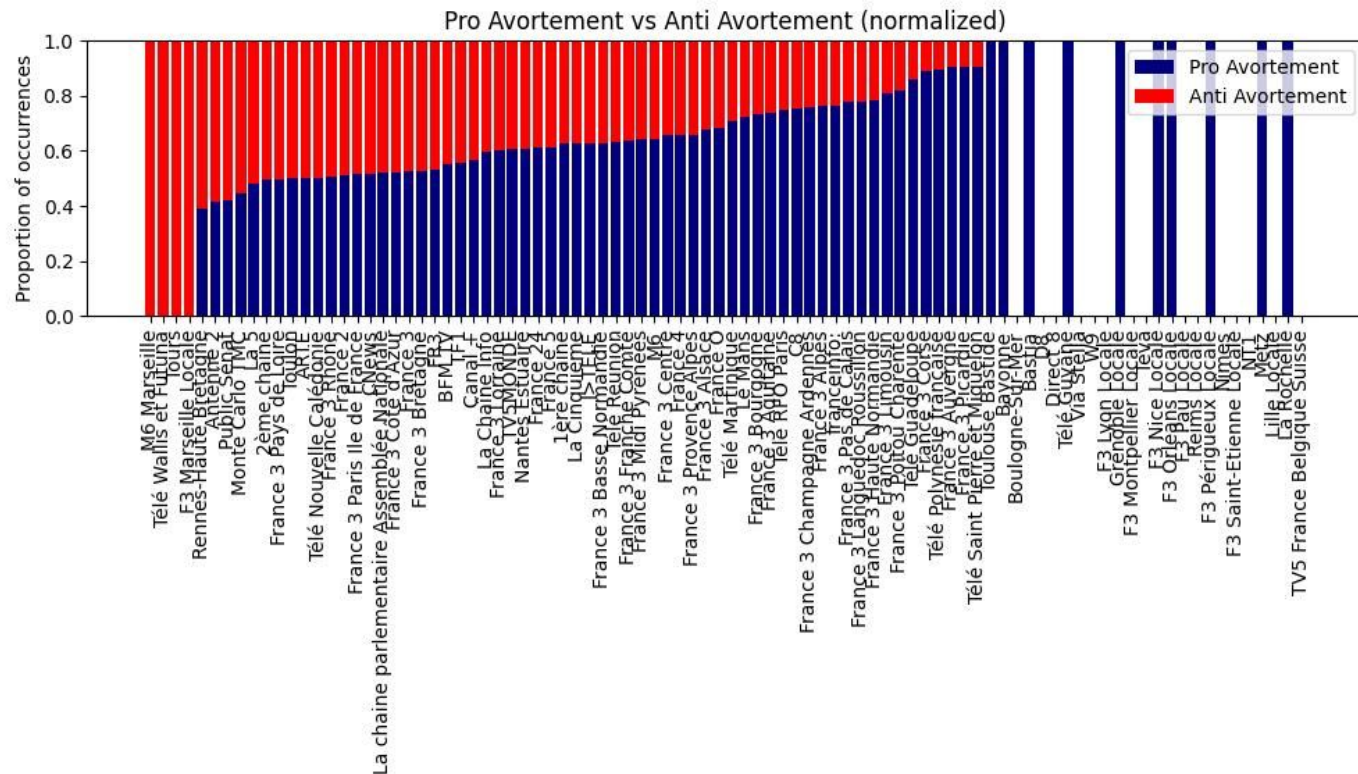


Fréquence relative de discours pro ou anti avortement en fonction du temps

- Globalement, les chaînes ne sont pas clivantes.
- Quelques marginaux très orientés :

Boulogne-Sur-Mer  
, F3 Lyon locale,  
Rennes locale, F3  
Marseille locale,  
F3 Périgueux  
locale,  
Rennes-Haute  
Bretagne.

- La première chaîne nationale la plus anti-avortement est CNews.



Fréquence relative de discours pro ou anti avortement par chaînes

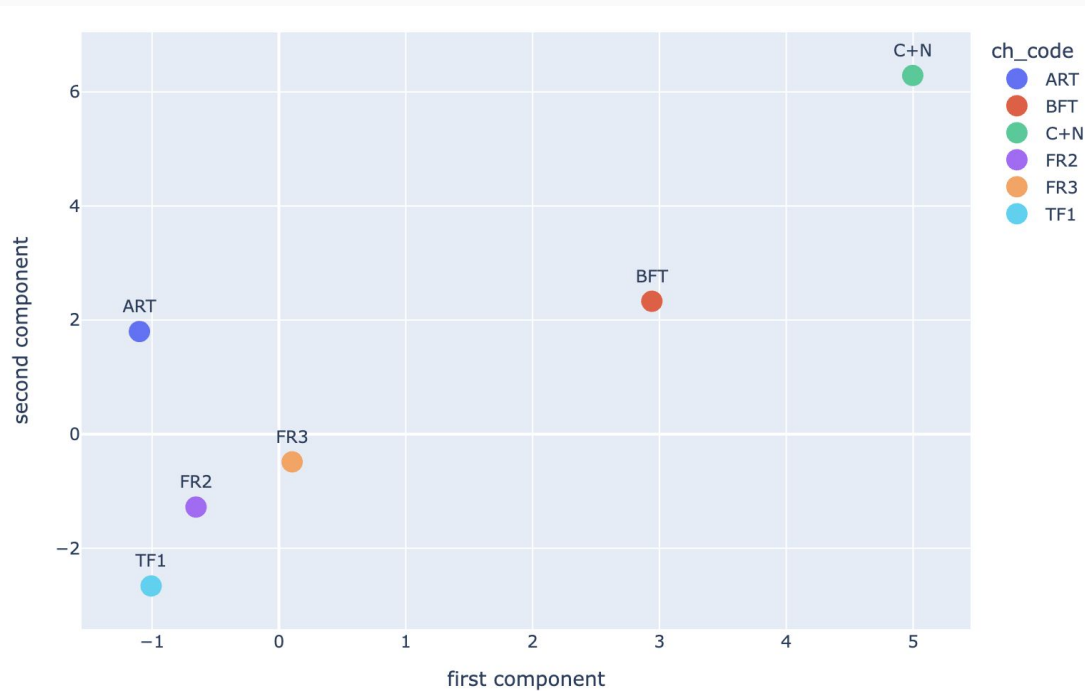
**But : Etudier les similarités entre les discours télévisés traitant de l'avortement**

**Différents types d'embeddings testés**

**Première méthode : modèle all-MiniLM-L6-v2**

- Embeddings de textes avec un modèle transformers : all-MiniLM-L6-v2
  - Avantage : fenêtre contextuelle non disponible dans les modèles classiques de word embedding
  - Etapes :
    - Division de chaque transcription en plusieurs blocs de maximum 300 mots (input max de all-MiniLM) de manière à ce qu'une phrase ne soit pas séparée entre plusieurs blocs.
    - Moyenne des embeddings des différents blocs de chaque sujet de JT
- On obtient un embedding de dimension 384 par sujet de JT et magazines TV

## Premiere méthode : modèle all-MiniLM-L6-v2



Embeddings moyen par chaîne, modèle all-Mini-LM et réduction TSNE

Méthode de visualisation :

- Moyenne d'embeddings de tous les sujet de JT et magazines par chaînes
- Réduction de dimensionnalité TSNE
- Visualisation en 2D

Résultat :

- L'axe 1 (abscisses) peut être interprété comme un axe d'orientation politique. Cnews et BFMTV sont les plus à droite, ARTE et TF1 les plus à gauches France 2 et France 3 sont assez proches entre eux
- Plus de mal à interpréter le second axe

## Deuxième méthode : embedding de mots via Fasttext

- Embeddings de mots pour chaque sujet de JT et magazine TV (300 dimensions)
- Moyenne des mots de chaque sujet de JT et magazines (obtention d'un vecteur par sujet de JT et magazine)

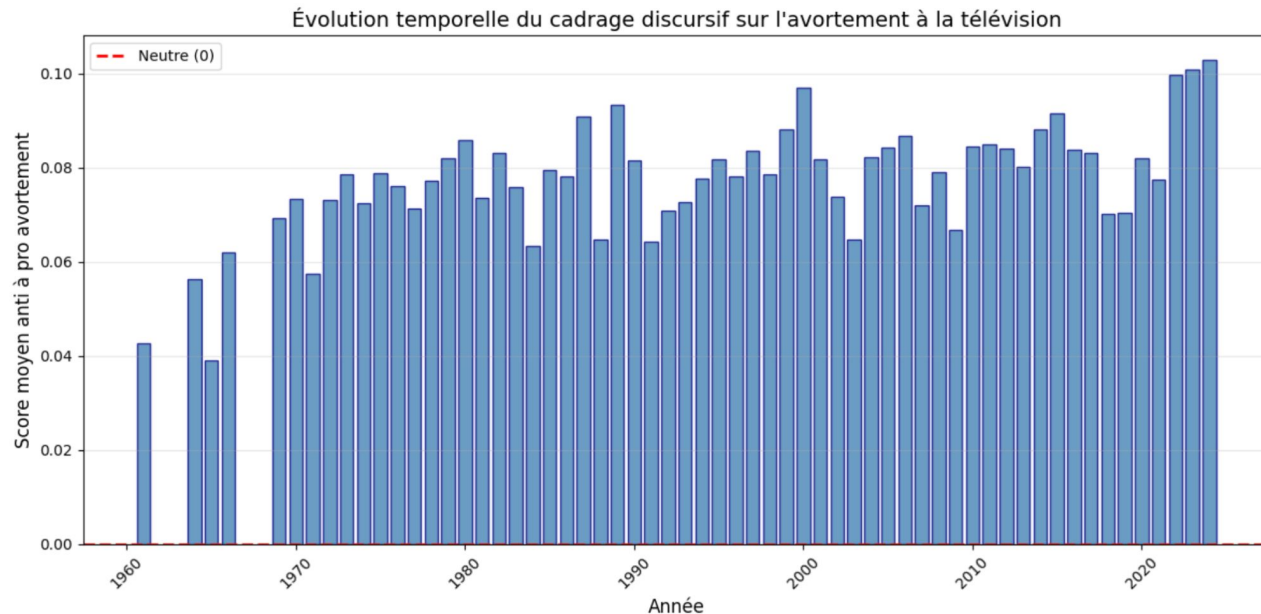
## Création d'un axe anti-avortement à pro-avortement à partir des embeddings

→ A partir de mots caractéristiques :

- mots anti-avortement (version criminalisante du discours) :  
["meurtre", "interdiction", "assassinat", "immoral", "suppression", "elimination", "destruction", "homicide"]
- mots pro-avortement (orienté vers le droit et les soins) :  
["droit", "choix", "liberte", "autonomie", "sante", "contraception", "acces", "soutien", "prevention"]

→ Attribution de scores à chaque Sujet de JT et magazines en projetant les embeddings des textes sur l'axe anti-avortement à pro-avortement

## Création d'un axe anti-avortement à pro-avortement à partir des embeddings



Augmentation des contenus pro-avortement dans le temps, un pic ces trois dernières années.

Ce pic peut être expliqué par des événements médiatiques autour de l'ivg :

- Constitutionnalisation de l'IVG en mars 2024
- 50 ans de la loi Veil en 2025

Evolution dans le temps du score anti-avortement à pro-avortement

## Extraction de contenu traitant de l'avortement

### Problématique :

- Problème inhérent aux données de l'INA concernant les transcriptions des sujets de JT et magazine :
  - Transcriptions sont produites à partir des horaires de diffusion associées aux notices. Seulement, les horaires "réels" peuvent différer des horaires prévisionnels  
→ décalage dans les transcriptions
- Du bruit peut être présent, un sujet de JT peut aborder l'avortement pendant 2 minutes puis consacrer le reste du sujet à d'autres actualités

### But :

→ Extraire des transcriptions les passages parlant spécifiquement de l'avortement

**Méthode :**

- Conserver uniquement les phrases contenant un des mots "avortement", "ivg" et " interruption volontaire de grossesse", la phrase précédente ainsi que les deux phrases suivantes.
- Extraction à la main des morceaux pertinents de 50 sujets de JT et magazines pour comparaison

**Résultats :**

- Précision moyenne : 0.775
- Rappel moyen : 0.608
- F1-score moyen : 0.681

**Limites :**

- ne prend pas en compte les variations dans la quantité de contexte donné
- match le mot "avortement" même quand ce dernier ne fait pas référence à une ivg. Exemple : "l'avortement du projet...".

**Poursuites :**

Utilisation d'un LLM (en local ou serveur protégé) tel que Mistral afin d'extraire les extraits pertinents de manière plus précise et contextuelle.



**Merci pour votre attention !!**

