# 💰 Trillion Dollar Words in Julia

JuliaCon 2024

Patrick Altmeyer

Thursday, July 11, 2024

# TrillionDollarWords.jl

![docs stable](https://img.shields.io/badge/docs-stable-blue) ![docs dev](https://img.shields.io/badge/docs-dev-blue) ![CI passing](https://img.shields.io/badge/CI-passing-green) ![codecov 88%](https://img.shields.io/badge/codecov-88%25-yellow) ![code style blue](https://img.shields.io/badge/code%20style-blue-blue)

A light-weight package providing Julia users easy access to the Trillion Dollar Words dataset and model (Shah, Paturi, and Chava 2023).

> **Disclaimer**
>
> Please note that I am not the author of the Trillion Dollar Words paper nor am I affiliated with the authors. The package was developed as a by-product of our research and is not officially endorsed by the authors of the paper.

# Context

ICML 2024 paper *Stop Making Unscientific AGI Performance Claims* (Altmeyer et al. 2024) (preprint, blog post, code):

- Even simple models can distill meaningful information that predicts external data.

- Humans are prone to seek patterns and anthropomorphize.

# Basic Functionality

The package provides the following functionality:

- Load pre-processed data.

- Load the model proposed in the paper.

- Basic model inference: compute forward passes and layer-wise activations.

- Download pre-computed activations for probing the model.

# Loading the Data

## Sentences

40,000 time-stamped sentences from

- meeting minutes

- press conferences

- speeches

by members of the Federal Open Market Committee (FOMC):

```
1  using TrillionDollarWords
2  load_all_sentences() |>
3    x -> names(x)
```

```
8-element Vector{String}:
 "sentence_id"
 "doc_id"
 "date"
 "event_type"
 "label"
 "sentence"
 "score"
 "speaker"
```

## All Data

Merged data includes economic indicators

- Consumer Price Index (CPI)

- Producer Price Index (PPI)

- US Treasury (UST) yields

```
1  load_all_data() |>
2    x -> names(x)
```

```
11-element Vector{String}:
 "sentence_id"
 "doc_id"
 "date"
 "event_type"
 "label"
 "sentence"
 "score"
 "speaker"
 "value"
 "indicator"
 "maturity"
```

# Loading the Model

- Can be loaded with or without the classifier head.

- Uses Transformers.jl to retrieve the model from HuggingFace.

- Any keyword arguments accepted by `Transformers.HuggingFace.HGFConfig` can also be passed.

```
1  load_model(; load_head=false, output_hidden_states=true)
```

# Basic Model Inference

## From Scratch

Layer-wise activations can be computed as follows:

```
1  df = load_all_sentences()
2  mod = load_model(
3    load_head=false,
4    output_hidden_states=tru
5  )
6  n = 5
7  queries = df[1:n, :]
8  layerwise_activations(
9    mod, queries
10 )
```

## From Artifacts

We have archived activations for each layer and sentence as artifacts:

```
1  using LazyArtifacts
2
3  artifact"activations_layer
```

OK, but why would I need all this? 🤔

# Motivation

- $A_1$: $enc($„It is essential to bring inflation back to target to avoid drifting into deflation territory.")

- $A_2$: $enc($„It is essential to bring the numbers of doves back to target to avoid drifting into dovelation territory.")

# Motivation

- $A_1$: $enc($ „It is essential to bring inflation back to target to avoid drifting into deflation territory." $)$

- $A_2$: $enc($ „It is essential to bring the numbers of doves back to target to avoid drifting into dovelation territory." $)$

> "They're exactly the same."
>
> — Linear probe $\widehat{cpi} = f(A)$
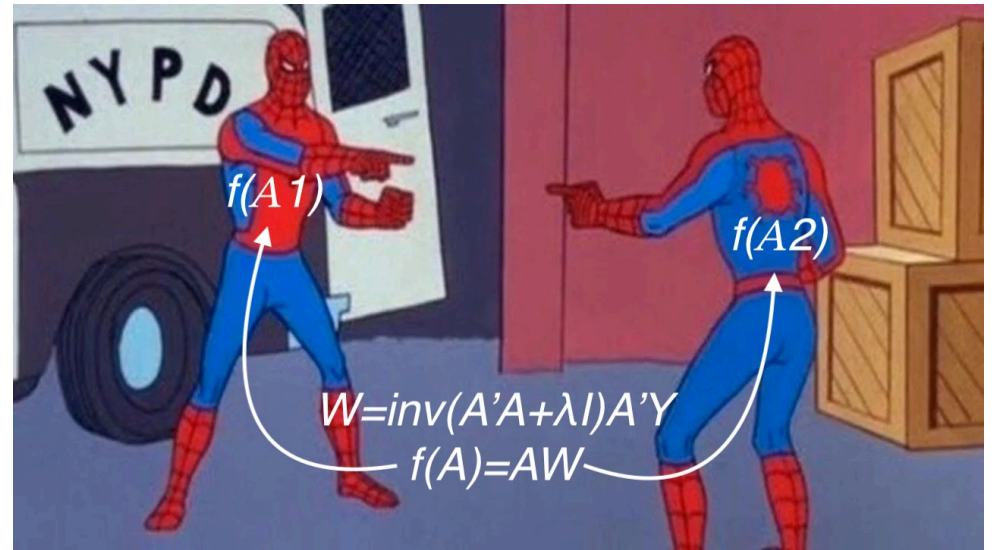
# Embedding FOMC comms

- We linearly probe all layers to predict unseen economic indicators (CPI, PPI, UST yields).

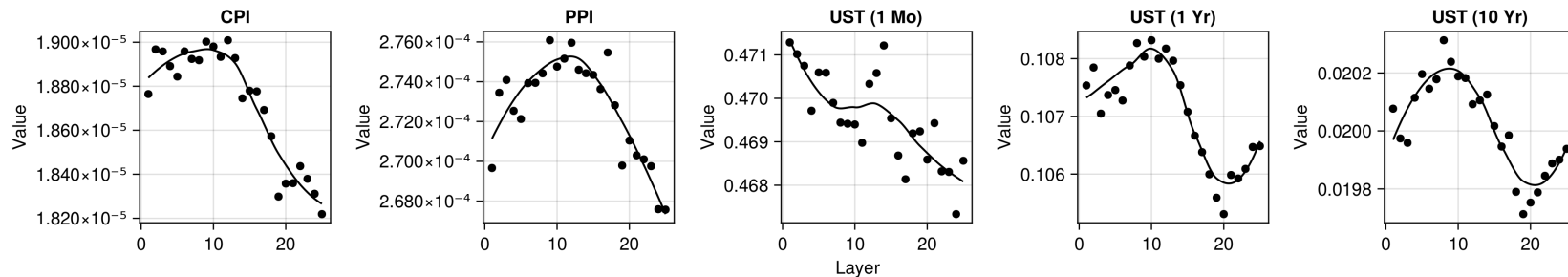- Predictive power increases with layer depth and probes outperform simple AR($p$) models.



Figure 1: Out-of-sample root mean squared error (RMSE) for the linear probe plotted against FOMC-RoBERTa's n-th layer for different indicators.

# Sparks of Economic Understanding?

If probe results were indicative of some intrinsic 'understanding', probe should not be sensitive to unrelated sentences.
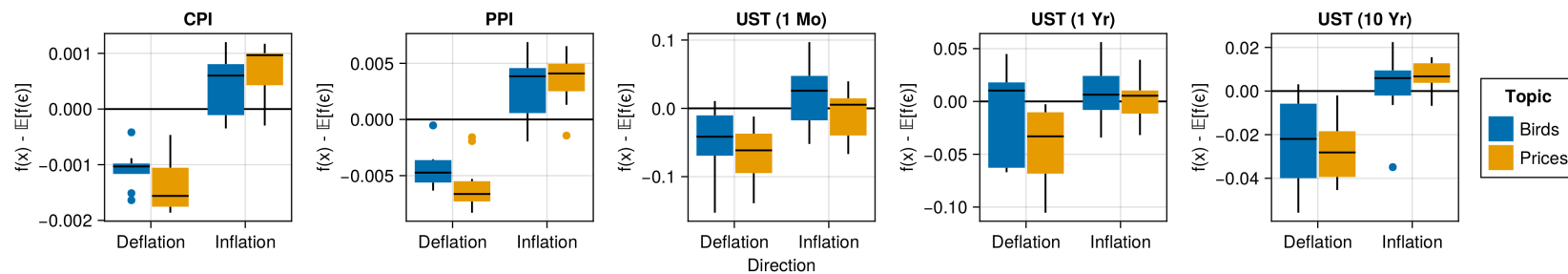


Figure 2: Probe predictions for sentences about inflation of prices (IP), deflation of prices (DP), inflation of birds (IB) and deflation of birds (DB). The vertical axis shows predicted inflation levels subtracted by the average predicted value of the probe for random noise.

# Intended Purpose and Goals

Good starting point for the following ideas:

- Fine-tune additional models on the classification task or other tasks of interest.

- Further model probing, e.g. using other market indicators not discussed in the original paper.

- Improve and extend the label annotations.

Any contributions are very much welcome.

# Questions?

With thanks to my co-authors Andrew M. Demetriou, Antony Bartlett, and Cynthia C. S. Liem and to the audience for their attention.

# References

Altmeyer, Patrick, Andrew M. Demetriou, Antony Bartlett, and Cynthia C. S. Liem. 2024. "Position: Stop Making Unscientific AGI Performance Claims." https://arxiv.org/abs/2402.03962.

Shah, Agam, Suvan Paturi, and Sudheer Chava. 2023. "Trillion Dollar Words: A New Financial Dataset, Task & Market Analysis." *arXiv Preprint arXiv:2310.02207v1*. https://arxiv.org/abs/2305.07972.

# Image sources

- Leonardo DiCaprio: Meme template by user on Reddit

# Quote sources

- "There! It's sentient"—that engineer at Google (probably!)