Augmented Journalism

WAN-IFRA Data Science Expert Group

12 December 2024

Le Télégramme

UMR IRISA

Soutenu par

MINISTÈRE DE LA CULTURE

Liberté
Égalité
Fraternité

A collaboration between

Xavier Marc, Data Engineer / DevOps
Yoann Péron, Data Science Manager
Guillaume Vasse, Director Numérique, Video and Media B2B

Laurent Amsaleg, CNRS senior researcher, Director of LinkMedia research team
Vincent Claveau, CNRS Researcher
Nicolas Fouqué, CNRS Research Engineer

financed by the Ministère de la Culture

Soutenu par

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

Table of contents

1. Presentation of Le Télégramme & Irisa
2. Origin and definitions
3. Modeling and Infrastructure
4. Current usage
5. Demo

Le Télégramme & Irisa

Le Télégramme

Distributed in **Bretagne**, west part of France

2nd most printed newspaper in France: 150 000 daily prints

800 000 daily digital visits

19 local editions + articles in breton

Groupe Télégramme also owns companies like HelloWork, OC Sports …

UMR IRISA

one of the largest French research center in computer science

1975
IRISA was established

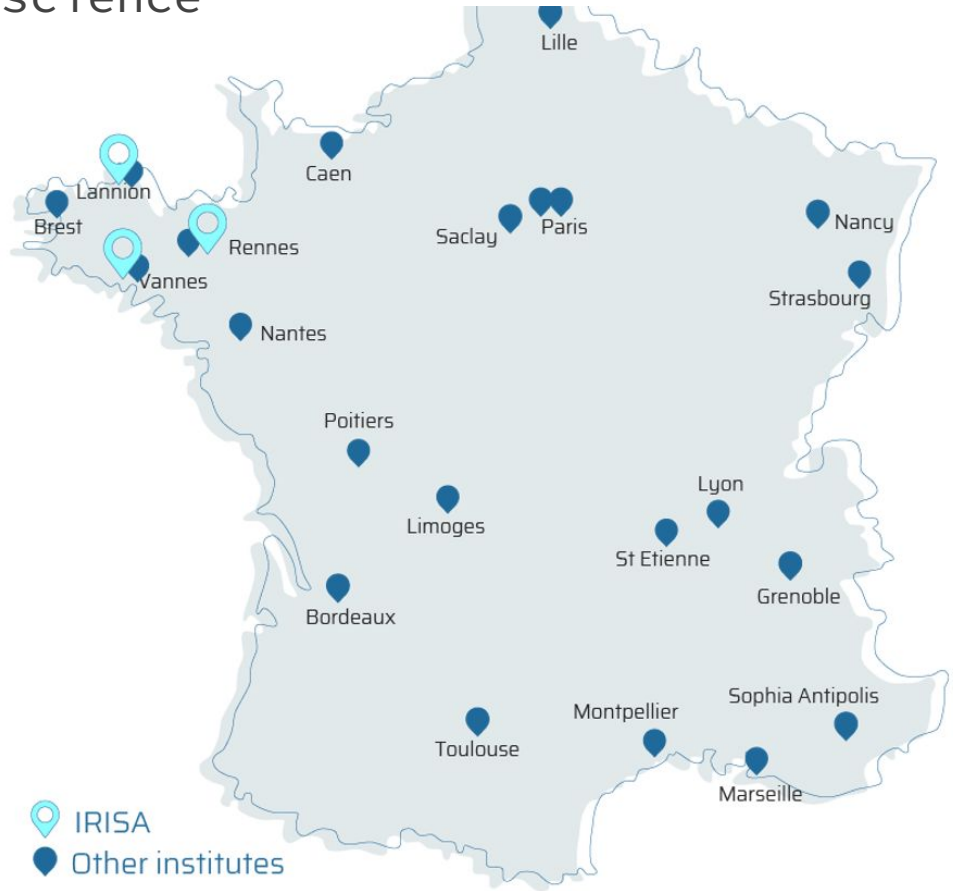
3
SITES in Brittany

+850
WOMEN and **MEN**

56
NATIONALITIES represented

9
TRUSTEES

Université de Rennes
ESS rennes
INSA
cnrs
Inria
Inserm
CentraleSupélec
IMT Atlantique
UBS



UMR IRISA

DEPARTMENTS

- Large Scale Systems
- Networks, telecommunication and services
- Architecture
- Language and software engineering
- Virtual reality, Virtual humans, interactions and robotics
- Signal, Image, Langage
- Data and knowledge management

~ 40
RESEARCH TEAMS

TRANSVERSES THEMES

CYBERSECURITY

ARTIFICIAL INTELLIGENCE

FRUGALITY & GREEN IT

ARTS, CULTURE & HERITAGE

BIOLOGY & HEALTH

ENVIRONMENT, ECOLOGY & AGRI/AGRO

SMART TERRITORIES

Origin and definitions

Origin of the project

Thanks to digital activities, we have a **feedback** on what read our readers, their behaviours, and therefore our work.

The project aimed at helping journalists during the writing of their articles:

- provide information about the future success of their articles
- provide suggestions on their articles
- propose new angles from that article

This last part has not been a success as we had difficulties to identify precisely what is an angle during the time of the project.

2-year project, during the release of ChatGPT

How to predict success?

First, how do we define success? Success depends on the usage and characteristics of the articles. A free article will not drive subscriptions, but can still be a success as it brings audience to our website.

We created **capacities** of an article. It represents how well it will perform on different parts of the conversion funnel. They are also defined to be able to act on the predictions of the capacity.

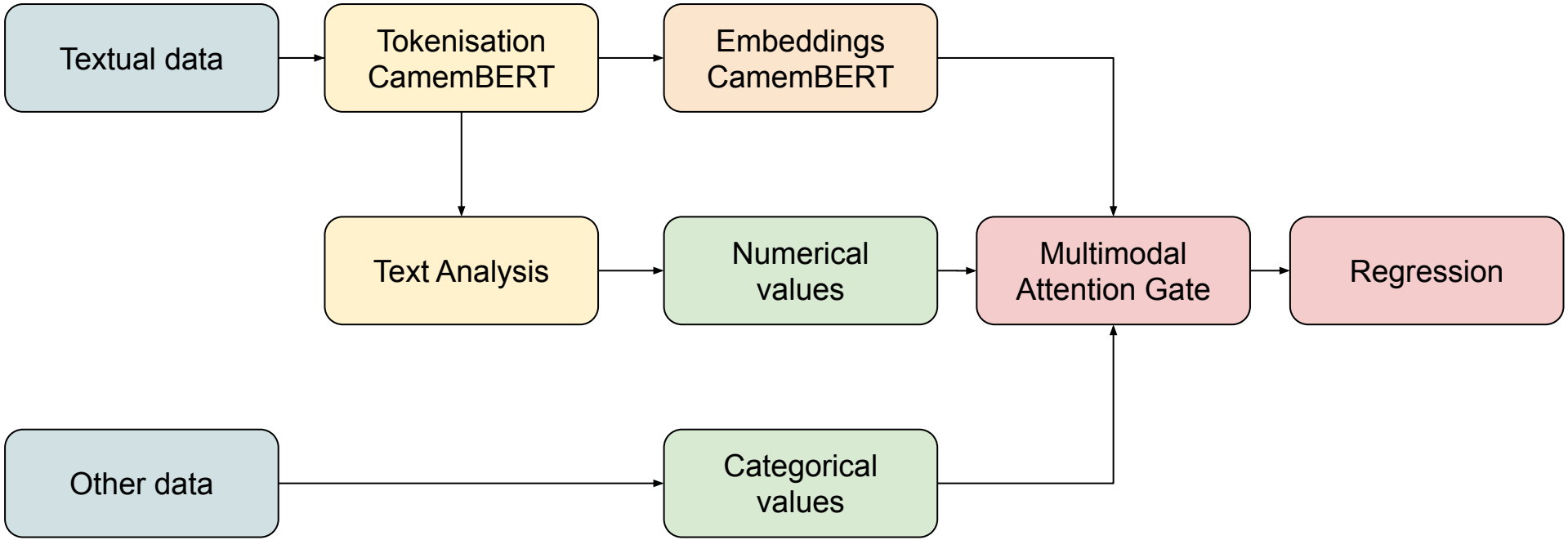
Audience: Number of page views on different times (first 96-hours, cold views)

Engagement of subscribers: click-rate on a newsletter

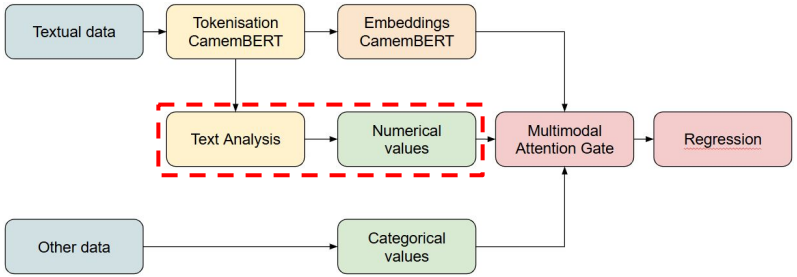
Engagement of younger audiences: reactions on social networks

Modeling & Infrastructure

Specific architecture for multimodal data



Text Analysis and Numerical data



Goal: improve models with additional data under numerical form

Named entity analysis: count the number of entities per type, normalised by the length of the article

Sentiment analysis: count the number of words attached to an emotion (joy, sadness, fear …), normalised by the length of the article

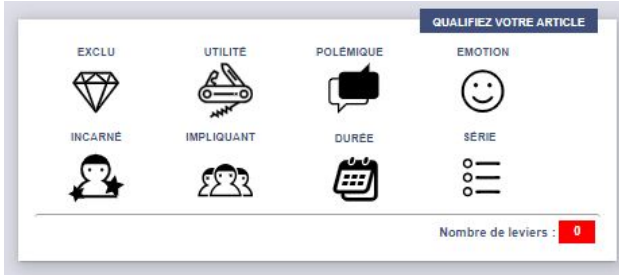
Uniqueness measure: based on similarity between articles, measure how unique an article is

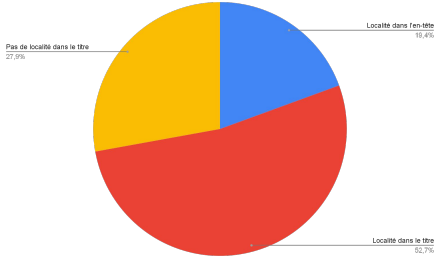
Length of article and number of **quotes**

Looking for improvements

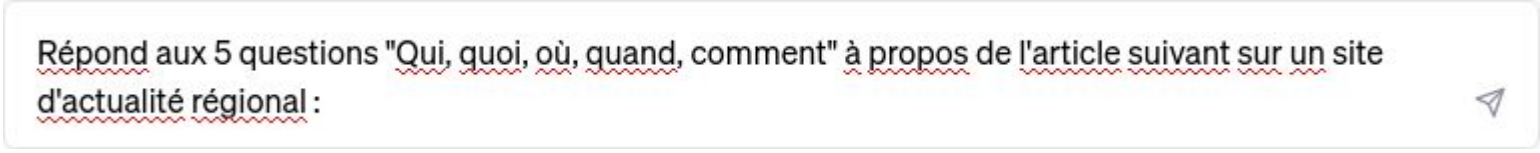
We have tested other inputs for our models

- use of journalist classification (levers)
- elements in title: presence of a verb, name of city at the beginning, question mark …





ChatGPT was also released during the time of the project. We used and compared it to our model for NER, definition of an angle … but ended with worse results than our home-models and were faced with hallucinations



How to explain our models

We experiment different methods for explaining our models. Ad-hoc methods (like SHAP and LIME) were tested.

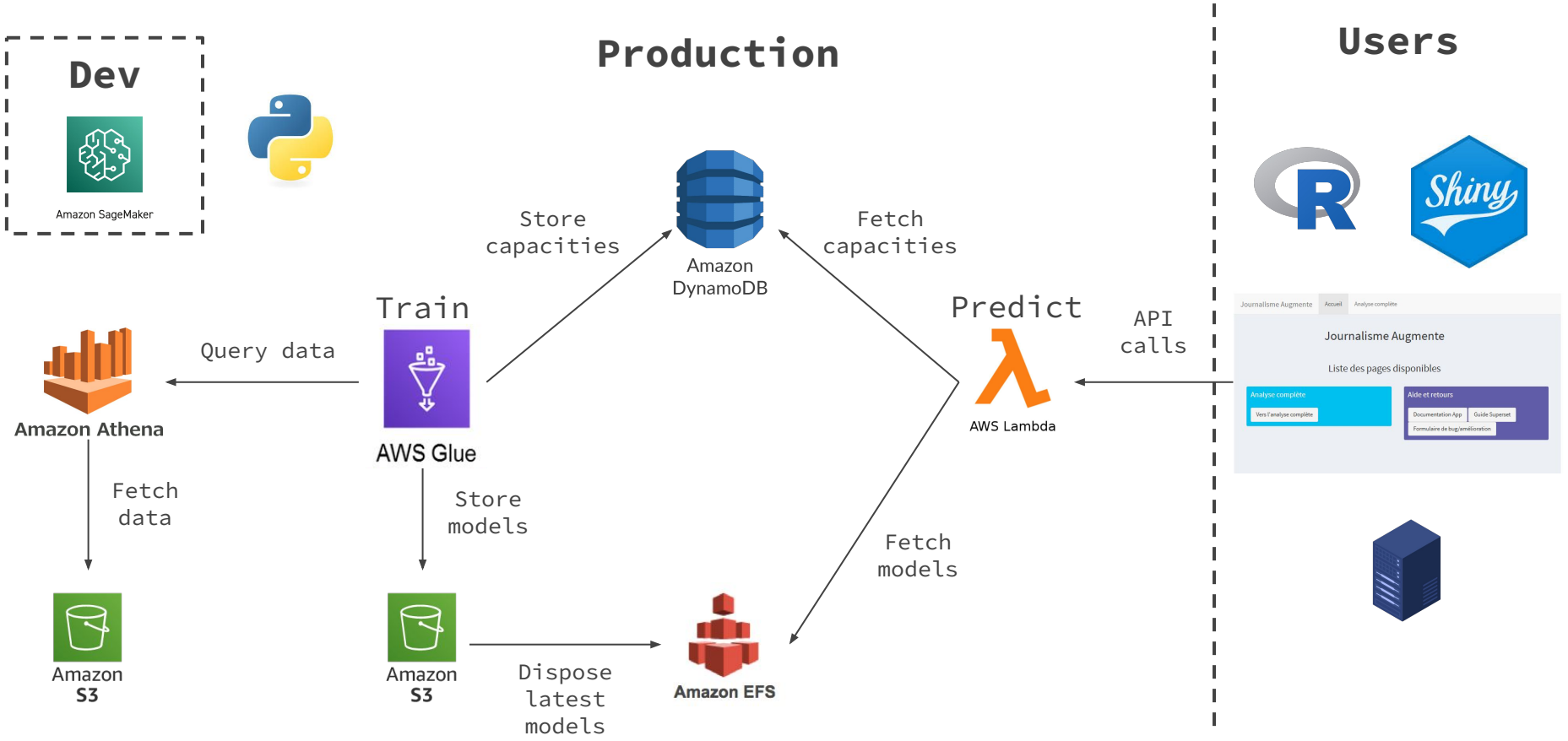
Advantages:

- Precised elements identified
- Compatibles with any models, even specific ones

Disadvantages:

- No guarantee on real usage by model
- Hard to read for non-specialists
- Costly

Infrastructure



Current usage

Inputs

« Je crois qu'après avoir vu ça, on peut mourir tranquille » : vos réactions à la victoire historique du Stade Brestois

Et le trophée de la meilleure vanne sur le Stade Brestois est attribué à...

Photo Le Télégramme/Nicolas Créach

Crispés ou euphoriques, les supporters du Stade Brestois aussi ont livré un beau spectacle ce mardi, sur le terrain des réseaux sociaux. Petite compilation des réactions les plus marquantes.

Manon Fontaine

Il y a ceux qui ont eu du mal à survivre à la deuxième mi-temps, ceux qui veulent ériger une statue à l'effigie de Marco Bizot, ceux qui trouvent ça « pas crovab' »... Pendant que le Stade Brestois nous régalait sur le terrain ce mardi 10 décembre 2024 en remportant 1 à 0 son match de Ligue des champions contre le PSV Eindhoven à Guingamp (22), vous nous avez régalés sur les réseaux sociaux avec vos plus beaux memes (*).

Title

Chapo

Text

Main section

Renseignez le titre

Renseignez le chapo

Renseignez le texte

Rubrique initiale

Bretagne

Aide à l'input

Similar articles for comparison

A single value (for example number of pages views) will not be useful. An article on a local subject can be a success compared to other similar articles.

We created several models of similar articles, based on different approaches.

- **TF-IDF**: based on words
- **NER**: based on named entities

	check	ID	Date de publication	Titre	Lien
1	☐	6553025	2024-03-27	À Penmarc'h, Sixun de retour sur scène après dix ans d'absence	Lien
2	☐	6523602	2024-02-11	Alain Souchon va repartir en tournée avec ses fils Ours et Pierre	Lien
3	☐	6463269	2023-11-06	The Fleshtones et Santa-Claws ont mis le feu à La Grande Ourse à Saint-Agathon	Lien
4	☐	6461398	2023-11-02	Découvrez la nouvelle chanson des Beatles, sortie avec l'aide de l'intelligence artificielle	Lien
5	☐	6443833	2023-10-07	La Passerelle à Saint-Brieuc : « La fin du début » au pays de l'absurde	Lien

These similar articles will also be used in recommender systems later next year on our website.

Predictions of capacities

We compare the prediction of a capacity with the capacities of known articles

Predicted capacity for current article

Known capacity for similar articles

Suggestions

For suggestions, we decided to compare the similar articles between themselves.

We classify them according to one capacity

We observe the characteristics of the best ones, and compare them to the characteristics of the worst ones.

We display only the characteristics that are different in both groups. If no difference is significant, we prefer to not display anything.

Rubriques dans les articles similaires

Show 5 entries Search:

	check	rubrique
1	☐	tb:www0f:local:ille-et-vilaine:rennes
2	☐	tb:www0f:culture-loisirs
3	☐	tb:www0f:local:cotes-d-armor
4	☐	tb:www0f:fil-info:bretagne

Demo

What's next?

This app is available for all journalists, but we need to improve the inference time before an implementation in Swing, on the ObjectPanel.

Other improvements are possible

1) Improve models: we follow performances of our models and observe a lesser performance on the click-rate model
2) Add conversions: has the article the same characteristics as articles which were part of a conversion path?
3) Help on suggestions: characteristics of a title

Thank you for your attention

Le Télégramme

UMR IRISA

Soutenu par

MINISTÈRE DE LA CULTURE

Liberté
Égalité
Fraternité