# The Narratives of the War in Ukraine in Czech News Media

Jakub Stauber
Charles University, Prague, Czechia

2nd Budapest Methods Workshop on Large Language Models and GenerativeAI

November 21, 2024

# GLOWIN

Global Flows of Political Information

# GLOWIN

- mapping information ties among states by tracking the content of news media
- **big data**: online news media in close to 200 states, across 60+ languages
- **text-as-data/natural language processing (NLP)**: dictionary, supervised ML, generative AI
- applications e.g. on states' media visibility (Parizek and Stauber 2024), media coverage of IOs (Parizek 2024), EU in the context of COVID-19 and the War in Ukraine (Rauh and Parizek 2024), and on ***the war in Ukraine and its narratives***

# V4 Media Slant

- using inovative NLP methods to map media bias in V4 countries
- new database of news media articles from Czechia, Hungary, Poland and Slovakia
- Media Dashboard

# Research goals

- Validate the SML models developed within the GLOWIN project and apply them to the new data.
- Present reflection of the war in Ukraine by the Czech media as a case study for a narrative detection task to uncover possible bias across diffent outlets and over time.

# Theoretical background

# Media bias

- Various media outlets may favor specific actors (e.g., political parties, politicians, economic organizations,…).
- The systematic long-term favoring of particular actors in reporting can be referred to as media bias (Rodrigo-Ginés, Carrillo-de-Albornoz, and Plaza 2024).
- This can be measured using NLP techniques to detect tonality, emotions or reproduction of specific narratives.

# Narratives of the war

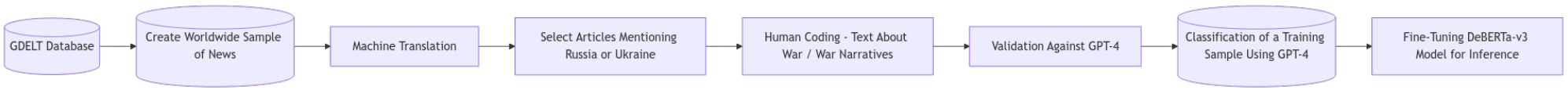
- *strategic* narratives of the war key for the war itself and the broader context (Freedman 2006; Miskimmon, O'Loughlin, and Roselle 2013; Schmitt 2018)
- *news media* a key carrier of these narratives

Two main competitors:

1. **Western narrative**: The war presented as an illegal, unprovoked aggression by Russia.
2. **Russian narrative**: The war presented as a defensive response to Western expansion to Russia's sphere of influence.

# Research Design

# Tracking the narratives of the war with generative AI: Workflow




1. Have GPT code a small sample of data (KWICS on the war)
2. Validate against human-labeled data
3. Have GPT code 10 000 KWICS, create a balanced *training* dataset
4. Fine-tune a DeBERTa model with this data
5. Deploy this fine-tuned model on the dataset

# Our *prompt*

In 2022, Russia attacked Ukraine. The **Russian view** is that Russia conducts this special military operation because it needs to defend itself against Western expansion and threats to its security from NATO and the United States and that Ukraine only serves the interests and hegemonic ambitions of the West. The **Western view** is that the conflict is an illegal and unprovoked invasion, that the aggressive war Russia wages violates Ukrainian sovereignty and undermines European security and that Russia attacks innocent civilians and commits crimes. In the following, I will provide you with a short text. Please write in less than five words which of the two views the text is closer to – Russian (write view 'R') or Western (write view 'W'). Many texts will not contain any of these views. If you do not recognize any of these two views, please write 'not applicable'.

# ChatGPT (GPT-4) performance on a small hand-labelled dataset

| True vs. Predicted | NA | R | W |
|---|---|---|---|
| NA | 42 | 3 | 2 |
| R | 2 | 19 | 1 |
| W | 2 | 1 | 8 |

- solid performance (F-1>0.8), the use of chatGPT seems validated for the task
- close to half of text segments *no* narrative

# Using GPT-4 instead of human coders

- balanced sample of chatGPT-labeled data (6,426 KWICS) to fine-tune a deberta-v3-large model (F-1 = 0.83)
- solid performance on the small hand-labelled dataset

| Human vs. DeBERTa | NA | R | W |
|---|---|---|---|
| NA | 43 | 0 | 4 |
| R | 4 | 16 | 2 |
| W | 1 | 3 | 7 |

# The Czech Case

# Empirical data

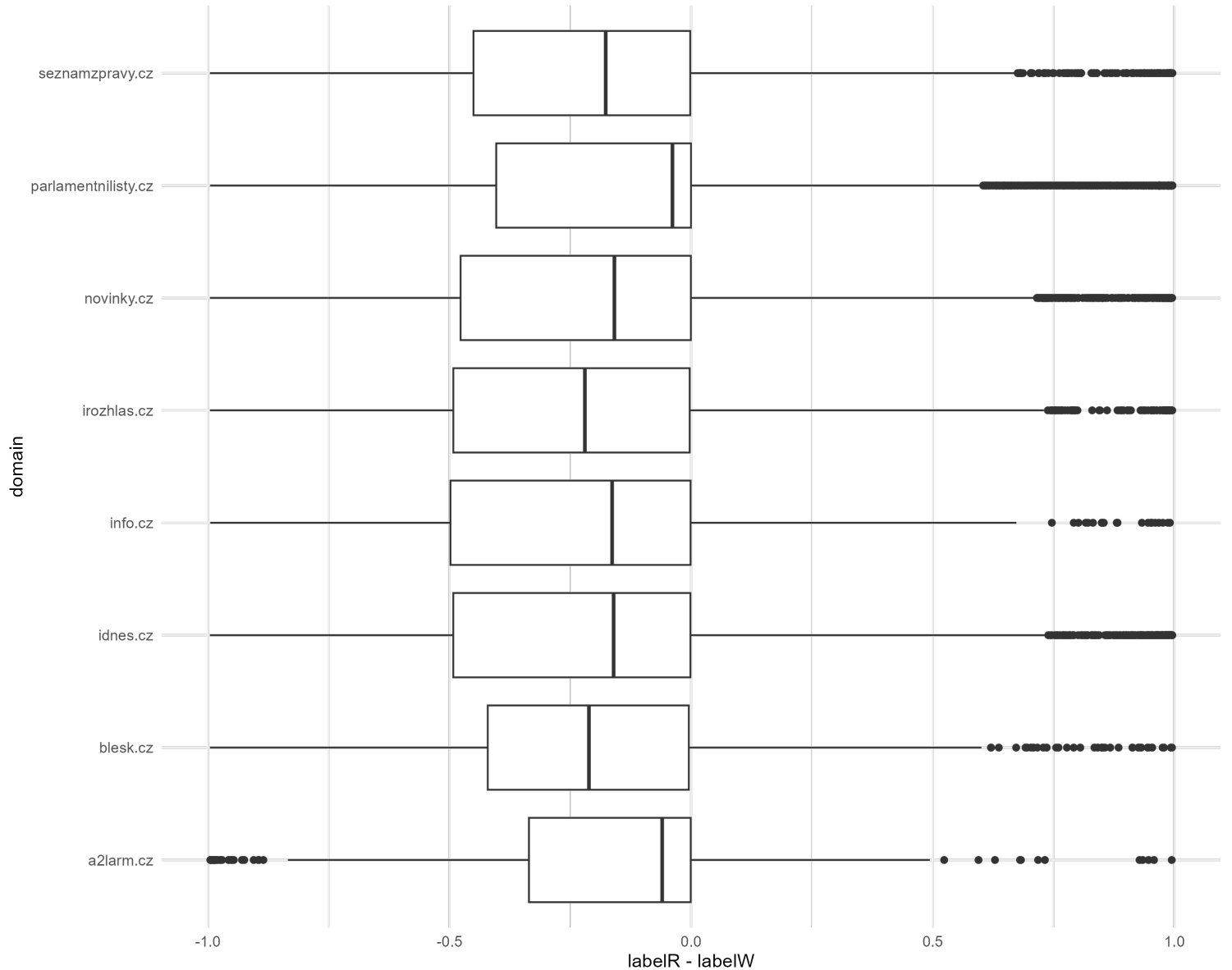
- News media articles from Czechia
- Sample covers wide spectrum of media outlets with different political positions and ownership structure
- Czech media landscape
  - A2larm: critical, left-wing
  - Blesk: tabloid
  - iDnes: mainstream, center-right, previously owned by Andrej Babis
  - info.cz: conservative, right-wing
  - iRozhlas: public radio broadcaster
  - Novinky: mainstream, center-left
  - Parlamentni Listy: desinformation
  - Seznam Zpravy: mainstream, center-right

# Pre-processing

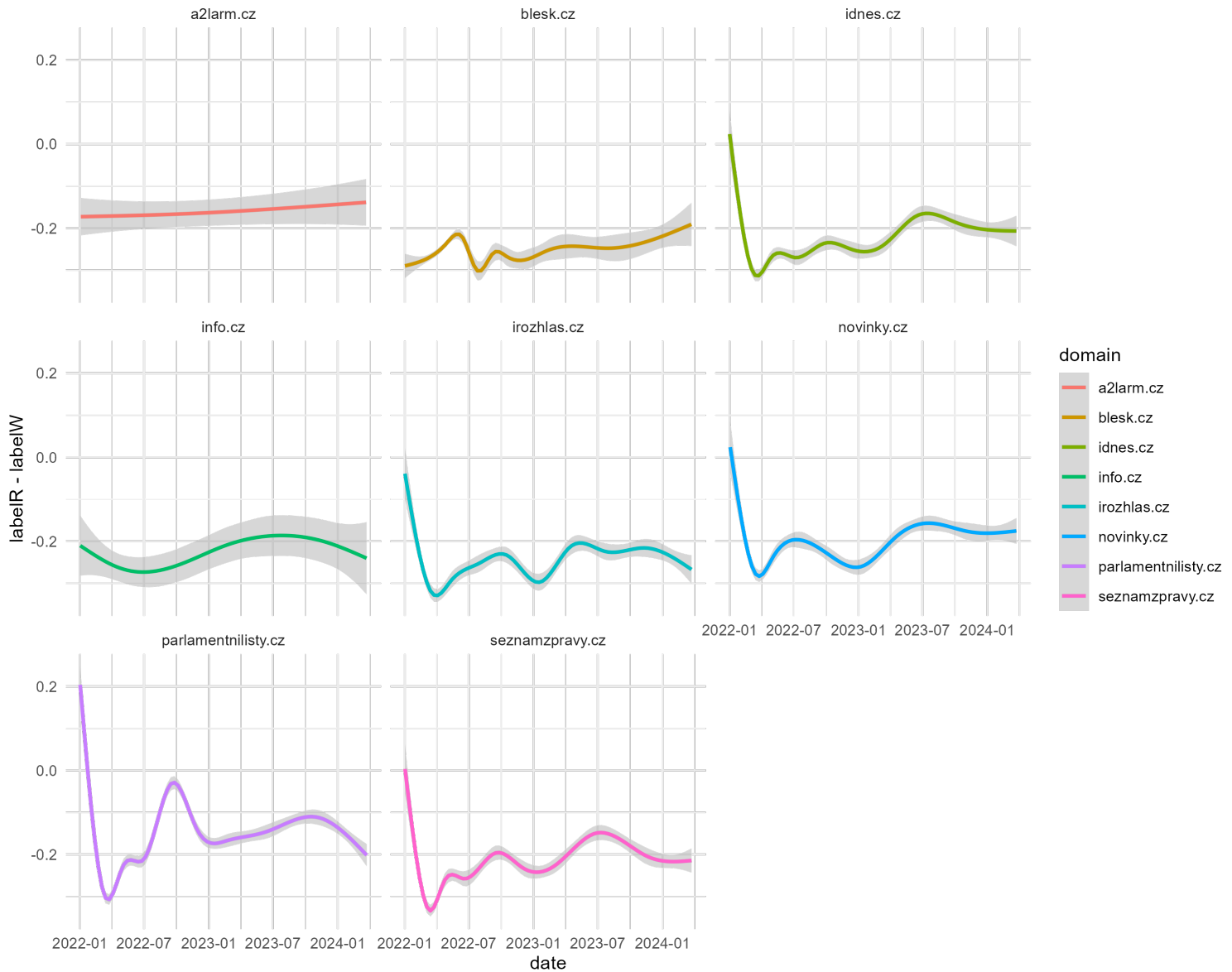
- The sample was splited into sentences
- Using regular expressions sentences refering to Russia AND/ OR Ukraine were selected
- The resulting subsample machine-translated to English using open source model OPUS
- The sentences associated with the war selected through a fine-tuned DeBERTa model (trained with 1000+ human-labelled KWICS, F-1>0.9)

# Empirical analysis

# Descriptive statistics



# Evolution of media bias over time



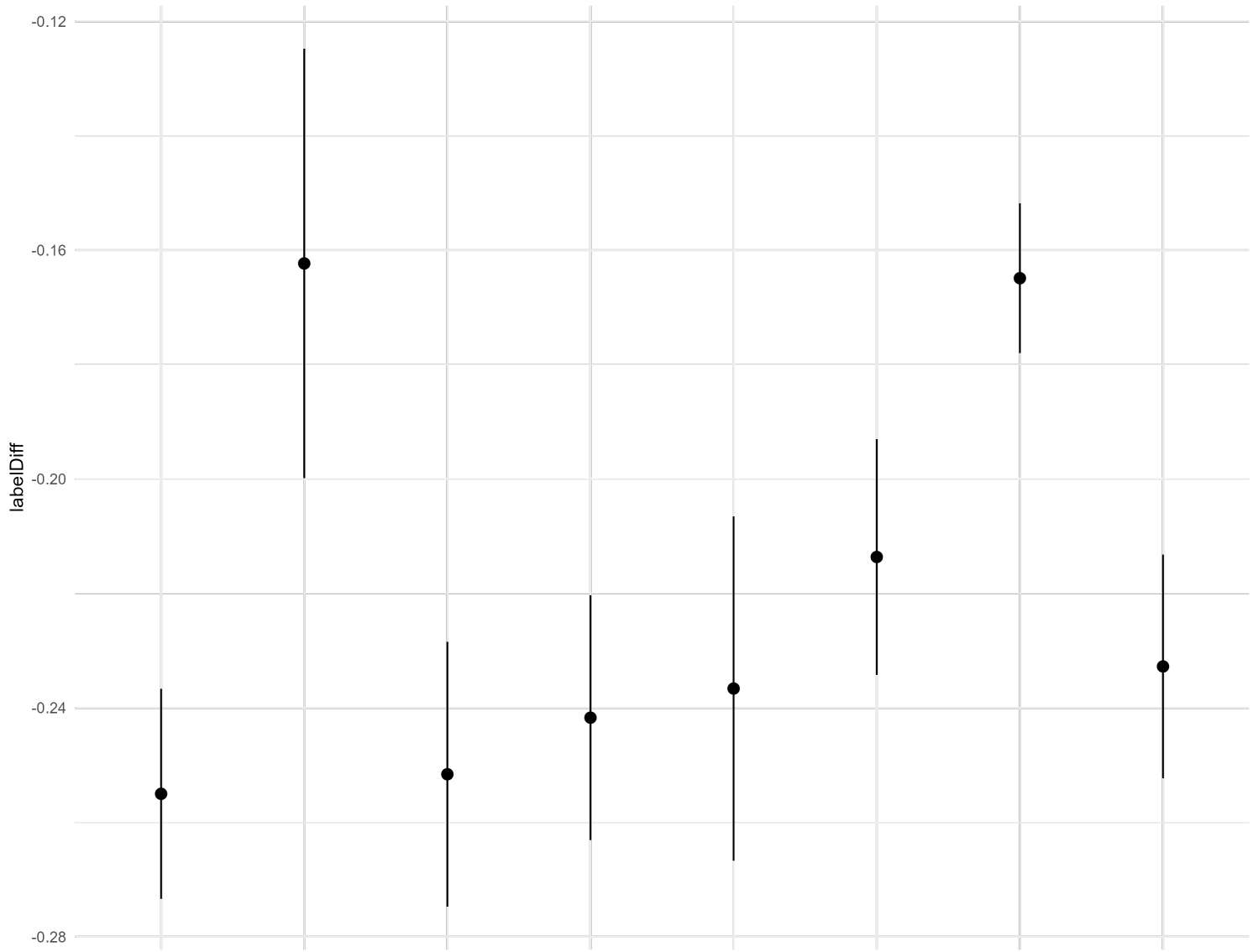
# Estimation of the bias?

- Aggregate on the article level (mean of probabilities over sentences refering to Russia)
- OLS model with weekly fixed-effects
- Models predicting mean probability of presence of Russian and Western narrative and difference between them

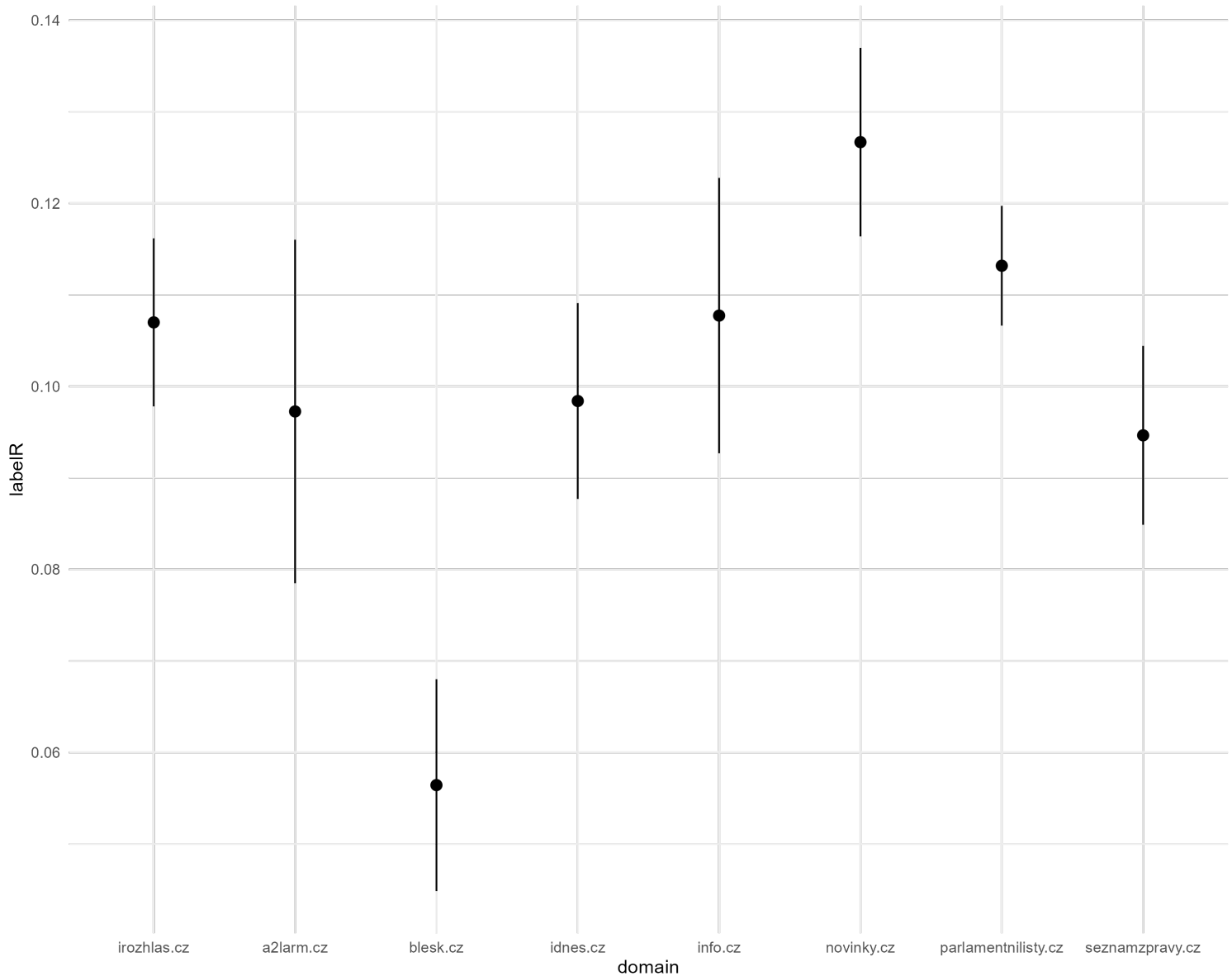
```
feols(Predicted Probability ~ Domain + Neutral | Date)
```

# Full period

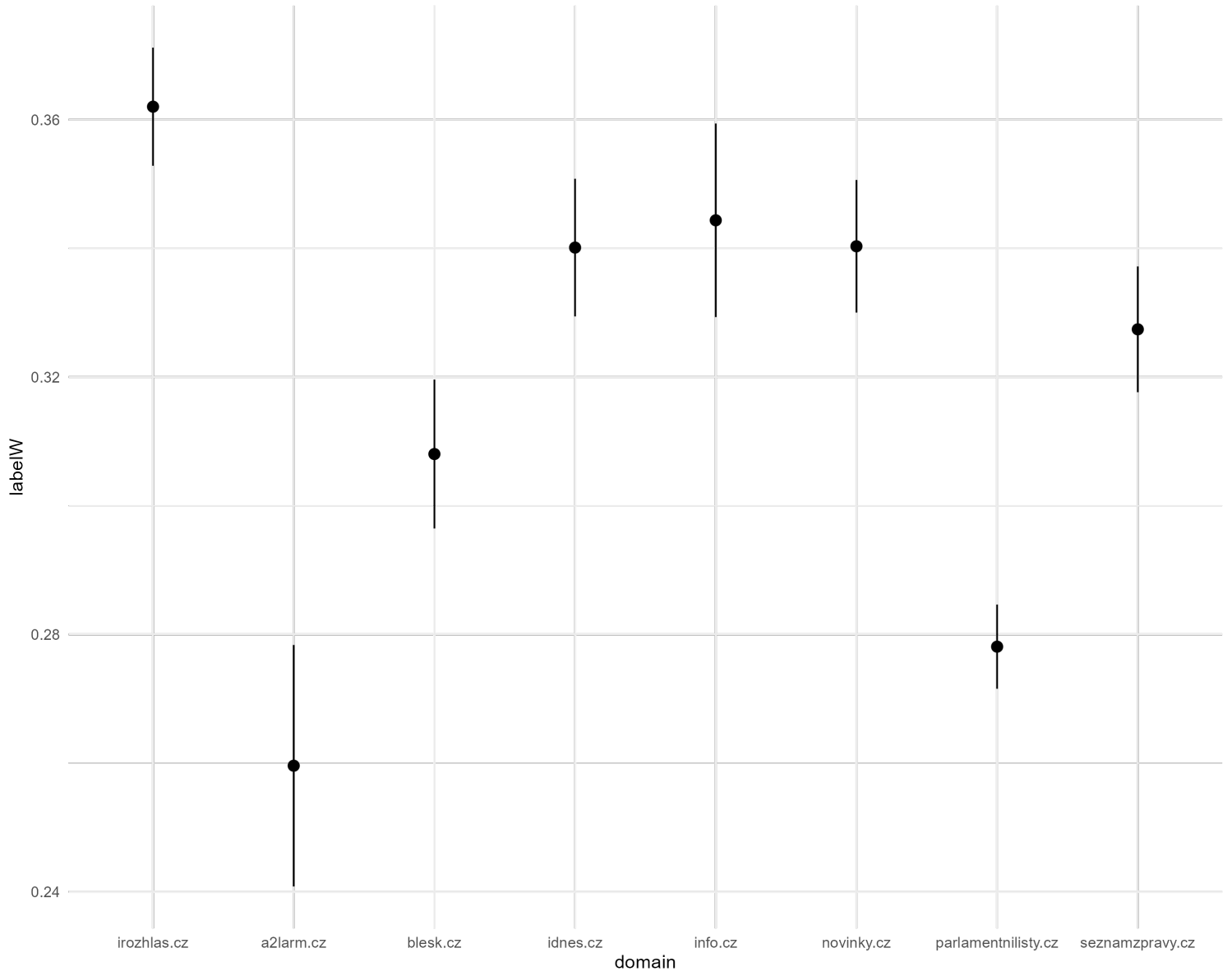
# Predictions from the model (Difference)



# Predictions from the model (Russia)

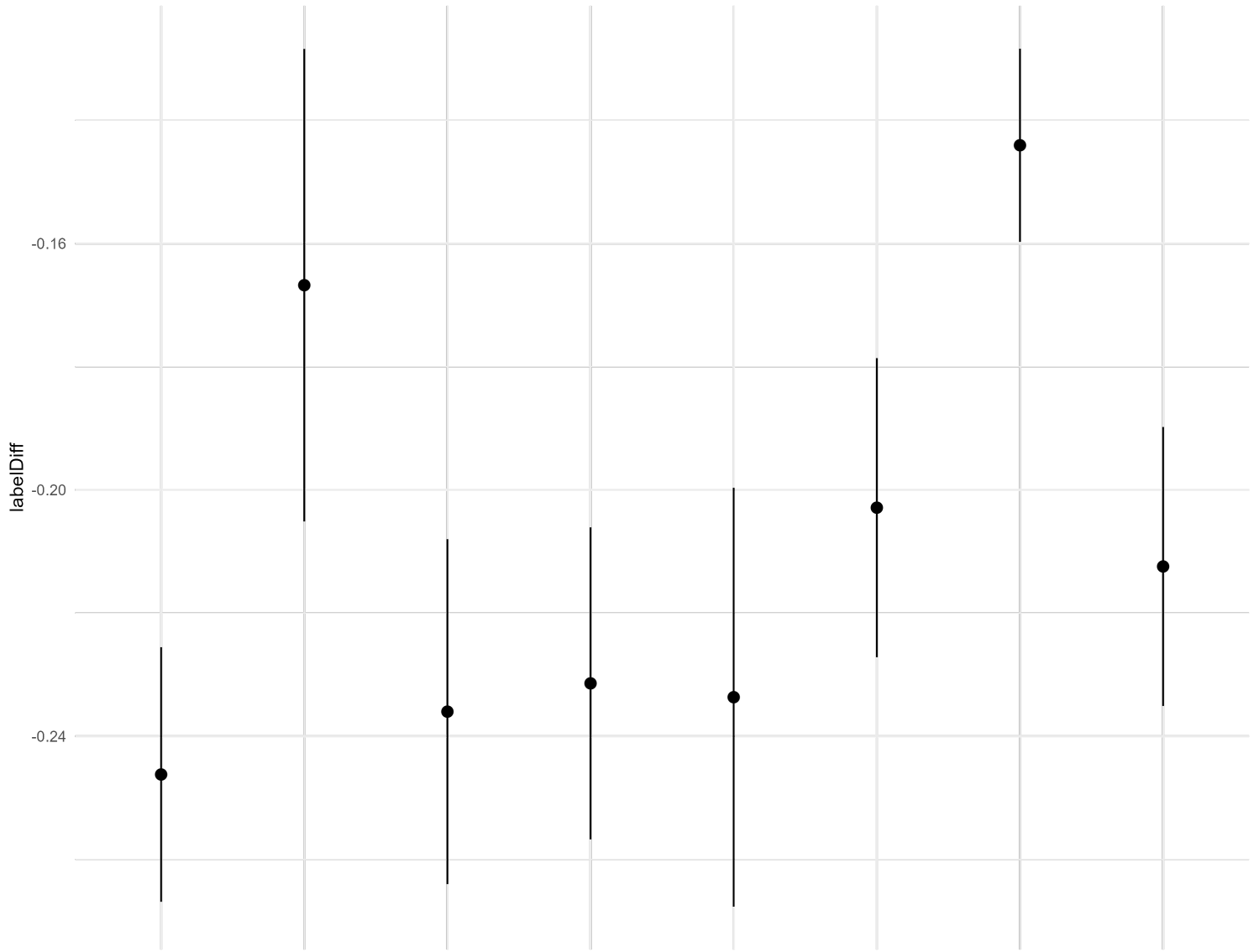


# Predictions from the model (West)

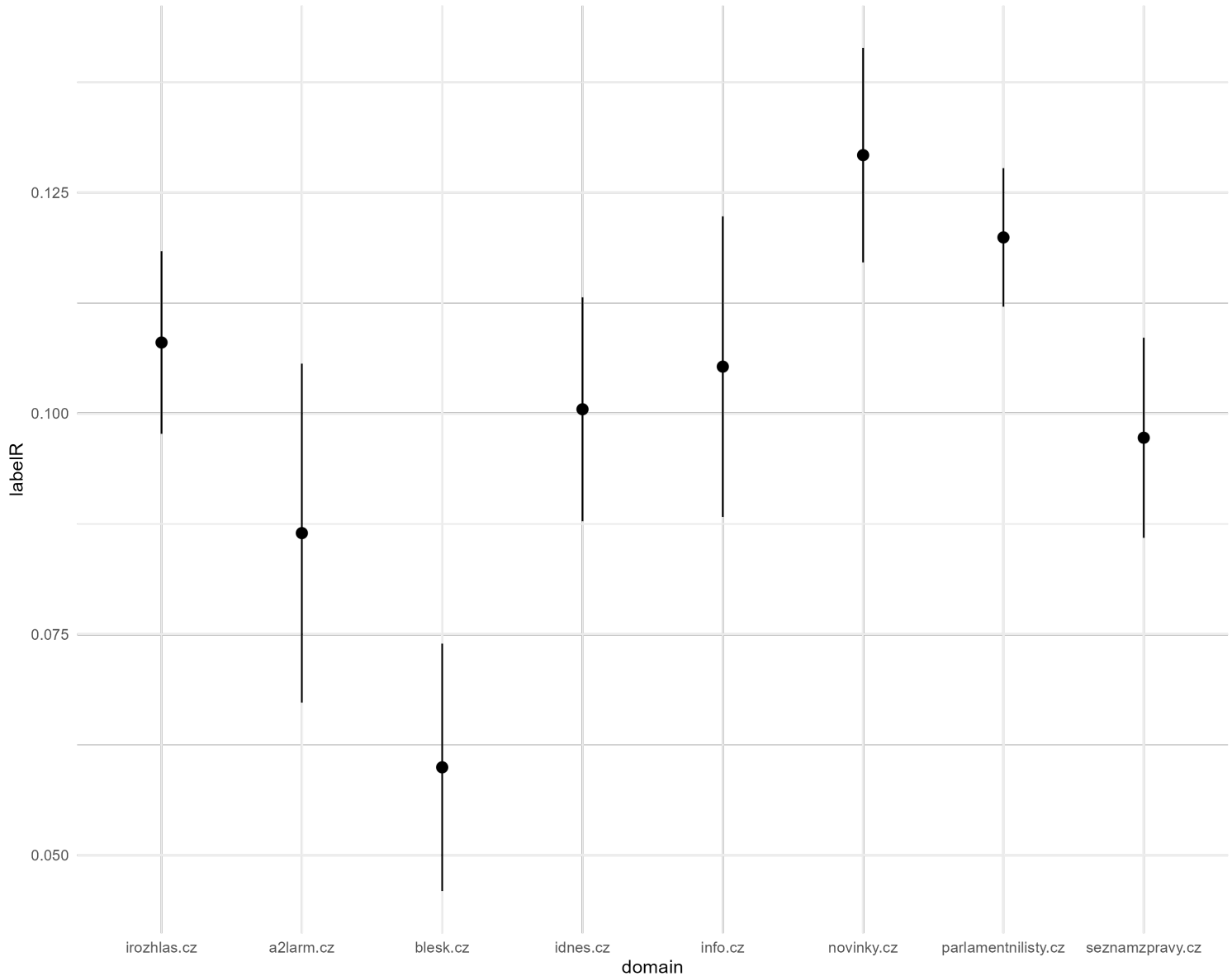


# Reduced sample (after 04-2024)

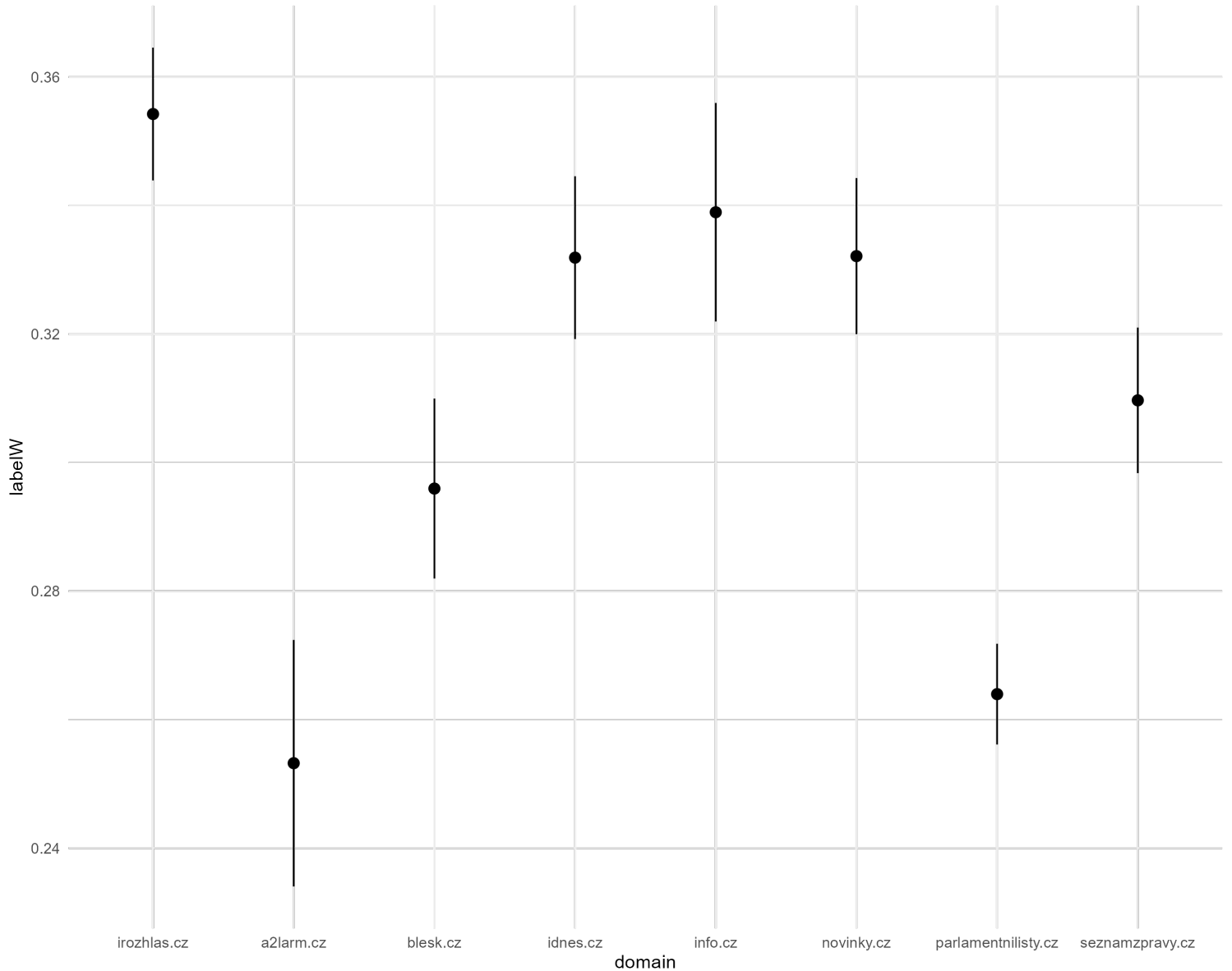
# Predictions from the model (Difference)



# Predictions from the model (Russia)



# Predictions from the model (West)



# Conclusions and next steps

- unique data on news media reporting of the war: significant differences across outlets
- The empirical strategy was able to capture a Russian-leaning bias in the case of disinformation actor
- NLP tools for detecting the narratives of the war
  - validated use of GPT-4 for coding
  - GPT-labelled data used to fine tune DeBERTa
- need to differentiate between reporting / citing and endorsing the narrative
- The results should be further validated using topic models to understand the actual content and evolution of the narratives

# References


Freedman, Lawrence. 2006. “Networks, Culture and Narratives.” *The Adelphi Papers* 45 (379): 11–26. https://doi.org/10.1080/05679320600661640.

Miskimmon, Alister, Ben O’Loughlin, and Laura Roselle. 2013. *Strategic Narratives: Communication Power and the New World Order*. New York: Routledge. https://doi.org/10.4324/9781315871264.

Parizek, Michal. 2024. “Less in the West: The Tangibility of International Organizations and Their Media Visibility Around the World.” *The Review of International Organizations*. https://doi.org/10.1007/s11558-024-09551-6.

Parizek, Michal, and Jakub Stauber. 2024. “The Global Media Visibility of States Dataset V1.” *OSF*. doi:10.17605/OSF.IO/2X9R3.

Rauh, Christian, and Michal Parizek. 2024. “Converging on Europe? The European Union in Mediatised Debates During the COVID-19 and Ukraine Shocks.” *Journal of European Public Policy*, April. https://www.tandfonline.com/doi/abs/10.1080/13501763.2024.2344849.

Rodrigo-Ginés, Francisco-Javier, Jorge Carrillo-de-Albornoz, and Laura Plaza. 2024. “A Systematic Review on Media Bias Detection: What Is Media Bias, How It Is Expressed, and How to Detect It.” *Expert Systems with Applications* 237 (March):

121641. https://doi.org/10.1016/j.eswa.2023.121641.
Schmitt, Olivier. 2018. “When Are Strategic Narratives Effective? The Shaping of Political Discourse Through the Interaction Between Political Myths and Strategic Narratives.” *Contemporary Security Policy* 39 (4): 487–511. https://doi.org/10.1080/13523260.2018.1448925.