



# Étudier le journalisme avec de l'IA

ESJ + Université Laval  
17 novembre 2023



# Introduction

## Qui sommes nous?

- Une équipe de recherche en sociologie (politique/computationnelle)
- Qui utilise l'IA dans ses recherches
- Dont des projets sur le journalisme



Emma Bonutti  
d'Agostini



Rubing Shen



Étienne Ollion

# Introduction

## **Objectif de cette session**

- Une initiation à la logique de la recherche appuyée par de l'IA
- Réfléchir ensemble à l'écriture du journalisme
- Présenter quelques ressources pour aller plus loin

# Introduction braconnière à l'IA

**L'IA, une longue histoire**

# Introduction braconnière à l'IA

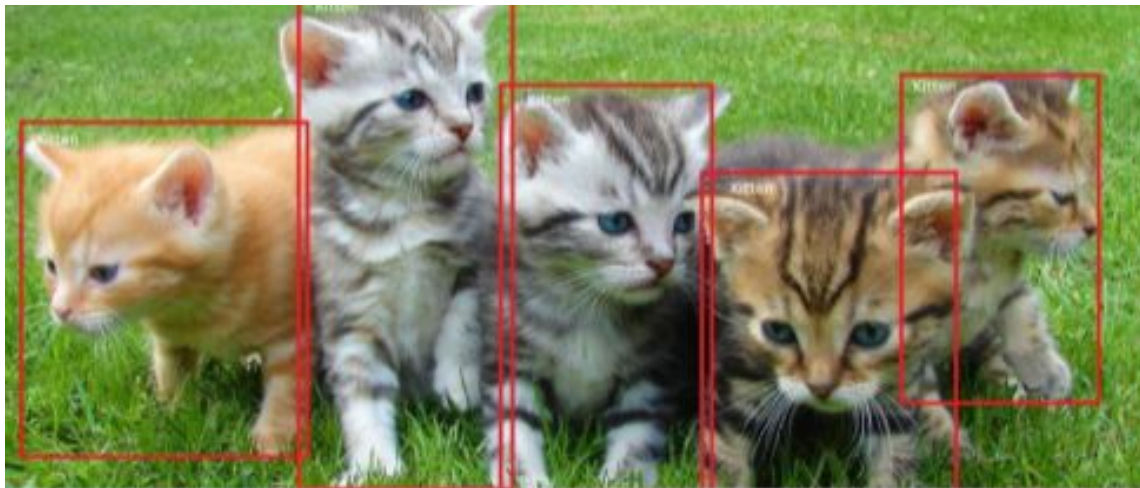
## Un tournant : les LLMs

- Un des outils centraux de l'extraction de contenu
- Qui s'appuie sur une architecture : les *Transformers*
- Des usages omniprésents (chatGPT, GPT 3.5, Bard, Claude, etc.)

# Introduction braconnière à l'IA

## Objectif

- Une analyse fine, et massive
- L'idée, pas le signe



# Introduction braconnière à l'IA

## Méthode

- Un apprentissage “supervisé”
- Une représentation mathématique du langage

## Plus d'informations :

- S'augmenter. Présentation disponible [sur mon site](#)
- *The Augmented Social Scientist*, [article de recherche](#), 2022

# Une expérience

**Les sciences sociales et l'intersectionnalité**



# Une expérience

## **Les sciences sociales et l'intersectionnalité**

### *0. Déterminer un corpus*

Ici : les résumés des articles de sciences sociales françaises depuis 2000

- 120 revues, 40 000 résumés

# Une expérience

## Les sciences sociales et l'intersectionnalité

### 0. Déterminer un corpus

Ici : les résumés des articles de sciences sociales françaises depuis 2000



The screenshot shows the CLIO interface for an article. At the top left is a green back arrow. Next to it is the article title 'La supériorité sportive féminine soviétique, un enjeu de guerre froide' by Sylvain Dufraisie. Below the title, it says 'DANS CLIO. FEMMES, GENRE, HISTOIRE 2023/1 (n° 57), PAGES 113 À 131'. To the right of the title are three buttons: 'CITER OU EXPORTER', 'AJOUTER À UNE LISTE', and 'SUIVRE CETTE REVUE'. Below the title bar are four tabs: 'ARTICLE', 'RÉSUMÉ' (which is active), 'PLAN', and 'AUTEUR'. To the right of the tabs is a blue 'ACHETER' button. The main content area shows the word 'Français' followed by a paragraph of text. A red arrow points from the right side of the page to the word 'égalité' in the text.

**La supériorité sportive féminine soviétique, un enjeu de guerre froide**  
Sylvain Dufraisie  
DANS CLIO. FEMMES, GENRE, HISTOIRE 2023/1 (n° 57), PAGES 113 À 131

CITER OU EXPORTER  
AJOUTER À UNE LISTE  
SUIVRE CETTE REVUE

ARTICLE **RÉSUMÉ** PLAN AUTEUR ACHETER

**Français** Durant la fin des années 1940, les sections de sport et le comité national olympique soviétique intègrent les fédérations et le Comité international olympique, ce qui leur permet de prendre part aux compétitions et à la communauté mondiale des sportifs. Très vite, les championnes d'URSS remportent de francs succès. La participation soviétique y a une double fonction dans la « politique extérieure de l'image » (R. Frank). D'une part, elle donne à voir les succès en matière d'égalité femmes-hommes. D'autre part, elle permet d'obtenir des médailles, essentielles à la mesure de la suprématie soviétique en matière de sport. Cet article étudie la manière dont l'administration et les médias soviétiques ont construit – en pratique et en représentations – la supériorité sportive féminine soviétique en interaction avec l'étranger. Loin d'être isolés, les acteurs du sport soviétique s'adaptent aux règlements internationaux, aux réactions que les performances des athlètes d'URSS suscitent et aux controverses qui surgissent dans la presse de l'adversaire.

# Une expérience

## Les sciences sociales et l'intersectionnalité

### *1. Définir une catégorie et son périmètre*

Ici: Le genre comme rapport social entre des sexes

- Le sexe comme variable (statistique)
- Les rapports hiérarchiques entre sexes
- Le genre comme performance
- L'orientation sexuelle

Note: qu'on choisit nous même ce qu'on inclut ou pas

# Une expérience

## Les sciences sociales et l'intersectionnalité

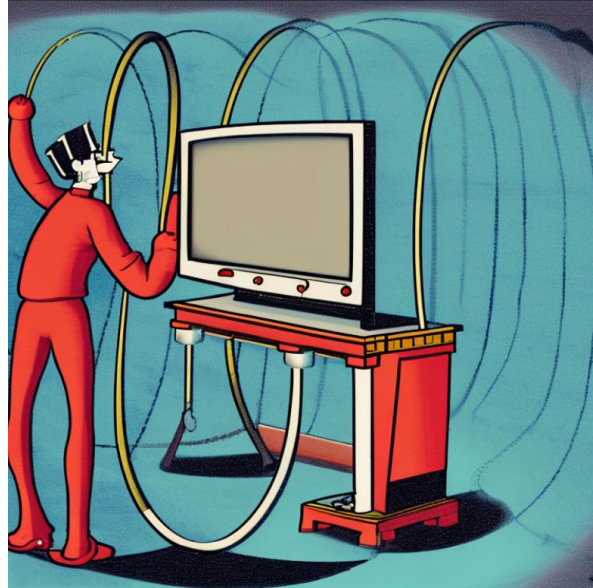
*2. Annoter X exemples pour entraîner un modèle*



# Une expérience

## Les sciences sociales et l'intersectionnalité

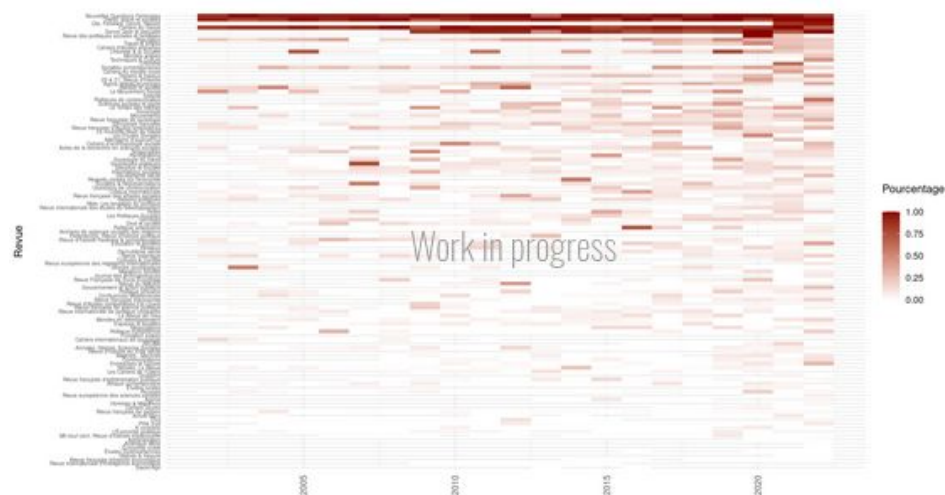
### *3. Entraîner un modèle*



# Une expérience

**Les sciences sociales et l'intersectionnalité**

*4. Valider*



# Une expérience collective



# Une expérience collective

## L'écriture du journalisme

### Corpus:

- 5 quotidiens nationaux
- Service politique
- Tous les articles signés, 1998-2018

# Une expérience collective

## L'écriture du journalisme

### 1. Deux tâches

- Le "off the record"

*Trouver les phrases qui l'introduisent*

Le Bingo des sources fiables

souffle un proche	chuchote un élu	glisse un comparse	assène un ministre
susurre un visiteur du soir	confie un acolyte	laisse entendre un ténor	croit savoir un confident régulier
vitupère un contempteur	Grommelle un porte-flingue	trompette un éléphant	s'alarme un militant
risque un ministrable	balaie un Marcheur	rétorque un fidèle	s'inquiète un compagnon de route
s'emporte un 'socio'	lâche un habitué des lieux	proteste un de ses amis	cingle un Macroniste
lâche un parlementaire du premier cercle	souligne un proche du dossier	souffle un pont de l'industrie	vocifère un conseiller
confie-t-on en interne	pronostique une source dans l'exécutif	Ironise un faiseur de rois	Croit-on savoir dans le premier cercle

# Une expérience collective : annotation

## L'écriture du journalisme

### 1. *Deux tâches*

- *Le "off the record"*
- *L'expression de la subjectivité*

### *Trouver des traces à différents niveaux*

- *Indicateurs micro-textuels (pronoms personnels, verbes etc.)*
- *Indicateurs sémantiques (termes spécifiques utilisés, connotés, etc. )*
- *Indicateurs structurels (raisonnement, logique de l'exposition)*

# Une expérience collective

## L'écriture du journalisme

### 2. *Votre mission*

- *Réfléchir (seul.e, en petits groupes)*
- *1er point collectif*
- *Une annotation via une interface*
- *Résultats et discussion collective*



# Une expérience collective

## L'écriture du journalisme

Environ 15'

- Réflexion (seul, en groupe)
- Appuis : sélection d'articles
- **Objectif**: définir des critères (un peu) précis en vue d'une annotation

# Une expérience collective

## L'écriture du journalisme

### *3. Tigger, une plateforme d'annotation*



Project / Scheme

Strategy

On

Tagged: 0 ; Untagged: 500

Tag at least 3 texts for 2 tags to train pred.

 Train



Project

Tagging

History

Stats

BERT

Export

1. Créez votre projet en lui donnant un nom

(Regex filter)

☐ Case


☒ Text

☐ Visualization

(New label)

Create

(Comment)

 No model

☐ Context

« Tactiquement, il vaut mieux souligner que Hollande reprend nos propositions pour le faire passer pour un président de droite et ainsi mettre la pagaille dans la majorité », confie un de ces modérés.



Project / Scheme

+ off

+ scheme0

Strategy

Random

On

Untagged

Tagged: 0 ; Untagged: 1500  
Tag at least 3 texts for 2 tags to train prediction model.

Train



Project Tagging History Stats BERT Export

Settings Sample Scheme

Current scheme



scheme0

Save

Write scheme description here

Rename tags

Current

Count

New

Décrivez ici votre stratégie d'annotation

- ce que vous cherchez en général
- quels indices vous mobilisez



Mettre random par défaut



Project / Scheme

+ Off

Strategy

Random

On

Untagged

Tagged: 0 ; Untagged: 500

Tag at least 3 texts for 2 tags to train pred.

+ etienne

Train



Project

Tagging

History

Stats

BERT

Export

(Regex filter)

☐ Case

☒ Text

☐ Visualization

(New label)

Create

(Comment)

No model

☐ Context

« Tactiquement, il vaut mieux souligner que Hollande reprend nos propositions pour le faire passer pour un président de droite et ainsi mettre la pagaille dans la majorité », confie un de ces modérés.

Création de labels

Texte à annoter



Project / Scheme

+ Off

+ etienne

Strategy

Random

On

Untagged

Tagged: 0 ; Untagged: 500

Tag at least 3 texts for 2 tags to train pred.



Project

Tagging

History

Stats

BERT

Export

(Regex filter)

☐ Case

☒ Text

☐ Visualization

(New label)

Create

(Comment)

No model

☐ Context

« Tactiquement, il vaut mieux souligner que Hollande reprend nos propositions pour le faire passer pour un président de droite et ainsi mettre la pagaille dans la majorité », confie un de ces modérés.

Commentez les textes que vous trouvez intéressants / ambigus / difficiles à classer

# Une expérience collective

## L'écriture du journalisme

### 2. *Votre mission*

- *Réfléchir (seul.e, en petits groupes)*
- *1er point collectif*
- *Une annotation via une interface*
- *Résultats et discussion collective*



# Une expérience collective

## L'écriture du journalisme

### *3. Tigger, une plateforme d'annotation*

**Off the record:** <https://tinyurl.com/esj-off>

**Subjectivité:** <https://tinyurl.com/esj-subj>

# Discussion collective

1. *Mettre à jour votre grille d'annotation (ce qui marche, ce qui ne marche pas)*
2. *Et discutons-en*

Résultats non validés rigoureusement, à ne pas réutiliser

# Résultats off

945 phrases annotées

```
==== Epoch 3 / 3 =====
Training...
Average training loss: 0.26
Training took: 0:00:08

Running Validation...
Average test loss: 0.28
Validation took: 0:00:00
```

	precision	recall	f1-score	support
0	0.96	0.92	0.94	72
1	0.77	0.87	0.82	23
accuracy			0.91	95
macro avg	0.86	0.89	0.88	95
weighted avg	0.91	0.91	0.91	95

Résultats non validés rigoureusement, à ne pas réutiliser

# Résultats off

Un déplacement décidément bien politique, car le choix de Dreux ne doit rien au hasard, reconnaît-on à l'Elysée.	2017	2017-02-28	Le Figaro	1	0.871101
Même s'il a admis que, si c'était à refaire, il le referait et même si son entourage a tenu à rappeler que, à l'époque, entre les deux tours de la présidentielle de 1988, la décision de donner l'assaut de la grotte d'Ouvéa avait été « prise avec l'accord explicite et les instructions directes » de François Mitterrand, Jacques Chirac a joué l'apaisement.	2003	2003-07-30	Le Figaro	1	0.672614
Pour toutes ces raisons, sans doute, mais aussi parce qu'il mesure les dégâts causés à l'image du parti écologiste par les guerres de courants, M. Hascoët joue l'apaisement.	2000	2000-05-21	Le Monde	0	0.932886
« Mais, note un haut fonctionnaire français, la vertu a été finalement récompensée.	1998	1998-04-15	Le Figaro	1	0.871579
« Dans ce contexte, l'augmentation des prix du pétrole sera sans conséquence sur les salaires.	2000	2000-04-01	Le Monde	0	0.932881

# Résultats

	F1-Score (en %)
Off	94
Pas Off	99

Pour le D-Day , d ' accord , mais pour une remise de décoration , inutile ,

s ' agace un collaborateur.

off

Certains visiteurs du soir l ' ont vu éteindre lui-même la lumière au palais ,  
alors qu ' il les raccompagnait. . . L ' obsession du détail est plutôt  
rassurante dans l ' organisation de l ' Etat , mais il ne faut pas faire rentrer  
les Français dans la cuisine ,

estime un communicant.

off

Beaucoup estiment ainsi qu

off

' un sommet fut atteint lors de l ' interview menée par Jean-Jacques Bourdin  
, pour le deuxième anniversaire de son élection , et surtout du dialogue  
organisé à cette occasion entre le président et les auditeurs de RMC .

## Résultats de notre équipe de recherche



# Ressources

- [Article](#)
- [Tutoriel](#), avec codes ([notebook 1](#), [notebook 2](#)) et explications générales
- [Tutoriel pas à pas](#) (préparation de corpus, gestion des problèmes)
- Le site de l'équipe : [www.css.cnrs.fr](http://www.css.cnrs.fr)
- Et bientôt : Active Tigger en version libre