

ADMANE Tania
MANCHE Giovanni
M2 IEF Quantitative Finance

Proposal submission - Natural Language Processing project

Our proposal is to train an NLP classifier capable of distinguishing whether a given economic text (article, report, speech, etc.) originates from a real economic figure (such as the ECB, the FED, economist columns, or academic papers) or from an AI-generated source. This project lies at the intersection of language analysis and economic reasoning, and directly relates to current concerns regarding disinformation and the authenticity of sources.

Our model will aim to detect whether an economic text is authentic or AI-generated, primarily by analyzing writing style, structure, and the nature of the expressed opinions. We will use two main data sources :

- For real economic texts : French and English articles, official documents from institutions like the ECB and the FED, and opinion columns from well-known economists.
- For AI-generated texts : A large set of synthetic economic texts created by LLMs, designed to sound plausible (they will not say absurd economic things on purpose) but still generated, in order to challenge the model's ability to recognize subtle stylistic differences.

One of our main challenges will be to ensure a balanced and representative dataset for both classes, and to avoid overfitting to superficial cues such as formatting or specific tokens. Another key difficulty lies in identifying and modeling the boundary between realism and artificiality, especially when AI mimics sophisticated economic reasoning. Within the group, responsibilities are initially distributed as follows :

- One member will be in charge of data collection and preparation.
- Another will focus on setting up a baseline model for comparison.
- Both will collaborate on the development and tuning of more advanced models.

We are enthusiastic about exploring this original and timely topic and look forward to your feedback. Thank you !

Giovanni MANCHE and Tania ADMANE