CSSMethods Talk 03/2025: Synthetic Surveys for Population Insights

Max Pellert (https://mpellert.at)

[PDF]

Classical traditional surveys

Beloved standard tool of the social sciences

Often considered the gold standard, “ground truth” (especially when working with large representative samples of a population)

But, classical survey methodologies increasingly suffer from problems

First line of the 2024 Book “Polling at a Crossroads: Rethinking Modern Survey Research”: *Survey research is in a state of crisis*

Last example: US presidential elections 2024

(Generally, I think strong method conservatism in the social sciences does not make sense)

‘Queen of polling’ J Ann Selzer quits after Iowa survey missed by 16 points

Pollster announces she’s moving on ‘to other ventures’ after poll wrongly predicted strong shift in state to Kamala Harris

?

Ann Seltzer had an excellent multi-decade track record of accurate polling

For example: In 2008, predicting that a virtually unknown senator, Barack Obama, would beat frontrunner Hillary Clinton in the Iowa caucuses

The widely publicized final poll of Iowa by Selzer & Company showed Harris leading by 3 percentage points (in Iowa!)

On Election Day, Trump won the state by 13 points

Laudable public effort by the pollster at an error analysis: “To cut to the chase, I found nothing to illuminate the miss.”

In defence

“Within the margin of error”, “We said it’s close”, “Predicting it at almost 50-50 means that this can happen”

The 2024 elections were not close, a decisive victory on all metrics

Relevance of survey research? You don’t need much sophisticated machinery (or money) to predict that a 2-party system as the US with deeply ingrained political beliefs of the population and a very peculiar electoral system will be a tight race

Main issue? Non-response

Less than 1% of people respond even in respected well-established surveys (NYT for example)

Non-response

By now, any change in the traditional polls may just mean a new pattern of non-response

Many reasons, spam calling and ping calls a recent one

Statistics can correct for some problems, but you need some basis to work from

At the same time that they don’t respond to surveys, people are extremely expressive on other, social media

We have massive amounts of text available: found data, digital traces or whatever you like to call it

Synthetic Surveys

Britain's mood, measured weekly

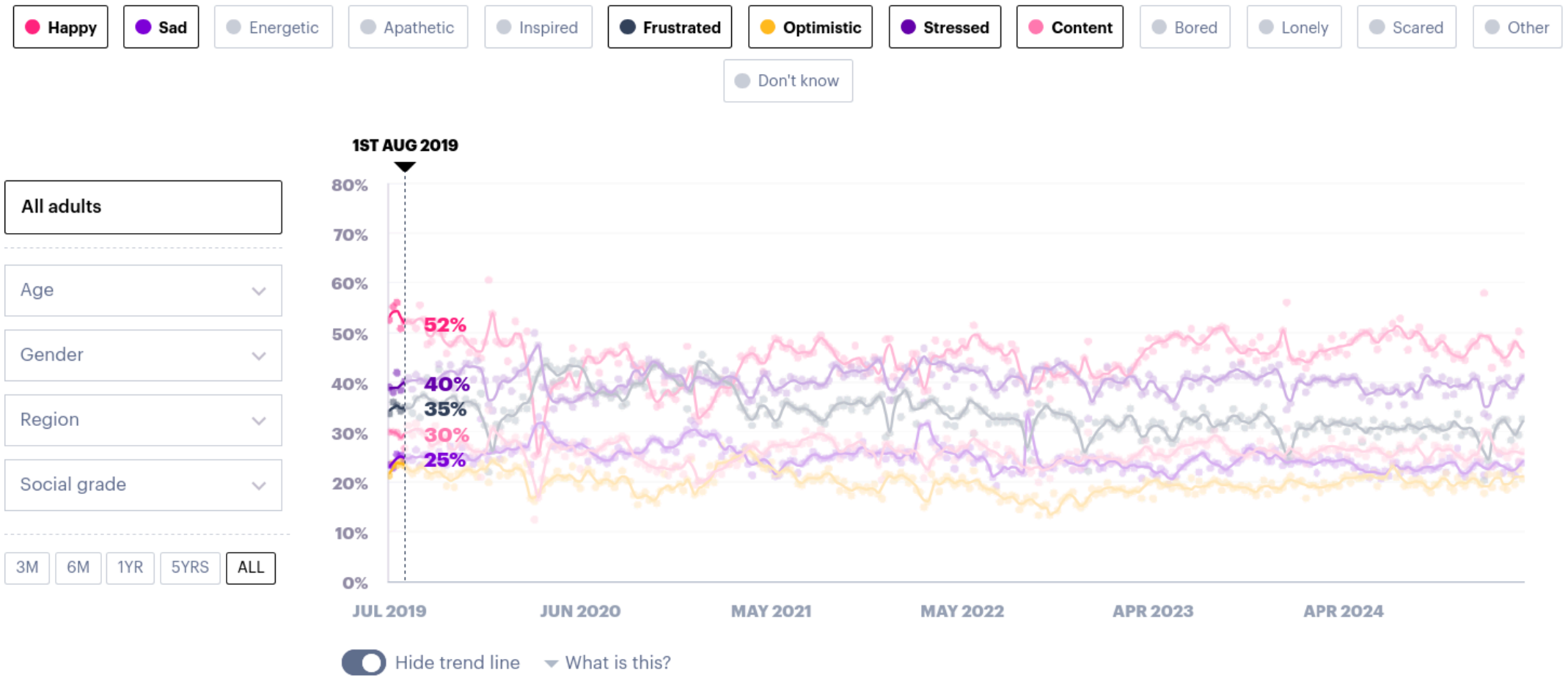
One example of an easily accessible, representative survey (UK) not directly in the political domain

https://yougov.co.uk/topics/politics/trackers/britains-mood-measured-weekly

Ahnert, G., Pellert, M., Garcia, D., & Strohmaier, M. (2024). Britain's Mood, Entailed Weekly: In Silico Longitudinal Surveys with Fine-Tuned Large Language Models. Companion Proceedings of the 16th ACM Web Science Conference, 47–50. https://doi.org/10.1145/3630744.3659829

Ahnert, G., Pellert, M., Garcia, D., & Strohmaier, M. (2025). Extracting Affect Aggregates from Longitudinal Social Media Data with Temporal Adapters for Large Language Models (arXiv:2409.17990). arXiv. https://doi.org/10.48550/arXiv.2409.17990

Britain's mood, measured weekly



FULL QUESTION

Broadly speaking, which of the following best describe your mood and/or how you have felt in the past week Please select all that apply

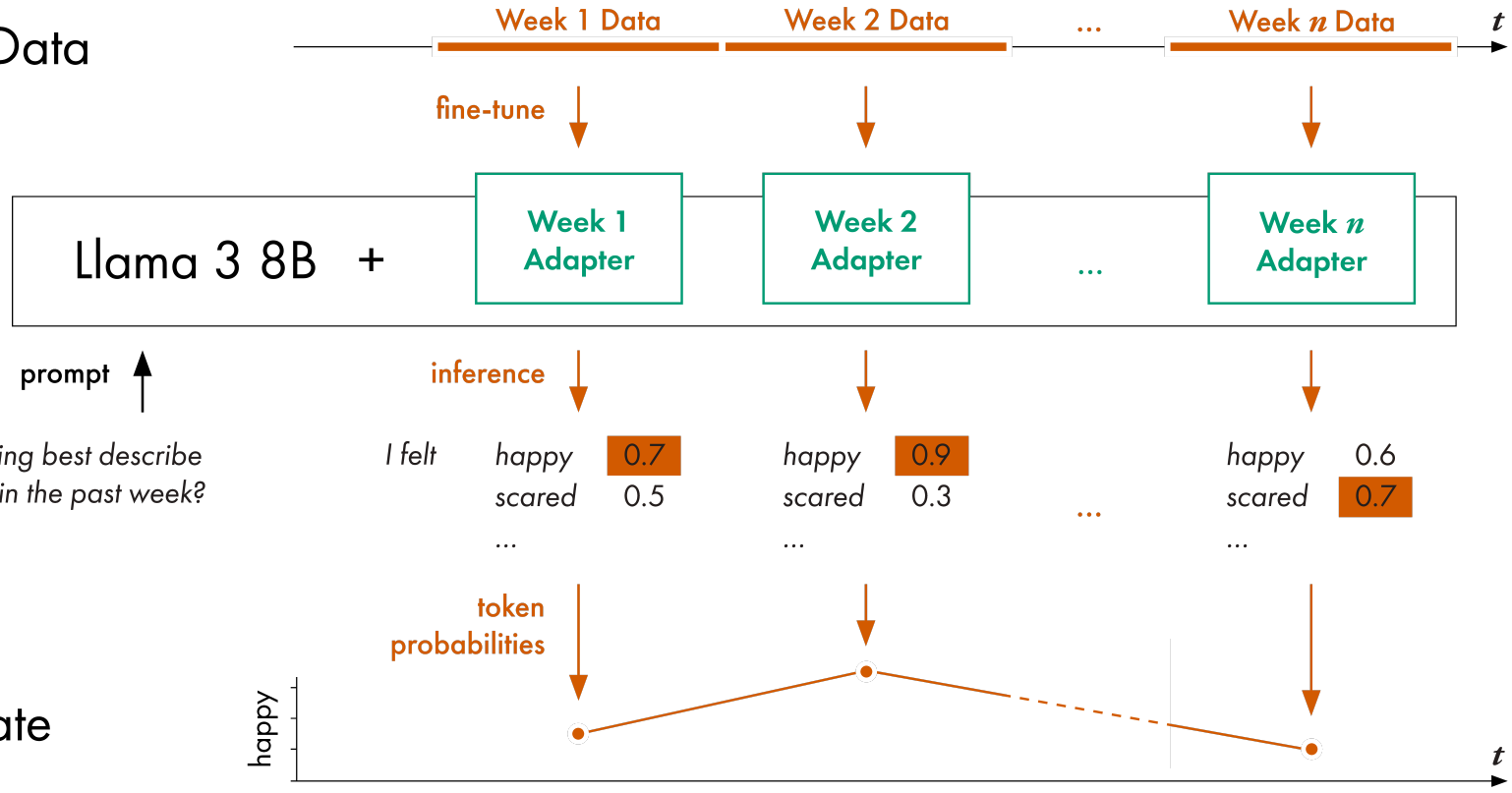
Weekly Social Media Data

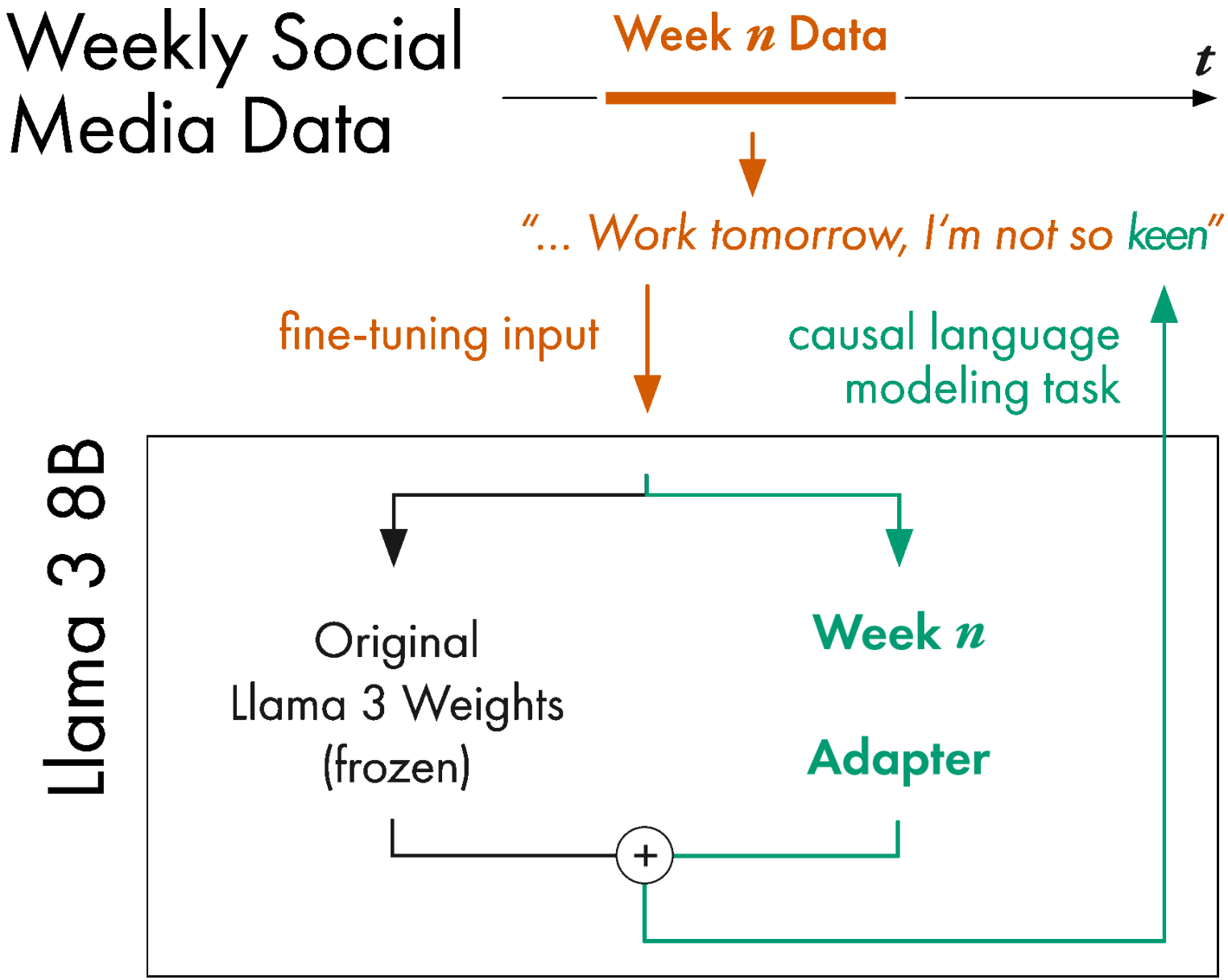
Temporal Adapters

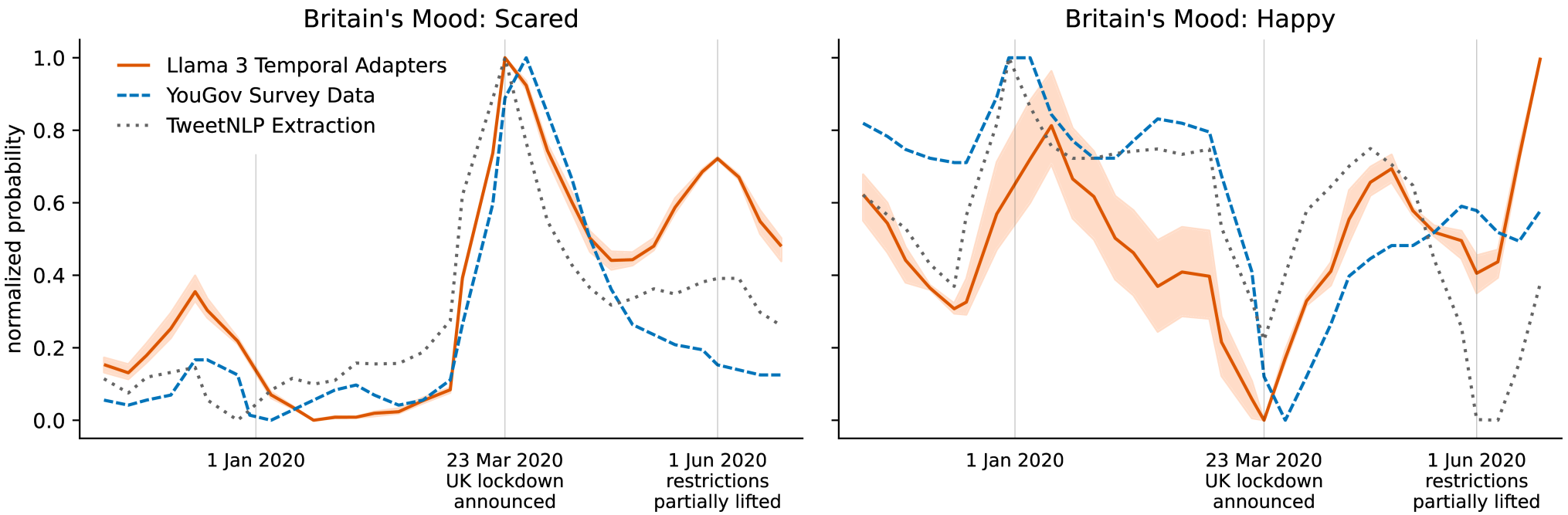
Survey Question

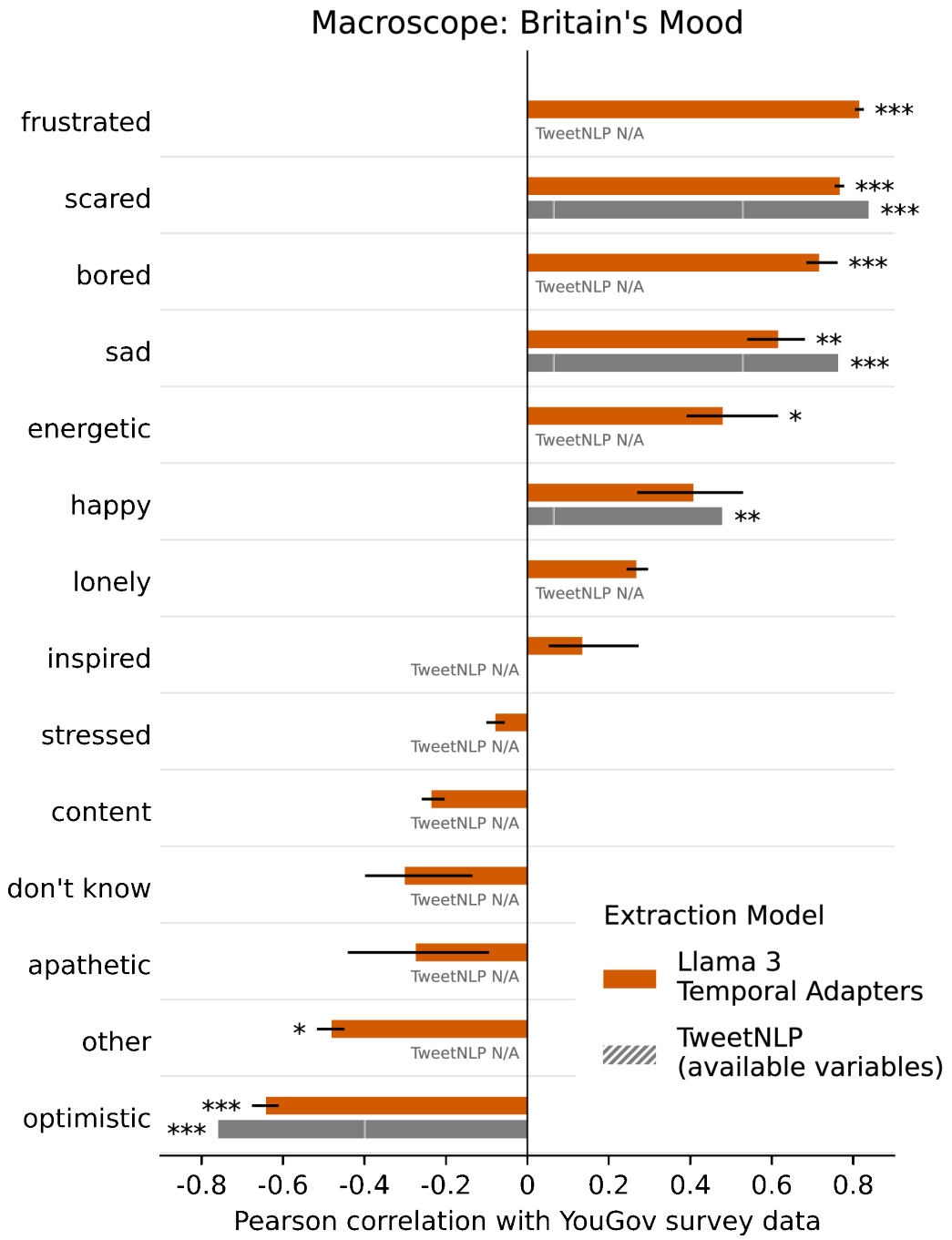
Broadly speaking, which of the following best describe your mood and/or how you have felt in the past week?

Weekly Affect Aggregate




Weekly Social Media Data
Week n Data
t
"... Work tomorrow, I'm not so keen"
fine-tuning input
causal language modeling task
Llama 3 8B
Original Llama 3 Weights (frozen)
Week n
Adapter
+


Britain's Mood: Scared
Britain's Mood: Happy
Llama 3 Temporal Adapters
YouGov Survey Data
TweetNLP Extraction
normalized probability
1.0
0.8
0.6
0.4
0.2
0.0
1 Jan 2020
23 Mar 2020
UK lockdown
announced
1 Jun 2020
restrictions
partially lifted


Macroscope: Britain's Mood
frustrated
scared
bored
sad
energetic
happy
lonely
inspired
stressed
content
don't know
apathetic
other
optimistic
***
TweetNLP N/A
***
***
***
TweetNLP N/A
**
***
*
TweetNLP N/A
**
TweetNLP N/A
TweetNLP N/A
TweetNLP N/A
TweetNLP N/A
TweetNLP N/A
*
TweetNLP N/A
***
***
Extraction Model
Llama 3
Temporal Adapters
TweetNLP
(available variables)
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
Pearson correlation with YouGov survey data

Results

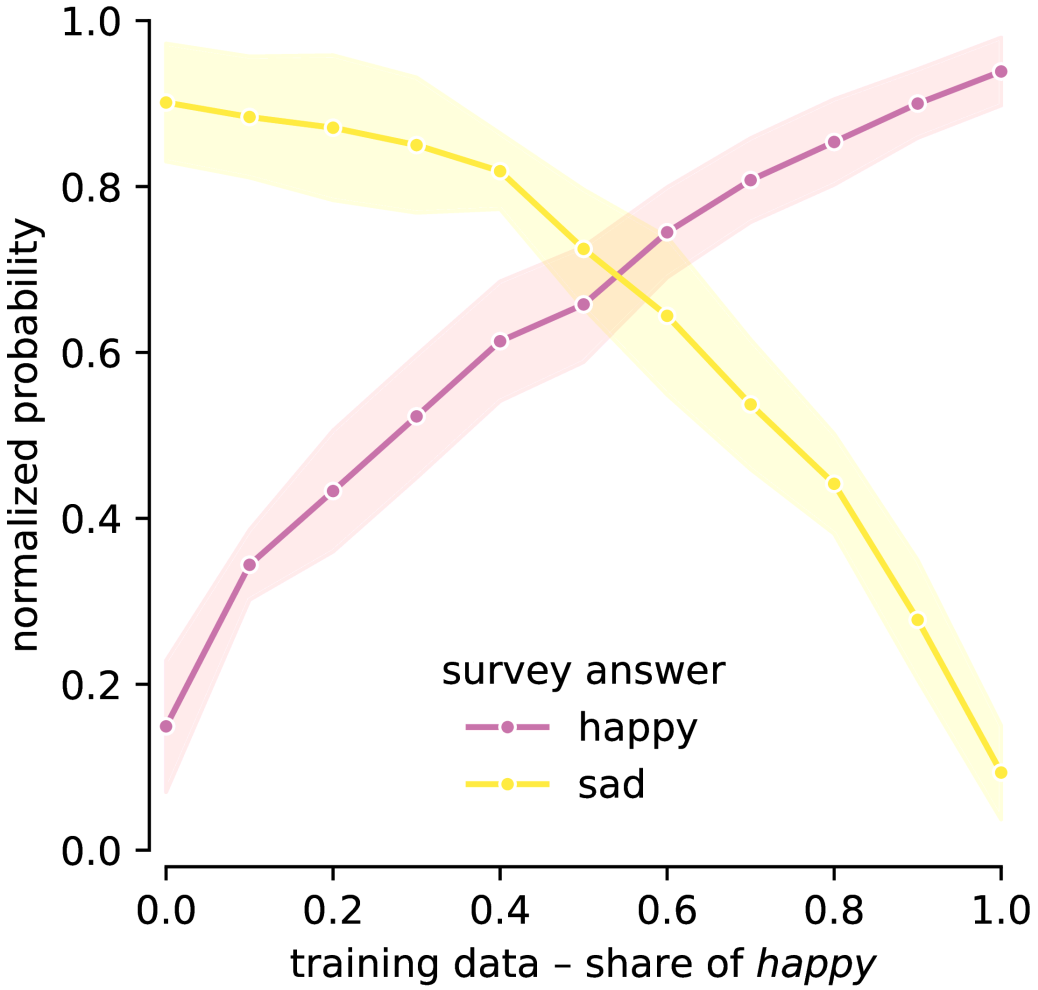
We can recreate dynamics with that approach of longitudinal adaptors

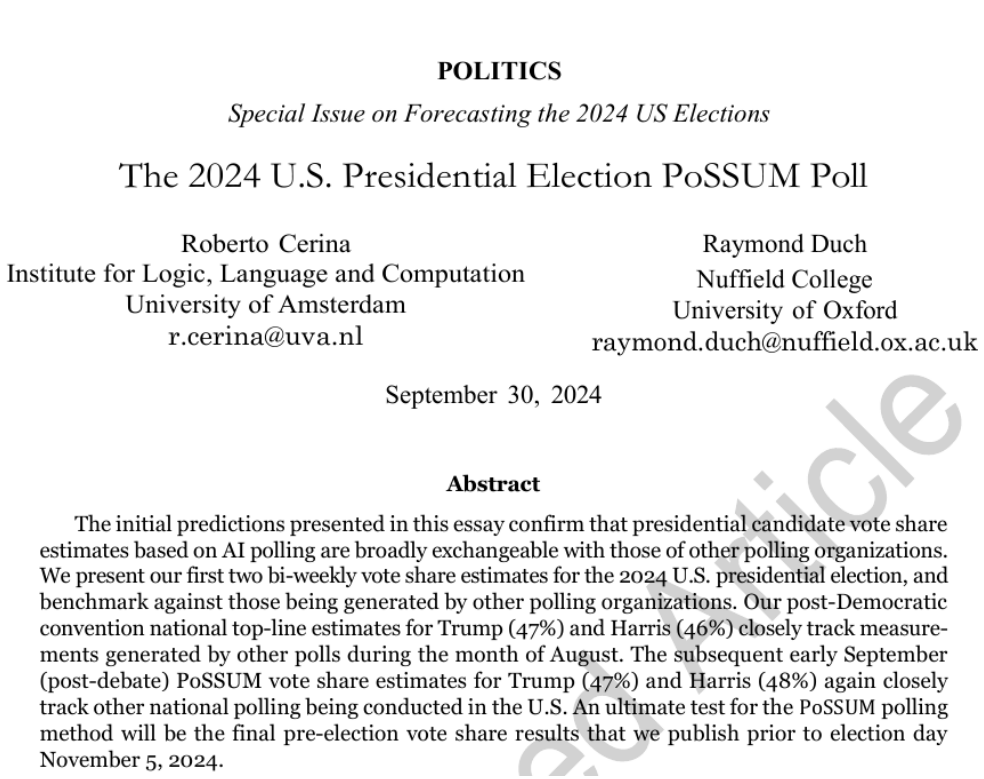
Not equally well for all constructs

Remember, our approach is just self-supervised next token prediction (no labels present as for example with the supervised text classification method of TweetNLP)

Our approach is very flexible, we can in principle ask any question and get survey-like responses for each week

Why does that work?


1.0
0.8
0.6
0.4
0.2
0.0
normalized probability
survey answer
happy
sad
0.0
0.2
0.4
0.6
0.8
1.0
training data – share of *happy*



POLITICS

Special Issue on Forecasting the 2024 US Elections

The 2024 U.S. Presidential Election PoSSUM Poll

Roberto Cerina
Institute for Logic, Language and Computation
University of Amsterdam
r.cerina@uva.nl

Raymond Duch
Nuffield College
University of Oxford
raymond.duch@nuffield.ox.ac.uk

September 30, 2024

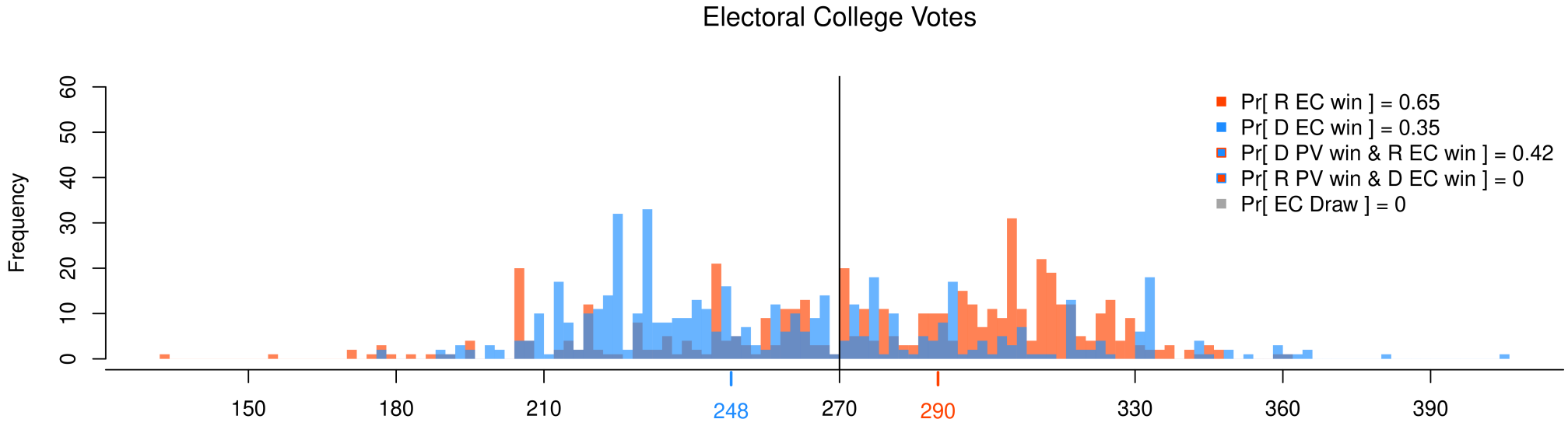
Abstract

The initial predictions presented in this essay confirm that presidential candidate vote share estimates based on AI polling are broadly exchangeable with those of other polling organizations. We present our first two bi-weekly vote share estimates for the 2024 U.S. presidential election, and benchmark against those being generated by other polling organizations. Our post-Democratic convention national top-line estimates for Trump (47%) and Harris (46%) closely track measurements generated by other polls during the month of August. The subsequent early September (post-debate) PoSSUM vote share estimates for Trump (47%) and Harris (48%) again closely track other national polling being conducted in the U.S. An ultimate test for the PoSSUM polling method will be the final pre-election vote share results that we publish prior to election day November 5, 2024.

Cerina, R., & Duch, R. (2023). Artificially Intelligent Opinion Polling (arXiv:2309.06029). arXiv. http://arxiv.org/abs/2309.06029

Cerina, R., & Duch, R. (2024). The 2024 U.S. Presidential Election PoSSUM Poll. PS: Political Science & Politics, 1–28. https://doi.org/10.1017/S1049096524000982

Cerina, R. (2025). PoSSUM: A Protocol for Surveying Social-media Users with Multimodal LLMs (arXiv:2503.05529). arXiv. https://doi.org/10.48550/arXiv.2503.05529


Electoral College Votes
Frequency
0
10
20
30
40
50
60
150
180
210
248
270
290
330
360
390
Pr[R EC win] = 0.65
Pr[D EC win] = 0.35
Pr[D PV win & R EC win] = 0.42
Pr[R PV win & D EC win] = 0
Pr[EC Draw] = 0

Wrap-up

I don't think we should be replacing survey research

Also with complementary synthetic methods we will need classical approaches for example to learn about the sampling frame

But we should be making use of the text that people are producing (and potentially other modalities too)

It's first steps for now and we strongly have to validate what we are doing

Huge potential: Low costs, scalability, unobtrusive observation, high temporal resolution, …

Bridging the gap between "qualitative" data and quantitative insights