

The Narratives of the War in Ukraine in News Media Worldwide

Michal Parizek & Jakub Stauber
Charles University



Global Flows of Political Information

GLOWIN

- target: map of the dyadic information ties among states
- **big data:** tracking content of online news media in close to 200 states and across 60+ languages
- **text as data:** dictionary, SML, generative AI
- empirical applications e.g. on news media visibility and coverage of IOs, the EU, on climate change/carbon neutrality, and on *the war and its narratives*

The concept

Two competing narratives of the war

1. **Western narrative:** Illegal, unprovoked aggression by Russia.
2. **Russian narrative:** War essentially a defensive response to Western expansion to Russia's sphere of influence.

RQs (informally)

- *Which of the two narratives prevails in which part of the world (especially beyond Europe and the West)?*
- *Can we develop a (language) model to capture these narratives?*

Empirical data

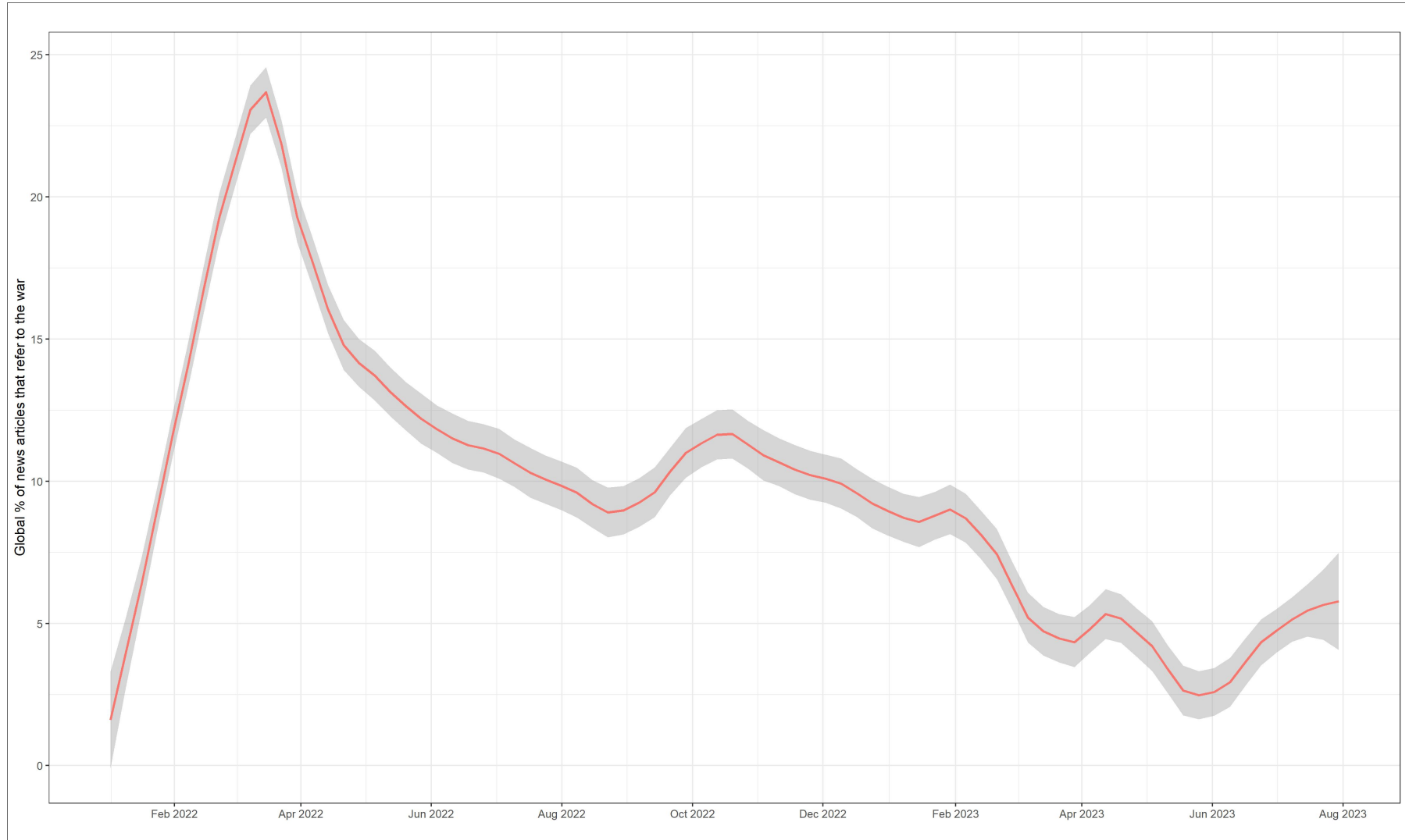
- 5 million news articles
- around 3000 media outlets from January 2022 to August 2023
- more than 4 million KWICs in close to 0.5 million unique articles in 1879 media outlets
- 136 states, 56 languages, non-English content (80%) translated to English

Dictionary techniques and the visibility of the war worldwide

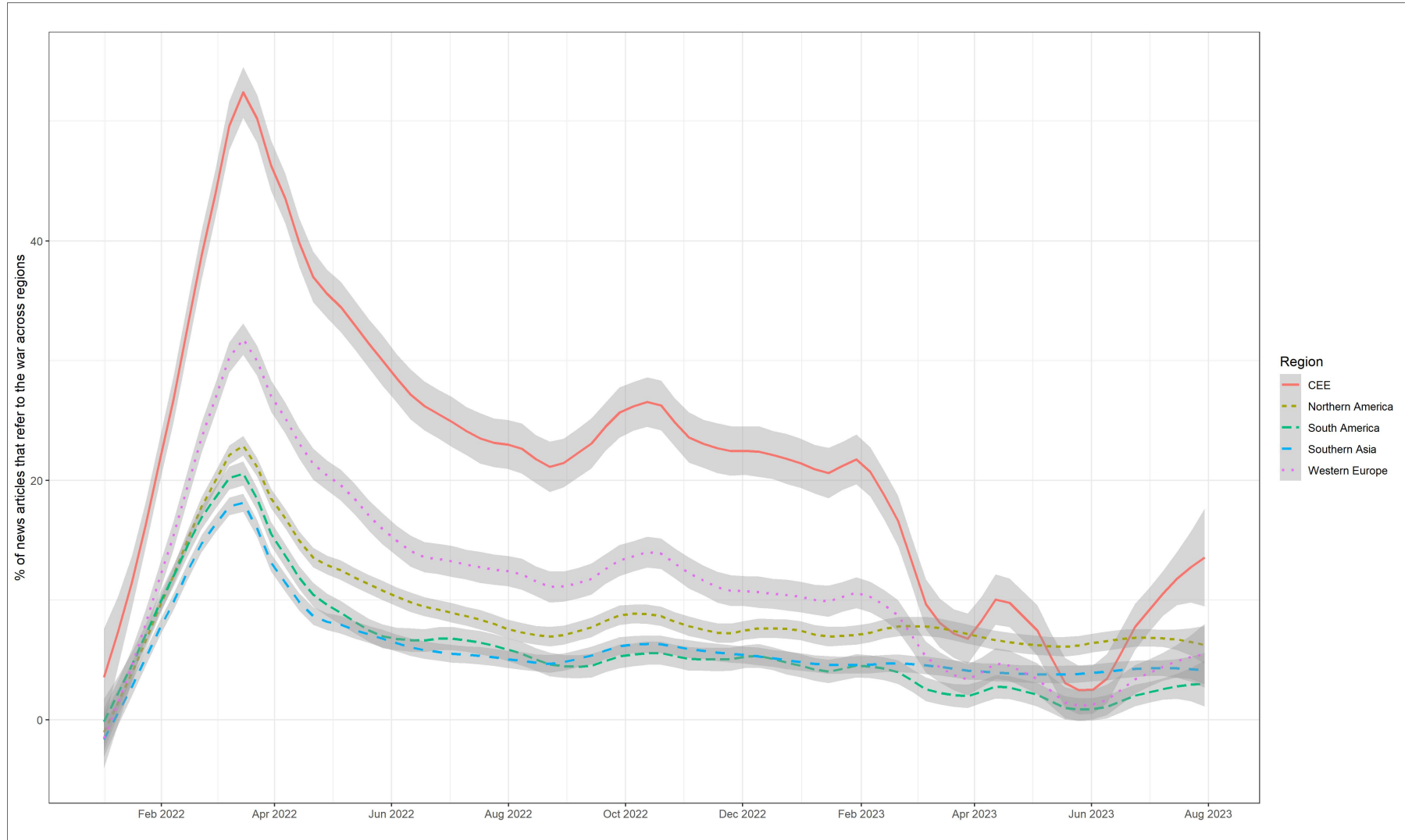
Dictionary terms to track the war

iso3	ent_searchterm_plain	regex
RUS	Russia	(\.: , ; \\A\\s)Russia[a-zA-Z]*
RUS	Putin	(\.: , ; \\A\\s)Putin(\.: , ; \\Z\\s)
RUS	Moscow	(\.: , ; \\A\\s)Moscow(\.: , ; \\Z\\s)
UKR	Ukrain	(\.: , ; \\A\\s)Ukrain[a-zA-Z]*
UKR	Zelensk	(\.: , ; \\A\\s)Zelensk[a-zA-Z]*
UKR	Kyiv	(\.: , ; \\A\\s)Kyiv(\.: , ; \\Z\\s)
UKR	Kiev	(\.: , ; \\A\\s)Kiev(\.: , ; \\Z\\s)

News media visibility of the war over time



News media visibility of the war across selected regions



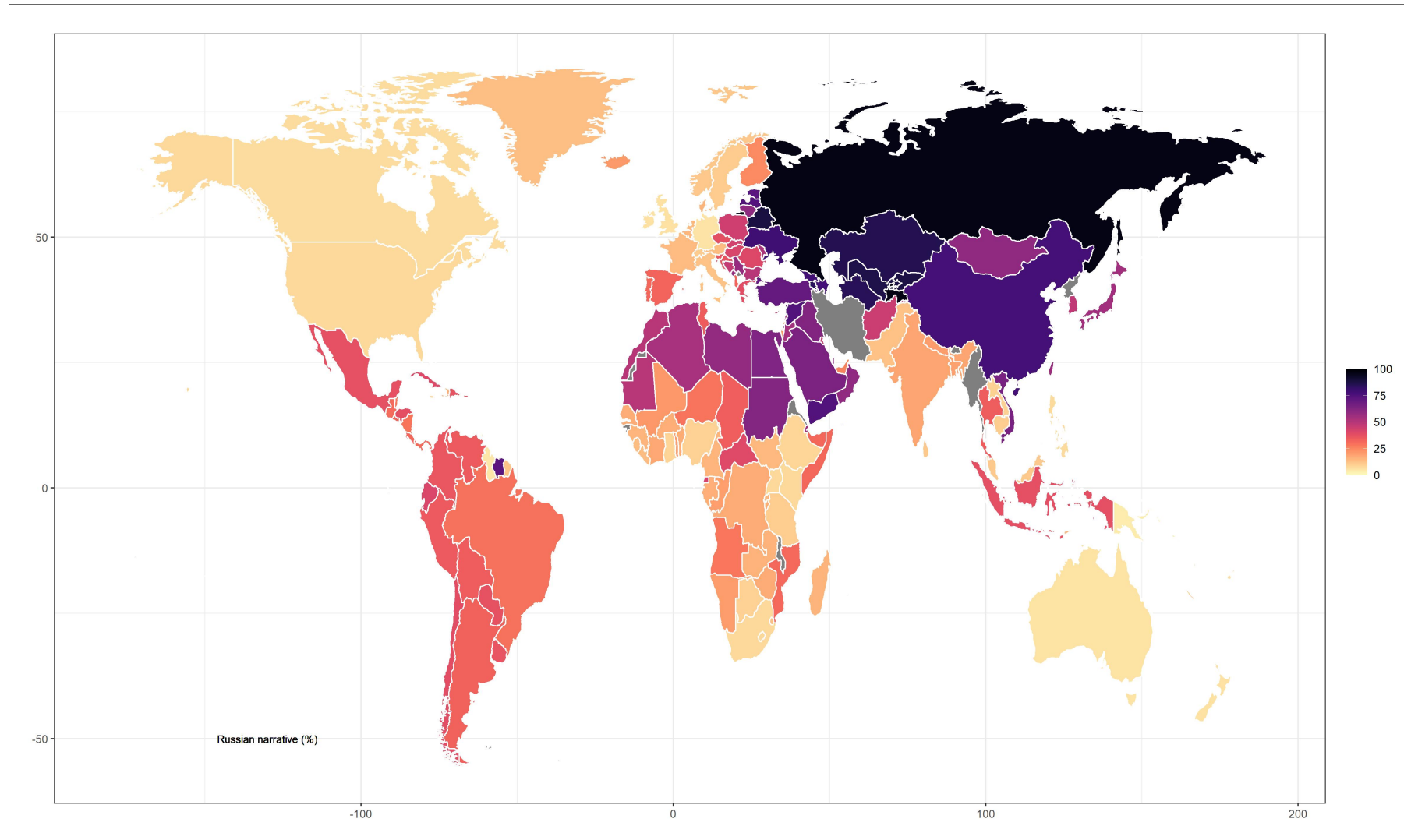
Supervised ML model to track the narratives

Training a model

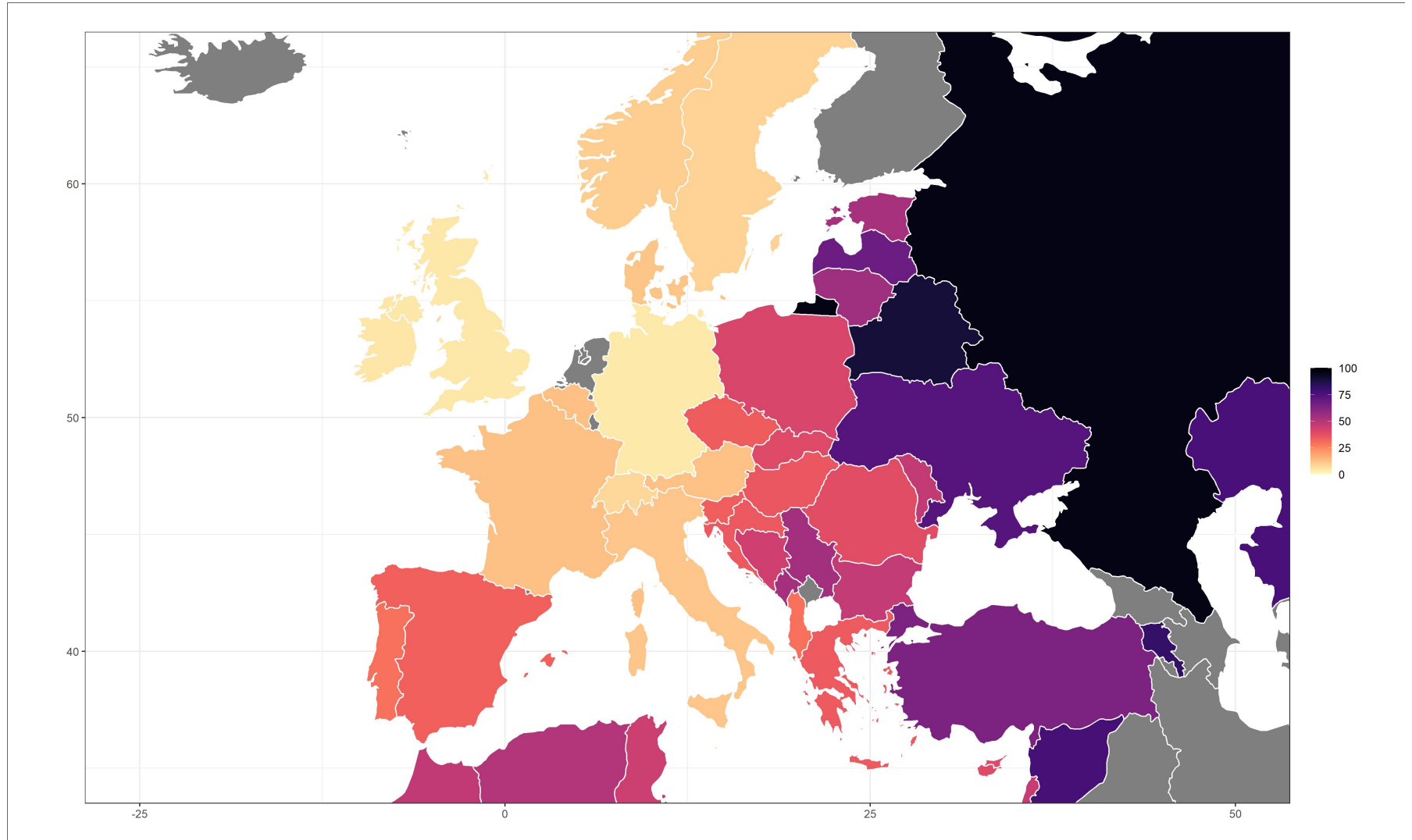
- representative sample of 5000 KWICS from *Western European media* (German, Italian, French)
- representative sample of 5000 KWICS from *Russian international media* (sputniknews, ria, rt, 5-tv)
- deBERTa-v3 large (F-1 score at 0.92)

	Predicted R	Predicted W
True R	911	71
True W	77	933

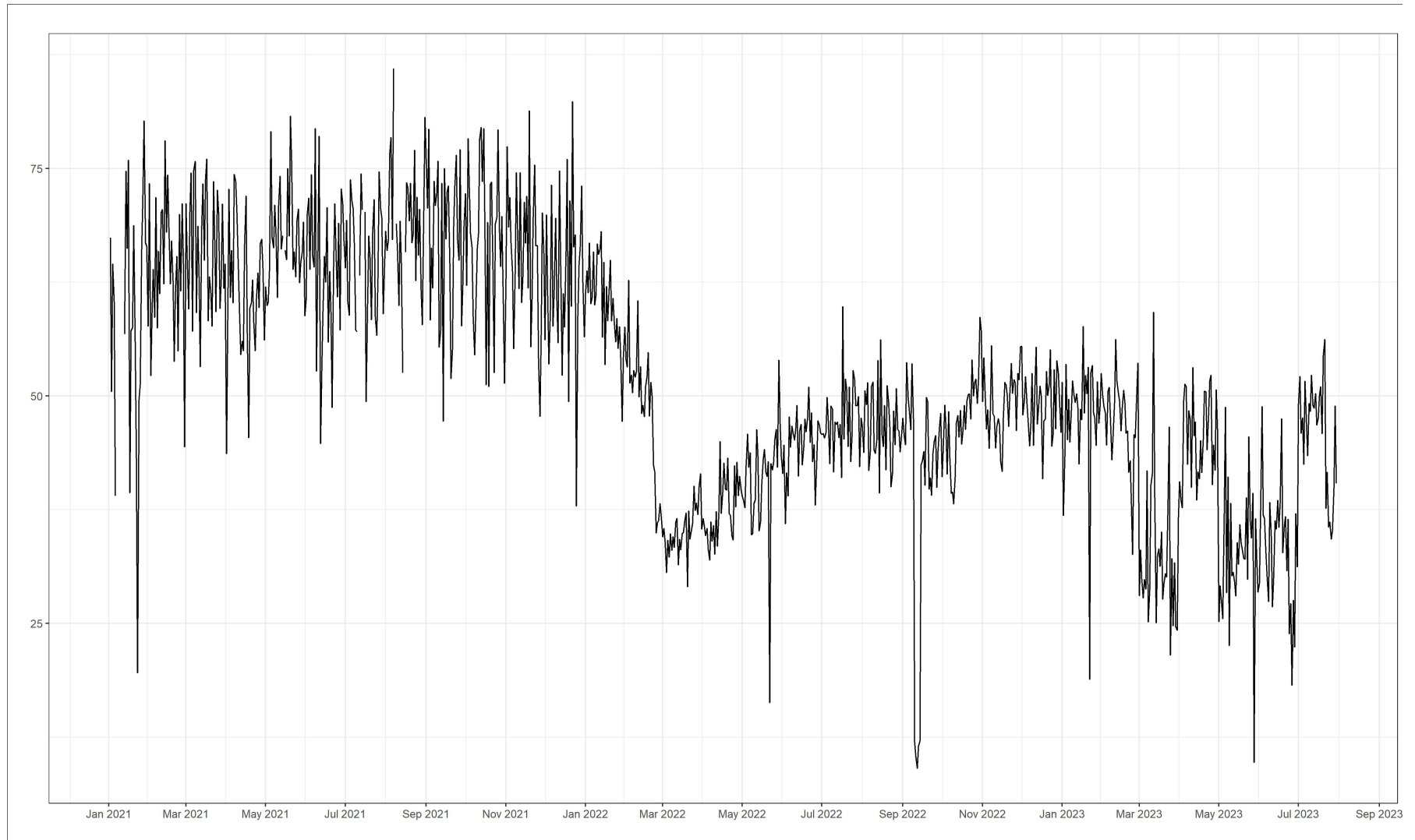
Prominence of the Russian narrative across the world



Prominence of the Russian narrative in Europe?



Change over time?



Generative AI

Generative AI

- Data
 - Hand-picked human-coded cases of the narratives about the war
 - Keywords in context mentioning the war
- Research design and validation
 - GPT vs supervised ML trained on source country
 - GPT vs human-coded baseline

Our prompt

In 2022, Russia attacked Ukraine. The **Russian view** is that in the conflict, Russia conducts a special military operation to defend itself against Western expansion, NATO, and the United States hegemony. The **Western view** is that the conflict represents an illegal unprovoked invasion of Ukraine, a violation of its sovereignty, and that in its aggressive war Russia is committing serious crimes. There is a third view which is neutral. This **neutral view** highlights the need for peace and stresses the costs of the war for other nations such as those in Africa, Asia and Latin America. In the following, I will provide you with a short text. Please write in less than five words which of the three views the text is closer to – Russian (write view ‘R’), Western (write view ‘W’), or neutral (write view ‘N’). If you do not recognize any of these three views, please write ‘not applicable’.

LLM performance on hand-picked dataset

Model	F1 micro	F1 macro
GPT-4	0.93	0.92
GPT-3.5	0.83	0.82
Platypus2 70B Instruct	0.89	0.89

GPT-4

- 10 000 texts about war classified with GPT-4 vs. our supervised ML model based on source-country dataset
- NAs (no narrative) removed (around 50%) -> remaining 5080 rows with Russian or Western narrative used
- F-1 score: >0.85
- It supports our sampling approach used in the supervised ML model above
- Preliminary human validation of 80 cases on all 3 categories (W, R, NA) - identification of NA category is surprisingly correct

Next step?

- How to include the NA category?
- Train supervised ML model using classification by GPT-4 to identify (W, R, and NA)
- Preliminary results:
 - 10K labelled texts, from these balanced sample of 6500
 - Trained supervised ML on balanced data from GPT itself - F-1 >0.80