Detecting Green Washing in French Text

FANG Tian, Elisabeth Olamisan,MEDJAED Aubin, Ahana Chattopadhyay

University of Lorraine
IDMC

February 18, 2025

GitHub Repository

Overview

- **Introduction:** Greenwashing definition and challenges.
- **Dataset:** Translation of ClimateBERT dataset, manual claim collection.
- **Methodology:** Fine-tuning Qwen models, NER masking experiment.
- **Observations:** Class imbalance, masked vs. unmasked data effects.
- **Results:** Model evaluation, performance comparison.
- **Ethical Considerations:** Environmental impact, fairness concerns.
- **Future Work:** Dataset improvement, better fine-tuning approaches.

What is Greenwashing?

Definition:
"Greenwashing—the act of misleading consumers about the environmental practices of a company or the environmental benefits of a product or service—is a pressing issue that requires robust detection methods across different languages. [United Nations Environment Programme (UNEP), 2023] "

ADEME Dataset

Emissions publication P1.1	Emissions publication P1.2	Emissions publication P1.3	Emissions publication P1.4	Emissions publication P1.5	Emissions publication P2.1
1049	2668			3379	2605
0	11308	0	27	0	735
279.2	12.5	0	27.4	0	36.5
0	295.73	0	84.17	0	48.4
0	0	0.05	0	0	13
					0.983
1459.37	60.86		61.31		259.73
9	11.7	149.5			
0	0.1	0	0	0	0
883.58	49.33		77.26		99.64
0.78	0.67	2.6	0	0	0.84

Our Approach: Fine-Tuning Qwen for Greenwashing Detection

Approaches

- Utilize **cross-lingual transfer learning** by adapting models trained on English greenwashing detection to French.
- Leverage **full fine-tuning** instead of adapter-based methods.

Methodology

- **Model Selection:** Fine-tune models from the **Qwen family**
- **Data Preparation:**
  - Translate the **ClimateBERT dataset** from English to French to form training, validation, and testing datasets.
- **Training Strategy:**
  - Fine-tune with/without PEFT
- **Evaluation:**
  - Collect **real-world greenwashing claims** for evaluation.
  - Compare model performance on both **translated dataset evaluation** and **manually curated claims**.

[4]

LE GROUPE SNCF EN 2023

1. PROFIL DU GROUPE SNCF

1.1 RAISON D'ÊTRE

SNCF, un leader mondial de la mobilité durable.

SNCF, 4 lettres pour une entreprise qui depuis 80 ans, accompagne la vie quotidienne de ses clients. Parfois critiquée mais toujours présente, SNCF est un patrimoine partagé avec l'ensemble des Français qui ont tous une histoire de train, une histoire en commun avec nous.

Pour autant, SNCF ne s'enferme pas dans son passé. Car si le train est le premier mode historique de transport sur longue distance, il est surtout un mode d'avenir, aux atouts évidents. Le ferroviaire est intégrateur du territoire et contribue à l'égalité des chances. Il propose le mode de déplacement le plus efficace sur un plan énergétique, garant de sobriété. Il minimise la consommation d'espace et les nuisances. Il est objet de prestige, vitrine du savoir-faire français et de son expertise technique sur la grande vitesse et le transport massifié, le fret ferroviaire et la gestion des infrastructures, les grandes dessertes européennes et les métros automatiques en France ou à l'autre bout du monde.

Surtout, SNCF est un groupe pétri de la fierté de ses 283 000 salariés, qui ont à cœur de proposer l'offre de transport la plus juste, sociale et écologique. C'est grâce à eux, au quotidien, que le Groupe peut réaffirmer son souhait de construire un monde meilleur et d'agir pour une société en mouvement, solidaire et durable.

Notre raison d'être

« Le Groupe SNCF a pour mission de contribuer à la vitalité de la société et de ses territoires. Nous offrons des solutions de mobilité et de logistique du 21e siècle, innovantes et centrales pour la décarbonation des transports.

La performance et l'intégration de l'ensemble de nos métiers autour du ferroviaire visent à optimiser le coût et l'impact global des transports pour les clients, les contribuables et les citoyens. Nos infrastructures et nos services, qui s'inscrivent dans le temps long, constituent un bien commun pour relever les défis sociaux, écologiques et économiques, et ainsi agir pour une société en mouvement, solidaire et durable. »

Nos 8 engagements

– Répondre aux nouveaux rythmes de vie et à l'évolution des modes de consommation et de production, en facilitant la combinaison des modes de transport.

– S'engager au quotidien pour améliorer les fondamentaux de la qualité de service, en y associant l'ensemble de nos clients, fournisseurs et partenaires.

– Investir dans les métiers et les compétences d'avenir pour faciliter l'insertion professionnelle, l'ascenseur social et les parcours de reconversion.

– Garantir la soutenabilité économique de nos activités, dans l'intérêt du système ferroviaire et de sa performance globale.

– Contribuer au dynamisme économique et social des territoires par nos décisions industrielles et le choix de nos fournisseurs.

– Renforcer notre rôle de catalyseur dans la transition

Preliminary Actions Taken

- **Dataset Preparation:**
  - Translated claims from the original ClimateBERT dataset into French using the DeepL API.
  - Saved the translated dataset for validation and analysis.
- **Model Validation:**
  - Tested the English ClimateBERT model on the translated French dataset.
  - Evaluated performance using classification metrics (accuracy, precision, recall, and F1-score).
- **Data Analysis:**
  - Analyzed dataset properties, including token and sentence counts.
  - Identified significant class imbalance impacting model performance.

Observations - English ClimateBERT dataset

Observations:

- **Severe Class Imbalance:**
  - Class 0 (non-claim): 75% of data (1,585 instances).
  - Class 1 (claim): 25% of data (532 instances).
- **Model Bias:**
  - The model favors the majority class, leading to excellent performance for Class 0 but poor detection of Class 1.
- **F1-Score Disparity:**
  - Class 0: Strong F1-Score of 86%.
  - Class 1: Weak F1-Score of 4%.

Guidelines for the selection of Greenwashing Claims

To begin with, we consulted the ADEME Anti Greenwashing Guide in order to gain more insight into this. They list 9 signs of greenwashing:

1. A blatant lie
2. An exaggerated promise
3. Vague words
4. Insufficient information
5. A misleading image
6. A fake label or false endorsement
7. An off-topic emphasis
8. Nonexistent evidence
9. A false exclusivity

Guidelines for the selection of Greenwashing Claims

We finally choose the following characteristics [5] to identify and to manually collect the potential greenwashing claims from websites, sustainability reports and Internet search in general.

1. Absence of explicit climate-related commitments and actions
2. Use of non-specific language
3. overly optimistic sentiment
4. lack of evasive or hedging terms

French Enviromental Report Claims Dataset (FER-C)

The manually curated Greenwashing and Non-greenwashing datasets were merged and the merged dataset was named as French Enviromental Report Claims Dataset (FER-C). Here's a brief description of the dataset:

- Columns
  - Source (link)
  - Text (string)
  - Label (0 or 1)
- Size and Content
  - Initial size : 157 greenwashing, 3 non-greenwashing
  - Final size : 200 greenwashing, 300 non-greenwashing
- Example
  - "Le Groupe s'est engagé à intégrer les enjeux climat dans sa stratégie [...]."

NER Masking

The translated dataset having almost 2k sentences contains company names in many sentences. Therefore, we decided to perform Named Entity Recognition masking of our dataset. Here are some reasons why NER masking is useful in this case:

- **Prevents Bias**
  - Avoids associating specific company names with greenwashing instead of analyzing actual language patterns.
- **Improves Generalization**
  - Ensures the model works well with unseen companies and industries.
- **Enhances Fairness**
  - Prevents the model from unfairly favoring or targeting specific companies.
- **Prevents Data Leakage**
  - Stops the model from relying on company names instead of linguistic features.

BLEU and Cosine Similarity Analysis

- **Objective:** Evaluate translation quality using BLEU score and cosine similarity.
- **Datasets:** Greenwashing dataset and translated claims validation dataset.
- **BLEU Score Calculation:**
  - Used `nltk`'s `corpus_bleu` for BLEU score.
  - References and candidates extracted and aligned.
  - Applied smoothing function for better results.
- **Findings:** BLEU score computed for the translated dataset is 0.0011 and for the masked dataset is 0.0035.

Cosine Similarity Computation

- **Vectorization:** Used `TfidfVectorizer` from `sklearn`.
- **Cosine Similarity:**
  - Computed between original and translated texts.
  - Measures semantic similarity between text pairs.
- **Findings:** Cosine similarity computed for the translated dataset is 0.23899 and for the masked dataset is 0.8803.
- **Conclusion:** In this case, cosine similarity is more suitable because BLEU fails when comparing texts in different languages since word order and phrasing differ significantly. Instead, semantic similarity (e.g., cosine similarity with Sentence-BERT) is a better approach.

Manual checking of the datasets

Firstly, the manually curated dataset as discussed in the slide 11 was randomized. The randomization is important for the following reasons:

- **Avoid bias:** Prevents the model from learning patterns based on order instead of content.
- **Improves generalization:** Helps the model perform better on unseen data.

Thereafter, a thourough manual checking was done by one of our team member who is a native French speaker.

- The translated dataset was checked manually to estimate the quality of the automatic translation. The task was especially challenging since there were plenty of financial terminology and corporate lingo.
- The merged dataset was meticulously reviewed to ensure that the greenwashing statements were accurately identified.

Training

The better performance of models without NER might be due to their learning of the clear association between certain company names and their greenwashing behavior. However, this does not exhibit good generalization ability, which is why models that use NER to mask company names may perform similarly to those that don't on the test dataset.

<table>
<tr><th>Model</th><th>Image</th></tr>
<tr><td>0.5B unmasked</td><td>
[136/165 20:46 < 04:29, 0.11 it/s, Epoch 4/5]
<table>
<tr><th>Epoch</th><th>Training Loss</th><th>Validation Loss</th><th>Accuracy</th></tr>
<tr><td>1</td><td>No log</td><td>0.613016</td><td>0.656604</td></tr>
<tr><td>2</td><td>1.737800</td><td>0.276957</td><td>0.890566</td></tr>
<tr><td>3</td><td>0.165900</td><td>0.368584</td><td>0.883019</td></tr>
<tr><td>4</td><td>0.165900</td><td>0.569220</td><td>0.879245</td></tr>
</table>
</td></tr>
<tr><td>0.5B masked</td><td>
[170/264 24:11 < 13:32, 0.12 it/s, Epoch 5/8]
<table>
<tr><th>Epoch</th><th>Training Loss</th><th>Validation Loss</th><th>Accuracy</th></tr>
<tr><td>1</td><td>No log</td><td>2.423565</td><td>0.750943</td></tr>
<tr><td>2</td><td>1.528200</td><td>2.481865</td><td>0.750943</td></tr>
<tr><td>3</td><td>0.107200</td><td>2.827964</td><td>0.750943</td></tr>
<tr><td>4</td><td>0.107200</td><td>3.146985</td><td>0.750943</td></tr>
<tr><td>5</td><td>0.022600</td><td>3.266309</td><td>0.750943</td></tr>
</table>
</td></tr>
</table>

Table: Comparison of 0.5B Models

Training

Model	Image			
	Epoch	Training Loss	Validation Loss	Accuracy
	1	No log	0.631473	0.694340
	2	0.722500	0.626848	0.713208
	3	0.709300	0.622548	0.713208
	4	0.709300	0.619015	0.716981
	5	0.697800	0.616853	0.705660
	6	0.705400	0.615124	0.705660
	7	0.705400	0.614020	0.709434
	8	0.702200	0.613463	0.716981
1.5B unmasked	9	0.694200	0.613282	0.716981
	Epoch	Training Loss	Validation Loss	Accuracy
	1	No log	1.231015	0.750943
	2	0.818600	1.172416	0.750943
	3	0.731500	1.139798	0.750943
	4	0.731500	1.119373	0.750943
1.5B masked	5	0.695300	1.114595	0.750943

Table: Comparison of 1.5B Models

Testing and evaluation

The testing was done on both the datasets. Here are the results based on the following evaluation metrics:

Model	Dataset	Accuracy	Recall	F1 Score
0.5B masked	FER-C	0.7445	0.5628	0.6364
	Translated	0.9057	0.8358	0.8175
0.5B unmasked	FER-C	0.7445	0.5528	0.6322
	Translated	0.9019	0.8358	0.8116
1.5B masked	FER-C	0.5868	0.0503	0.0881
	Translated	0.7019	0.0597	0.0920
1.5B unmasked	FER-C	0.5788	0.4322	0.4491
	Translated	0.6943	0.4478	0.4255

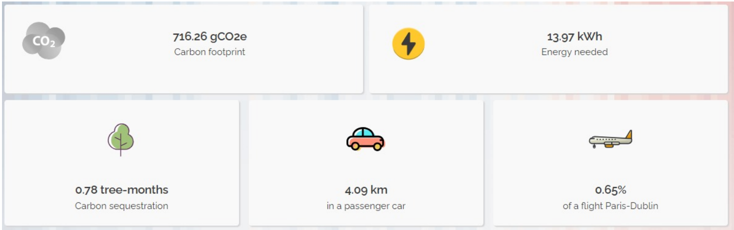
Figure: Testing Phase Results

Observations and Possible Explanation

- **The translated test dataset performs better than FER-C:** The non-greenwashing part of FER-C is mostly related to environmental issues, while the part of the translated test dataset is less related to the environment.
- **The small model (0.5B), whether using NER or not, demonstrated better performance:**
  - The size of the 1.5B model may not be sufficient to fully capitalize on the benefits brought by LoRA(increasing the rank size improved the model performance), but without using LoRA, it would run out of memory.
  - Smaller models, due to their relatively fewer parameters, may be better suited for handling less complex datasets.[3]

Carbon footprint and Ethical concerns

The amount of carbon footprint generated due to our experiment was calculated using this Online Carbon footprint calculator.



Carbon footprint

As for the ethical concerns, it is complex to define greenwashing.

Future Works

Constraint on time and resources :

- Training larger or more performing models
- Training past overfitting

Dataset quality :

- help from experts
- increasing size or quality

Other methods :

- Few Shot Learning already proved efficient. [2]
- Active learning, Prompt engineering.

Big interest on Greenwashing in the NLP sphere [1]

Reference

[1] Tom Calamai, Oana Balalau, Théo Le Guenedal, and Fabian M. Suchanek. Corporate greenwashing detection in text - a survey, 2025.

[2] Abhijeet Kumar Mridul Mishra Mayank Singh, Nazia Nafis. Measuring sustainability intention of esg fund disclosure using few-shot learning, 2025.

[3] Kangfu Mei, Zhengzhong Tu, Mauricio Delbracio, Hossein Talebi, Vishal M. Patel, and Peyman Milanfar. Bigger is not always better: Scaling properties of latent diffusion models, 2024.

[4] Unknown. Measuring greenwashing: the greenwashing severity index. *SSRN Electronic Journal*, 2023.

[5] Avalon Vinella, Margaret Capetz, Rebecca Pattichis, Christina Chance, and Reshmi Ghosh. Leveraging language models to detect greenwashing. *arXiv preprint*, 2311.01469, 2023.