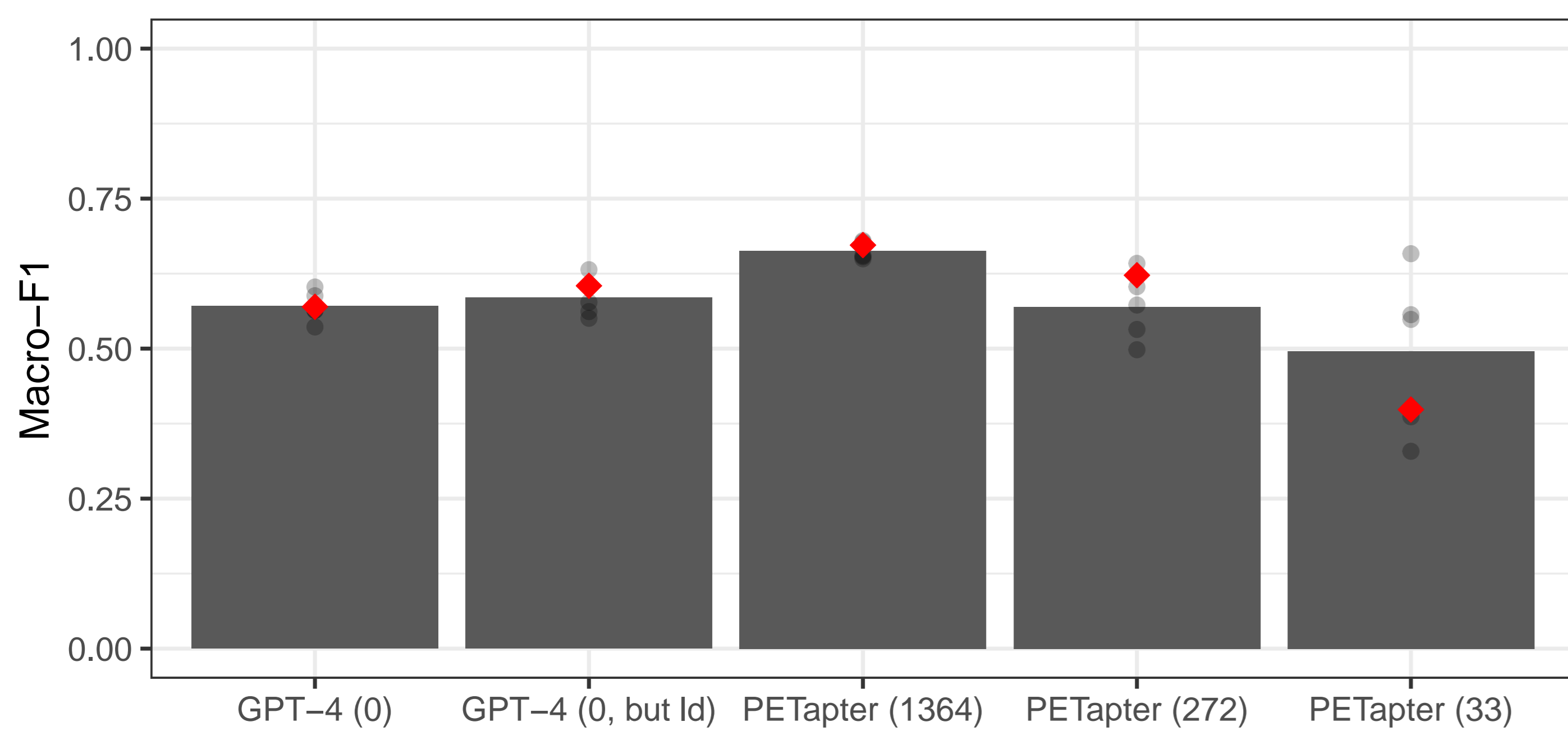


Few-shot learning for automated content analysis (FLACA) in the German media debate on arms deliveries to Ukraine

Jonas Rieger, Mattes Ruckdeschel, Kostiantyn Yanchenko, Gerret von Nordheim, Katharina Kleinen-von Königslöw, Gregor Wiedemann

1. Ildr; PETapter outperforms GPT-4

- GPT-4 has a basic zero-shot understanding / representation of claims and arguments
 - Performance is not satisfying (0.57 macro-F1)
 - Label distribution (ld) helps (0.57 vs. 0.55 macro-F1)
 - Stance prediction is easy (all runs > 0.77 macro-F1)
- Our tailor-made method (PETapter) outperforms GPT-4
 - Need of some training examples (at least 272 annotations)
 - Stance prediction is easy (0.9 macro-F1)



Glossary:

few-shot learning	using a small number of annotated examples
zero-shot learning	using no annotated data
supervised learning	using annotated data to learn relations
near-domain pre-training	using existing (and roughly related) annotated data
SOTA	state-of-the-art
NLP	natural language processing
(macro-)F1	evaluation measure to maximize

Paper:

Few-shot learning for automated content analysis: Efficient coding of arguments and claims in the debate on arms deliveries to Ukraine. Accepted for *Studies in Communication and Media*.

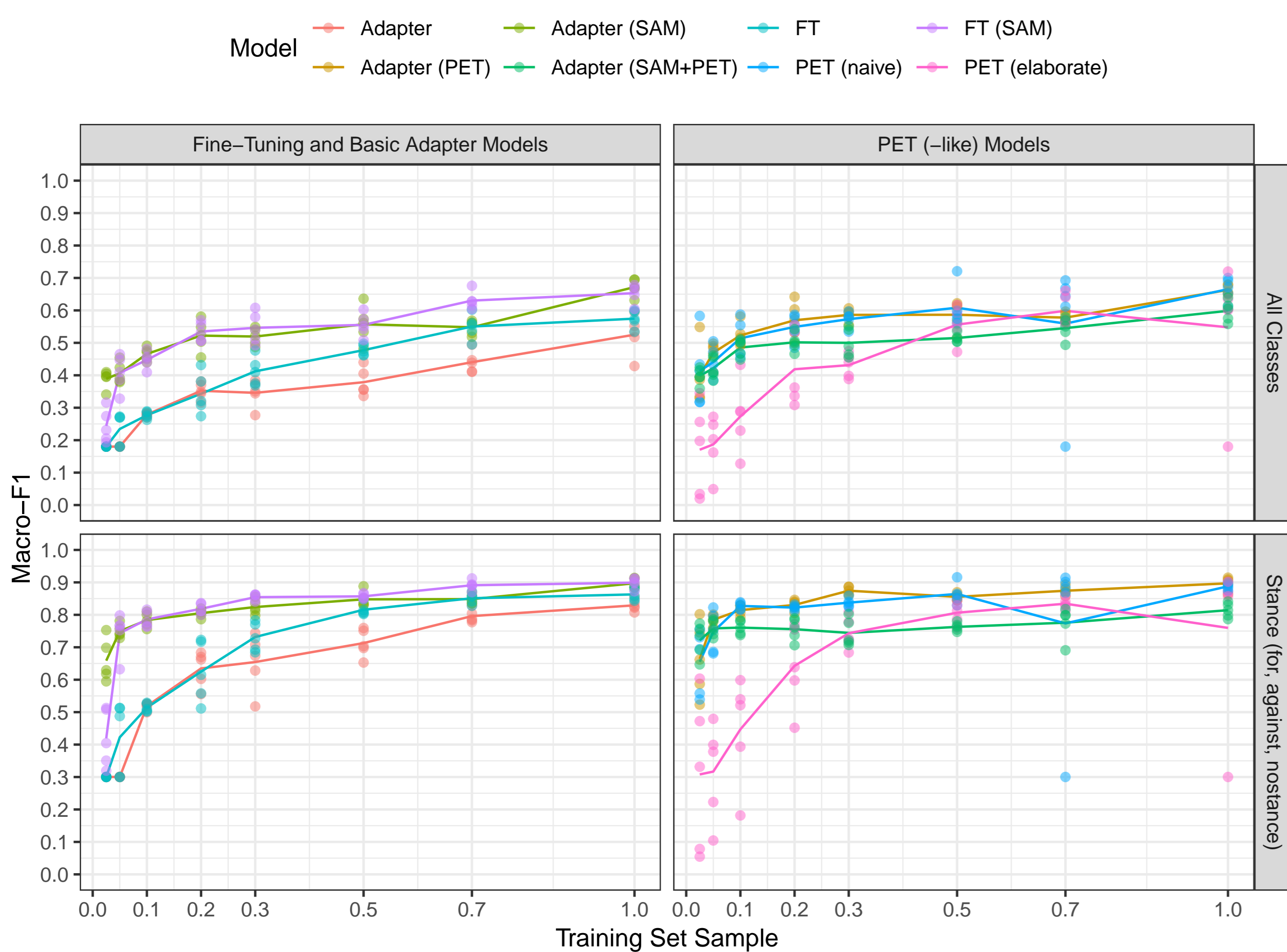
2. Utilizing supervised learning in viewpoint diversity [1, 2, 5]

- Coding** claims for/against and arguments for/against (+ no stance) at the sentence level
- Accounting for two **contextual sentences** before and after a target sentence
- Challenging** task even for humans (four coders: $\alpha_1 = 0.497$; $\alpha_2 = 0.667$; $\alpha_3 = 0.662$)
- Bonus:** Coding "onion-structured" claims and arguments (*not displayed on poster*)
- Case: **Waffenlieferung debate** in Germany [6, 8]
 - 01/01/2022 – 11/30/2022; 22 national newspapers; 7,301 articles

Search terms:

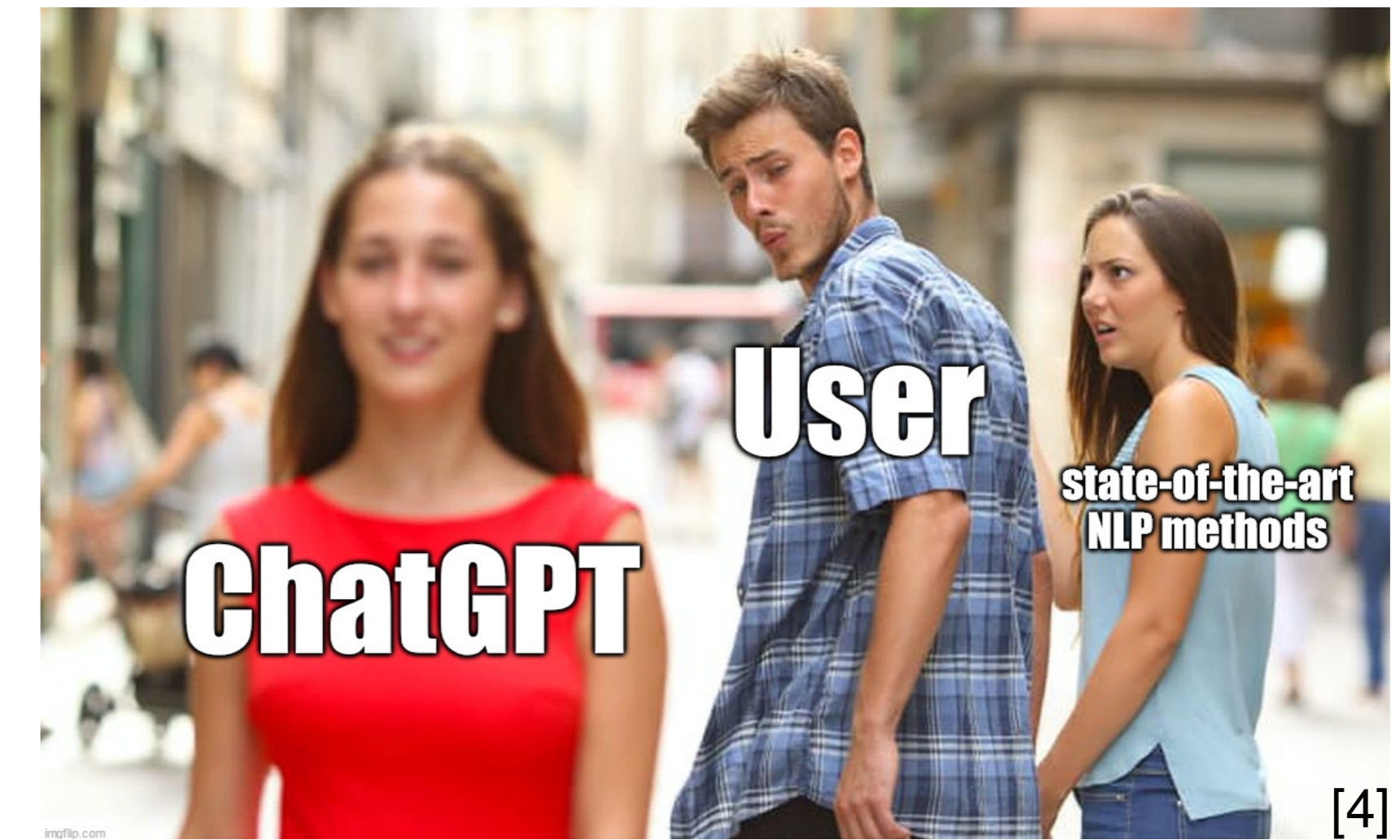
Waffen AND *L-/liefer* [deliver*] (in one sentence)
 OR militärisch* [military] & U-/unterstütz* [support] (in one sentence)

3. Few-shot learning using SOTA NLP methods [3]

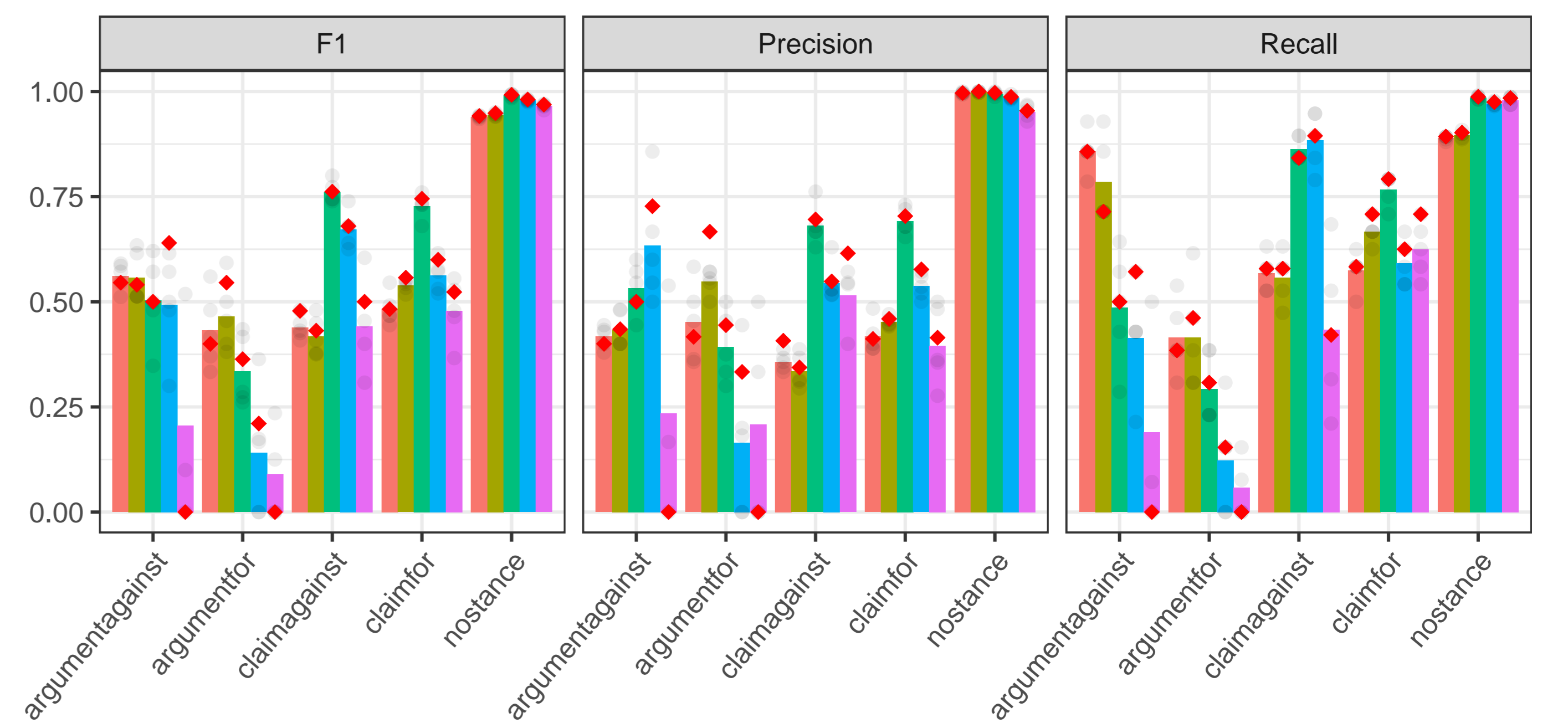


- Actors shuffling is an effective **pre-processing** step (*not displayed on poster*)
- Near-domain pre-training (cf., **SAM**) is promising
- PET-like models [9] are effective **few-shot** learners
- PETapter** combines few-shot technique [9] with parameter-efficiency [7]

4. Scrutinizing ChatGPT



Model: GPT-4 (0) GPT-4 (0, but ld) PETapter (1364) PETapter (272) PETapter (33)



- GPT-4 works far better than GPT-3.5 (*not displayed on poster*)
- GPT-4 works far better zero-shot than few-shot (*not displayed on poster*)
- GPT-4 performance more balanced over different labels
- PETapter with better performance for claims, worse for arguments

ChatGPT prompt:

In the frame of this coding task, we are interested in published viewpoints regarding the (in)appropriateness of arms deliveries to Ukraine. This includes any type of arms, e.g., defensive or heavy arms; any type of delivery schemes, e.g., direct or in cooperation with other countries; and any other aspects not mentioned here, e.g., purchase from the manufacturer or donation from the Bundeswehr stocks, etc. We assign one of the following five labels for each "target sentence": "argument for", "argument against", "claim for", "claim against", "no stance". We expect the following proportions per label: "argument for": 3%, "argument against": 4%, "claim for": 6%, "claim against": 5%, "no stance": 82%. An argument is a prescription or demand or viewpoint regarding (in)appropriateness of arms supplies from Germany to Ukraine with explanation(s) of such position. In contrast, a claim is a prescription or demand or viewpoint regarding (in)appropriateness of arms deliveries to Ukraine without explanation(s) of such position. For the decision which of the five labels is appropriate, we look at two sentences before and after a potential argument or claim if there is not enough information within the target sentence itself. As we work with the German newspaper discourse whose target audience are German readers, we consider general mentions of arms deliveries to Ukraine as relevant. Non-relevant ("no stance") are only those mentions of arms deliveries that name specific non-German actors, e.g., the US, France, Poland (i.e., arms deliveries that are explicitly not from Germany). We also consider relevant the mentions of arms deliveries from the entities and organizations that include Germany, e.g., the EU, NATO, the West, etc. Anything that is not arms deliveries should be labeled as "no stance", e.g. providing military training, sending helmets or humanitarian aid or troops, "closing the sky", the so-called "Ringtausch" etc. For each of the following 10 target sentences, give me the most appropriate of the five labels "argument for", "argument against", "claim for", "claim against", "no stance". If necessary, include the "context before" and the "context after" for your decision. Do not give me a reason or justification for your decision. Output the predictions in the following form:

- <prediction for first "target sentence">
- <prediction for second "target sentence">

Excerpt of test cases:

2. "target sentence": "Wir dürfen nicht nachlassen, die Ukraine zu unterstützen. "
 2. "context before": 12:43 Uhr - Nato rechnet mit langem Ukraine-KriegDer Krieg in der Ukraine könnte nach Einschätzung der Nato noch Jahre dauern. Darauf müsse man sich vorbereiten, sagte Nato-Generalsekretär Jens Stoltenberg der "Bild am Sonntag".
 2. "context after": Ähnlich äußerte sich der britische Premierminister Boris Johnson in einem Gastbeitrag für die "Sunday Times". 11:37 Uhr - Kasachstan ruft zur vollständigen Vernichtung von Atomwaffen aufDie Führung der Ex-Sowjetrepublik Kasachstan in Zentralasien hat vor dem Hintergrund des Ukraine-Kriegs dazu aufgerufen, bis 2045 weltweit alle Atomwaffen zu vernichten.

5. Still work to do

- Results are not satisfactory for individual classes
- Argument markers important for models; latent argumentative structures challenging
- (This) annotation task is difficult even for humans (Krippendorff's α : 0.66)
- Utilizing GPT-based models in the annotation process
- Identifying frames in debates

6. References

- C. Baden and N. Springer. Conceptualizing viewpoint diversity in news discourse. *Journalism*, 18(2), 2017.
- R. Benson. What makes news more multiperspectival? a field analysis. *Poetics*, 37(5), 2009. Fields in Transition, Fields in Action.
- A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov. Unsupervised cross-lingual representation learning at scale. In *ACL*. ACL, 2020.
- J. Kocoń et al. ChatGPT: Jack of all trades, master of none. *Information Fusion*, 99, 2023.
- A. Masini and P. V. Aelst. Actor diversity and viewpoint diversity: Two of a kind? *Communications*, 42(2), 2017.
- M. Maurer, J. Haßler, and P. Jost. Die Qualität der Medienberichterstattung über den Ukraine-Krieg. *Forschungsbericht für die Otto Brenner Stiftung*, 2023.
- J. Pfeiffer, A. Rücklé, C. Poth, A. Kamath, I. Vulić, S. Ruder, K. Cho, and I. Gurevych. AdapterHub: A framework for adapting transformers. In *EMNLP*. ACL, 2020.
- R. D. Precht and H. Welzer. *Die vierte Gewalt: Wie Mehrheitsmeinung gemacht wird, auch wenn sie keine ist*. S. Fischer, 2022.
- T. Schick and H. Schütze. Exploiting cloze-questions for few-shot text classification and natural language inference. In *EACL*. ACL, 2021.