Taking the green pill: Psychological distance and the climate crisis in the European Parliament*

Lorcan McLaren
School of Politics & International Relations
University College Dublin
lorcan.mclaren@ucdconnect.ie

Abstract

Climate change presents an existential risk to global society, yet public engagement has not risen to meet this challenge. Although research suggests that manipulating psychological distance could enhance the effectiveness of climate communication, it remains unclear whether politicians employ this strategy. This paper addresses this gap by developing and validating automated methods to measure psychological distance in political speech. I introduce a dataset of 35,000 speeches from the European Parliament on climate-related topics from 2014 to 2023, translated into English to bypass multilingual modelling challenges. I present a novel tool designed to streamline the annotation of long-form texts, enhancing efficiency and consistency among annotators. Finally, I explore and validate the use of generative language models to synthesise training data for fine-tuning bidirectional encoder models in order to measure psychological distance in text. This approach demonstrates strong performance and allows for scalable analysis of complex phenomena in natural language. It paves the way for future substantive research on the dynamics of European climate politics.

"Mr President, Hello? Hello? Are you aware that there is a climate crisis?"
— *Nikolaj Villumsen, GUE/NGL*

1 Introduction

In liberal democracies, climate action hinges on public buy-in, requiring citizens to engage in voluntary actions such as housing retrofits, switching to sustainable transportation, or reducing meat consumption. While policy instruments like taxation, subsidies, and loans can steer behaviour toward lower-impact choices, such measures can also become politically unpopular, risking electoral backlash for the implementing party. Instances of public resistance are already observable, such as the repeated destruction of cameras used to enforce ultra-low emissions zones in the UK.

Engagement with climate action is influenced by political ideology to such an extent that persuasion through presentation of evidence for climate change and its impacts cannot be relied upon. In fact,

*Working paper prepared for the 6th Annual COMPTEXT Conference, 2024

Luo and Zhao (2019) demonstrate a phenomenon of motivated attention, where political orientation (liberal vs. conservative) determines the level of attention paid to evidence for climate change. Furthermore, they show that evidence can have a backfire effect among conservatives, where exposure to motivationally inconsistent information not only fails to encourage pro-climate behaviours but may actually decrease support for climate action, exacerbating polarisation.

One psychological framework that offers useful insights for climate communication is Construal Level Theory [CLT] (Liberman, Trope, Stephan, et al., 2007; Trope & Liberman, 2010). CLT claims that the perceived psychological distance from an event or object influences how abstractly or concretely it is conceptualised. This "distance" can manifest across four dimensions: temporal, spatial, social, and probabilistic. Events perceived as distant are thought of in more abstract terms, whereas those seen as near are considered more concretely. Psychological distance refers to how removed an event or object feels from an individual's direct experience. Understanding the effects of this distance on perception and behaviour is crucial for enhancing public engagement with climate issues. For example, discussing climate change as a distant, abstract threat might result in generalised concern without action, whereas framing it as an immediate, local issue could stimulate concrete responses and behavioural changes.

Effective communication can bridge psychological distance. This can be achieved through sensory-rich language that makes an event feel immediate and real, or through the use of localised examples that resonate more personally with an audience. For instance, a politician might evoke a stronger response by discussing climate impacts within one's own community or constituency rather than referring to distant events.

The following excerpts from speeches in the European Parliament are examples of how the climate crisis can be discussed as psychologically proximate:

Karima Delli, Greens/EFA –
"Our house is burning, our lungs are clogging up and you'd rather look elsewhere."

Krzysztof Hetman, EPP –
"Europeans are experiencing the negative consequences of exceeded air standards more often and more severely, cities are fighting smog, and the number of diseases resulting from pollution is also increasing."

Martin Hojsík, Renew –
"Europe's forests, our natural wealth, are under threat. They are threatened not only by the climate crisis, but also by developers, mismanagement, fires and deadly chemicals. Their devastation is damaging our entire society."

Delara Burkhardt, S&D –
"Madam President, the people of Venice have seen in recent weeks that the climate emergency is not something abstract but a brutal reality here in Europe. The floods inundated 80% of the city and caused millions of euros worth of damage. These events show once again that we must not just pay lip service."

Despite a robust body of experimental research suggesting that manipulating psychological distance can significantly enhance the persuasiveness of climate communication — increasing both concern and the inclination to act — there appears to be no existing scholarship examining whether politicians actually employ this technique, consciously or otherwise. This paper is among the first to explore the psychological distance of the climate crisis in political speech. In particular, I focus on the impacts of climate change and pollution as the object, and seek to measure implied psychological distance from this object in parliamentary speech. I introduce a dataset of 35,000 speeches from the European Parliament on climate-related topics from 2014 to 2023, which are transformed into English using machine translation to circumvent the complexities of multilingual modelling. Additionally, I present a new tool designed to facilitate the annotation of long-form text. Finally, I demonstrate the effectiveness of employing generative language models to synthesise training data in order to fine-tune bidirectional encoder models. The performance of both the generative language models and bidirectional encoder model is validated against human-annotated ground truth. I conclude with a discussion of future technical and substantive work I intend to undertake in this area.

2 Related work

2.1 Defining psychological distance

Construal level theory posits that the psychological distance of an object or event influences its mental representation, where distance can be characterised across four dimensions: spatial distance, temporal distance, social distance, and probability (Trope & Liberman, 2010). According to CLT, events or objects that are psychologically near are represented more concretely, whereas distant events are represented more abstractly. This dichotomy affects how individuals predict, evaluate, and plan for future events (Liberman, Trope, McCrea, & Sherman, 2007). Importantly, the process is bidirectional: increasing the concreteness of an object's representation can reduce its perceived psychological distance, influencing behaviour (Maglio et al., 2013b).

Moreover, psychological distance influences emotional response; distant objects tend to reduce the intensity of emotions felt, while abstract thinking often leads to more positive reflections (Williams et al., 2014). Concomitantly, reductions in perceived distance are associated with stronger emotional reactions (Van Boven et al., 2010). The orientation of an object also matters: negative events or objects that are imagined as moving closer in spatial terms to the self elicit more negative responses and higher levels of arousal than negative events that are spatially static (Davis et al., 2011). There is also a bidirectional relationship between psychological distance and an individual's focus on causes versus consequences. Psychologically distant objects tend to elicit a focus on causes, particularly if they are temporally or socially distant, and thinking about causes generates a sense of psychological distance (Rim et al., 2013). Focusing on consequences should therefore have the effect of rendering an object more psychologically proximate.

2.2 Psychological distance and the climate crisis

The concept of psychological distance has been extensively studied in the context of climate communication. Survey research has shown that populations often perceive climate change as a distant, abstract problem, resulting in a reduced sense of personal concern and urgency (Leviston et al., 2014; Spence et al., 2012). However, psychological proximity is generally associated with higher levels of concern (Jones et al., 2017). In particular, spatial and probabilistic proximity predict higher levels of concern and engagement, while temporal distance is less significant (Singh et al., 2017). Strategies that reduce psychological distance can potentially enhance public engagement in climate issues. While reducing spatial distance by citing local examples of climate impacts can proximise climate change, perceived social distance can also be bridged by emphasising as a shared global identity to similar effect (Loy & Spence, 2020).

Despite these insights, efforts to manipulate psychological distance have yielded mixed outcomes. While some studies report increased engagement when psychological distance is minimised, others find no significant effect on behaviour change or support for climate action (Schuldt et al., 2018; Valkengoed et al., 2023; Wang et al., 2019). This variability may stem from differences in measurement approaches and whether psychological distance is viewed as a stable trait or a malleable perception (Brügger, 2020; Keller et al., 2022).

The effect of psychological distance may also depend on cognitive style, where psychological closeness coupled with an analytic cognitive mindset leads to elevated concern and behaviour intentions, whereas the effect was lessened when individuals are in a holistic mindset (Sacchi et al., 2016). The effect of psychological proximity on support for climate action is attenuated by belief in policy effectiveness (Brügger et al., 2015; Singh et al., 2017). Some research suggests that describing both the proximal and distal impacts of climate change simultaneously could be more effective than focusing on one type of impact alone (Brügger et al., 2016; Ejelöv et al., 2018). The credibility of the source and the perceived accuracy of the information also play crucial roles in how messages are received and the psychological distance felt by the audience (Maglio et al., 2013a). Interventions that manipulate psychological distance could overcome the issue of motivated attention, as the effect of political ideology on one's views is reduced for proximal objects compared to distal (Brügger, 2020; Chu & Yang, 2018).

The relationship between emotion and psychological distance in the context of the climate crisis is notable; increased distance can foster hope, whereas decreased distance often triggers fear and anger, which in turn can influence attitudes towards mitigation efforts (Chu & Yang, 2019). However, the use of fear-inducing communication must be approached with caution to avoid potential backfire effects (O'Neill & Nicholson-Cole, 2009).

Consensus on the utility of manipulating psychological distance for climate communication has not been reached as the literature continues to develop. Regardless of the outcome, it remains a potentially useful lens through which to study the framing of climate change.

2.3 Measuring psychological distance in text

Having established the relevance of psychological distance to climate communication, I now move on to the problem of measurement. Previous work on this topic has collapsed the four established dimensions of psychological distance into a single scale that runs from abstract to concrete.

The Linguistic Category Model [LCM] (Semin & Fiedler, 1988) is one framework for analysing language that reflects varying levels of abstraction, which may indicate different degrees of psychological distance. This model categorises verbs and adjectives into levels from concrete to abstract: descriptive action verbs, interpretive action verbs, state verbs, and adjectives. The granularity of language used in text provides insights into the speaker's or writer's perceived psychological distance from the content discussed. Johnson-Grey et al. (2020) introduce an automated method, Syntax-LCM, to score documents according to LCM.

Brysbaert et al. (2014) used crowd-sourcing to compile continuous numeric concreteness ratings for around 40,000 unigrams and bigrams, which serve as a resource for measuring psychological distance in written or spoken communication. Subsequent work by Snefjella and Kuperman (2015) and, more recently, by Yeomans (2021) has applied this set of ratings using a dictionary approach.

However, bag-of-words approaches such as dictionaries are unaware of the contextual nuances of word usage and can result in incorrect coding, particularly in instances of polysemy. Furthermore, they can struggle to maintain validity across domains other than those they were developed for. The results of applying Brysbaert et al.'s ratings with the `doc2concrete` R package (Yeomans, 2021) to my dataset, for example, exhibited very little variation. Yeomans (2021) concedes that the Brysbaert et al. measure should only be used as a starting point where good training data is not available. In light of this, my proposed method intends to establish and validate domain-specific training data. Furthermore, transformer models such as BERT (Devlin et al., 2019) offer significant advantages for text classification by capturing the contextual relationships between words, which leads to a deeper understanding of text nuances and more accurate classification results.

3 Data

The dataset of almost 35,000 climate-related speeches from 2014 to 2023 was constructed by combining ParlEE data, an existing dataset of parliamentary speeches, and web scraping. The speeches were then filtered to limit the dataset to climate-related topics using keywords. Finally, the speeches were translated to English using the DeepL API after a process of validation.

3.1 Data collection

ParlEE v2 (Sylvester et al., 2023) gathers parliamentary speeches from the European Parliament and Finish, Dutch and Croatian legislatures among others for the period 2009-2019. The dataset provides annotations at the quasi-sentence level which were not necessary for my use case, so speeches were reconstituted in their entirety. Speeches from more recent parliamentary sittings were scraped directly from the website of the European Parliament and parsed in order to align with the format and metadata of the ParlEE dataset.

In order to constrain the dataset to climate-related discourse, I filtered the speeches based on their agenda title using keywords. Keywords were determined by identifying frequent n-grams in the agenda titles of units labelled with the the Comparative Agendas Project's environmental code, which are provided at the quasi-sentence level in the ParlEE dataset. The keyword set was iteratively expanded in order to minimise false positives.

The resulting dataset contains 34,483 speeches ($\mu = 178.99; \sigma = 127.80$)[1] from the beginning of the eighth European Parliament in July 2014 up to the end of 2023. This period covers the lead-up to and ratification of the Paris Agreement as well as several significant packages of legislation including: the European Green Deal, the Fit-for-55 package, the Green Deal Industrial Plan, and the Nature Restoration Law. It consists of debates that explicitly reference core issues in climate action such as carbon emissions and biodiversity loss, but also spans all debates related to polluting or extractive industries – including farming, fishing, mining, energy production, waste management, and transport among others.

[1] Length in tokens

3.2 Machine translation

The European Parliament is a highly multilingual environment, with speakers permitted to participate in any of the 24 official EU languages. Interpreters translate each speech into other official languages in real time so that all MEPs can follow the discussion. However, speeches are transcribed only in the original language.

Though significant improvements have been made in the performance of multilingual models, performance can still vary across languages. Given the linguistic diversity in this dataset, including some relatively under-resourced languages, I opted to translate all data into English before proceeding.

Two industry-standard tools for machine translation are: Google Cloud Translate and DeepL. Both are available via API at a similar cost per token of input.

In order to compare their performance, I created 3,000-sentence random samples of parallel data for each language pair (<*source language*>-<*target language*>, with English being the target language in each case e.g. French-English, Slovak-English) from the Europarl corpus (Koehn et al., 2003), translated these with each tool and used automated measures to compare their performance.

Figure 1 shows translation performance across all language pairs measured in terms of BERTScore (Zhang et al., 2020), BLEU (Papineni et al., 2001), METEOR (Banerjee & Lavie, 2005), NIST (Doddington, 2002), ROUGE (Lin, 2004) and SacreBLEU (Post, 2018). In each case, the output of the translation tool is compared to the human-generated reference translation from Europarl. All metrics are scaled using the max norm for the purpose of visual comparison.

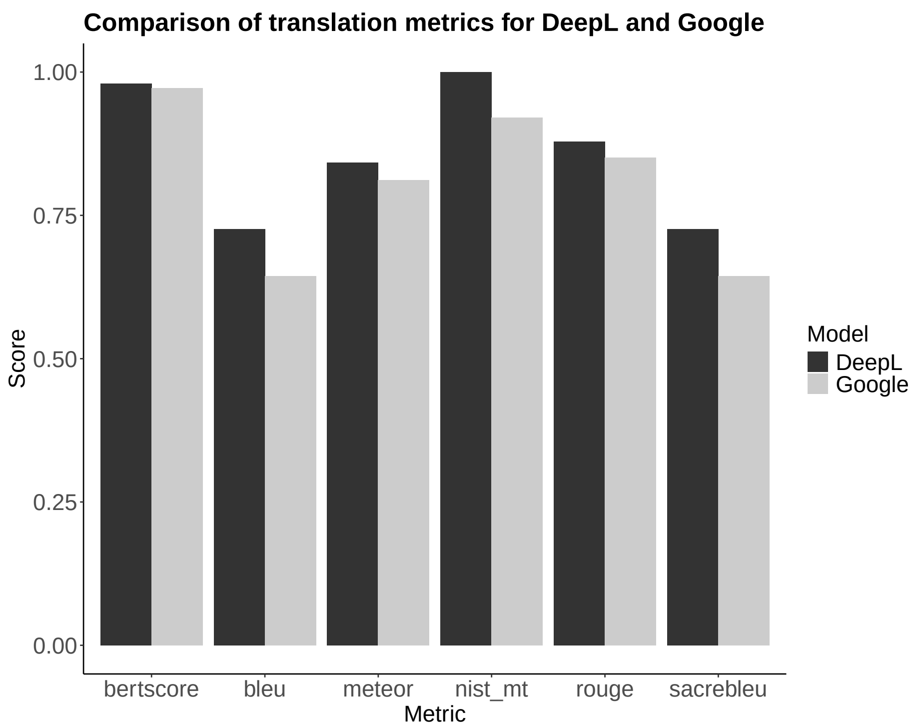


Figure 1: Machine translation performance by service

I find that DeepL outperforms Google Cloud Translate by every measure. Appendix A shows translation performance for each language pair, similarly demonstrating the superior performance of DeepL, with perhaps the exception of Italian-English. Manual inspection of the translations also gave the impression that DeepL generates a more coherent output. Consequently, DeepL was used to translate the dataset. Maltese and Irish are excluded from the Europarl corpus and thus cannot be validated.

Since MEPs frequently speak in multiple languages in a single speech, including languages other than those of the nation they represent, I performed language identification prior to translation. Each speech was segmented into sentences, the language of the sentence identified using the `langdetect` package in Python, and then consecutive sentences of the same language within each speech were concatenated together. These monolingual sequences, along with the identified language code, were then passed to the DeepL API. Sequences already in the target language, English, were excluded from this process. The translation output was then reconstructed into speeches.

The translated speeches, along with replication materials for scraping, translation and validation have been made available[2].

4 Methods

This paper contributes to a broader project that examines the interplay of moral and emotional language with psychological distance in political discourse on climate change.

Given the size of the dataset, the content analysis process must be automated (Grimmer & Stewart, 2013). However, creating a large and reliable training dataset for supervised machine learning presents challenges due to the complexity and scope of the annotation tasks involved.

This study seeks to assess whether generative language models can be used to synthesise training data for fine-tuning bidirectional encoder models (Meng et al., 2022; Møller et al., 2024; Wang et al., 2021). Since many of the leading generative language models are closed-source, local fine-tuning of models is important for replicability, interpretability and transparency. Moreover, this approach offers efficiency benefits by reducing the substantial energy and financial costs typically associated with generative language models, especially when applied to large datasets.

To ensure the methods employed are reliable, establishing ground truth is essential. The following section details the development and use of a novel text annotation tool, which supports the content analysis of long-form text.

4.1 Validation data

Extensive data annotation is needed to validate the computational techniques used throughout the broader project. Consequently, I developed a text annotation tool[3] to improve the process of establishing ground truth. Although similar tools are available, such as CCS Annotator (AnnoTinder), this tool provides a few key benefits. It enables rapid development and testing of codebooks in a no-code environment, using a graphical user interface, in hopes that it can be accessible to a broader array of scholars. It also permits swapping out datasets and annotation schemata via portable, interoperable file formats such as CSV or JSON. It is particularly useful where the unit of analysis is long-form text rather than quasi-sentences or tweets, which are more amenable to spreadsheet

[2] Available at https://github.com/LorcanMcLaren/europticon

[3] Available at https://green-pill.streamlit.app/

software for labelling. Figure 2 shows an example text and the accompanying instructions and annotation options used for this paper.

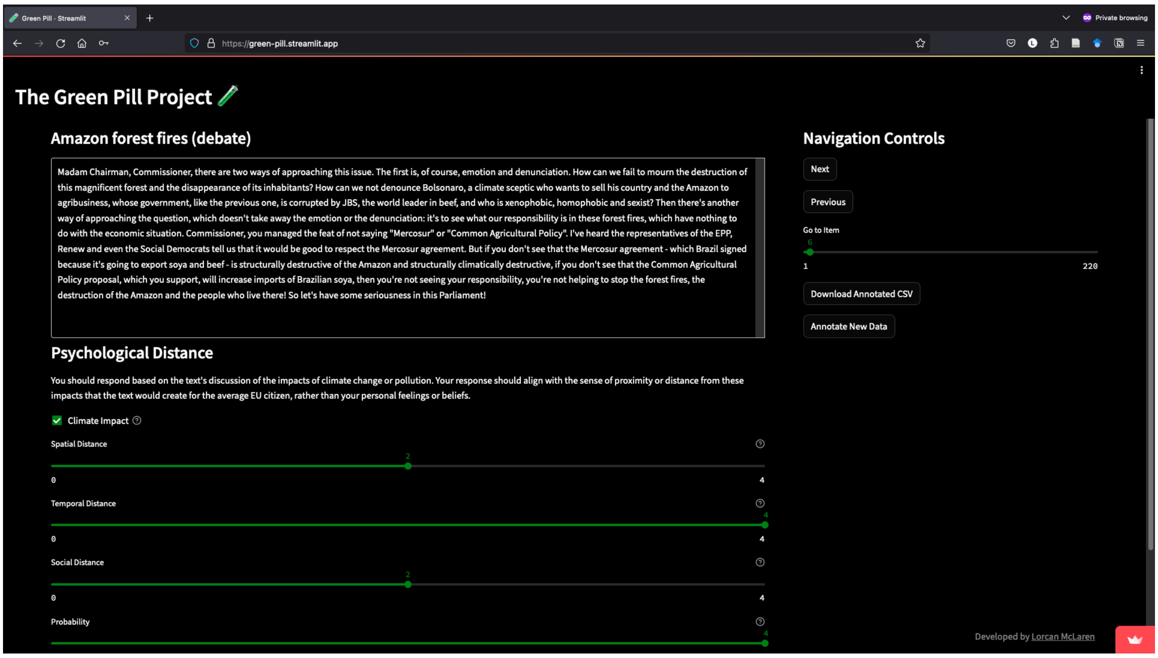


Figure 2: Annotation tool for long-form text

A human annotator labelled a set of 220 parliamentary speeches for psychological distance. First, they indicated whether each text discussed the impacts of climate change or pollution and, if so, they responded to the following items using a 4-point Likert scale, where 1 = 'Disagree', 2 = 'Slightly disagree', 3 = 'Slightly agree', and 4 = 'Agree':

- Spatial distance: "The climate impacts described in the text feel close to my current location."
- Temporal distance: "The climate impacts described in the text are occurring now, in the very near future, or have already occurred."
- Social distance: "The people or social groups impacted by climate change in the text are similar to me or my social group."
- Probability distance: "The climate impacts described in the text are likely to occur."

The annotator was instructed that their response in each case should align with the sense of proximity or distance from these impacts that the text would create for the average EU citizen, rather than their personal feelings or beliefs.

4.2 Synthesising training data with generative language models

To synthesise adequate training data for model fine-tuning, a two-stage generative language modelling process was used. This section describes how I leveraged GPT-3.5 (Brown et al., 2020) and GPT-4 [4] (OpenAI et al., 2024) to automatically annotate a subset of the data for psychological distance using the same codebook as the human annotator.

For each speech, the model was asked to assess whether the text discussed the impacts of climate change or pollution and, only if it responded in the affirmative, to assess the text for the four dimensions of psychological distance. Performance of GPT-3.5 and GPT-4 are compared.

Initial experimentation with different prompt styles had found that prompting the model with a statement and asking whether it agreed or disagreed with it led to results that were more in line with human responses. This was particularly true in the case of yes/no questions. For example, *"Does this text discuss the impacts of climate change or pollution?"* led to significantly more affirmative responses than *"This text discusses the impacts of climate change or pollution. Respond with a number where 1 = 'Agree' and 0 = 'Disagree'"*. The same codebook used for human annotation was transformed directly into prompts. See section 4.1 for the specific statements used.

After assessing the quality of the synthetic training data with respect to human-annotated validation data, I apply the generative language model using the same two-stage process to label 2,000 speeches. The next subsection details the training and validation of the bidirectional encoder model on this synthetically-annotated dataset.

4.3 Bidirectional encoder modelling

BERT (Devlin et al., 2019) is a large pre-trained language model based on a bidirectional encoder architecture that may be used for a variety of NLP tasks, including natural language generation, text summarisation, question answering, and classification. While it provides a general model of English, it must generally be adapted to a specific dataset and task through a process of parameter optimisation known as fine-tuning. This paper fine-tunes a subvariant of the original BERT model, DistilBERT. DistilBERT (Sanh et al., 2020) uses knowledge distillation during the pre-training phase to reduce the size of a BERT model while providing faster training and processing, and a majority of the same language understanding capabilities.

In this study, DistilBERT was fine-tuned for five epochs on the synthesised training data, using an 80/20 split where 1,600 speeches were used for training and 400 for validation. The performance of the model was then tested using the human-annotated data. Detailed performance metrics are reported in section 5.2.

[4] Specifically, `gpt-3.5-turbo-0125` and `gpt-4-0125-preview`, which are referred to as GPT-3.5 and GPT-4 throughout the remainder of the paper.

5 Results

5.1 Synthetic training data quality

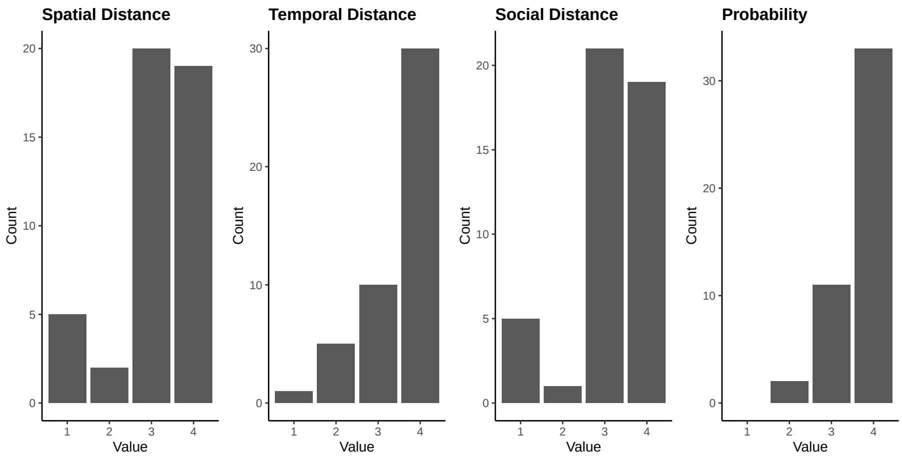
The quality of synthetic training data generated by the models is evaluated using macro F1 scores, as shown in Table 1. These scores reflect the correspondence between model outputs and human annotations across different psychological distance dimensions.

Table 1: Macro F1 scores across models

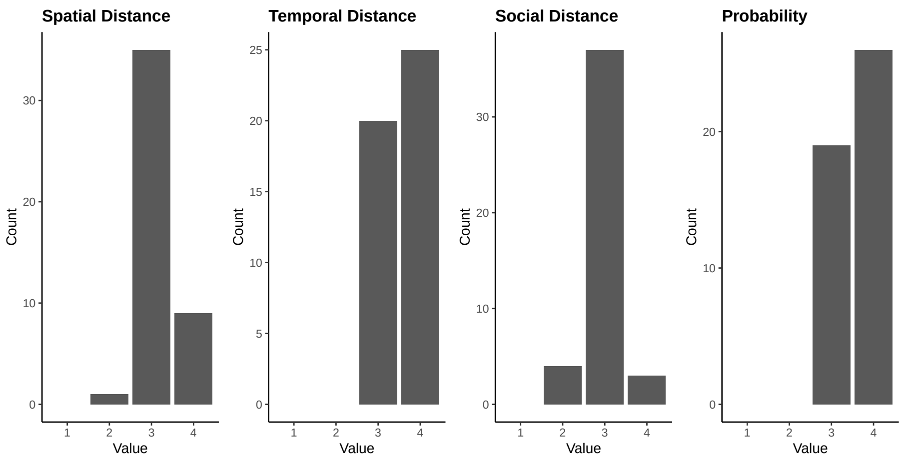
Annotation	**GPT-3.5**	**GPT-4**	**GPT-3.5 (Merged)**	**GPT-4 (Merged)**
Climate Impact	0.846	0.847	0.846	**0.847**
Spatial Distance	0.341	0.463	0.803	**0.875**
Temporal Distance	0.298	0.323	0.792	**0.824**
Social Distance	0.324	0.446	**0.801**	0.792
Probability	0.371	0.322	0.813	**0.837**

The intial performance, reported in first and second columns of Table 1, was underwhelming. Looking at Figure 3, we can see this is due to the notably different distributions in Likert-scale responses. Examining the distribution of responses from GPT-4, for instance, I find that it tends to respond more emphatically in either direction in the case of temporal distance and probability, returning only 1s or 4s. The human annotator, by contrast, also makes use of the 2 and 3 response options in these cases. Notably, GPT-3.5 and GPT-4 also produce consistently different response distributions from each other.

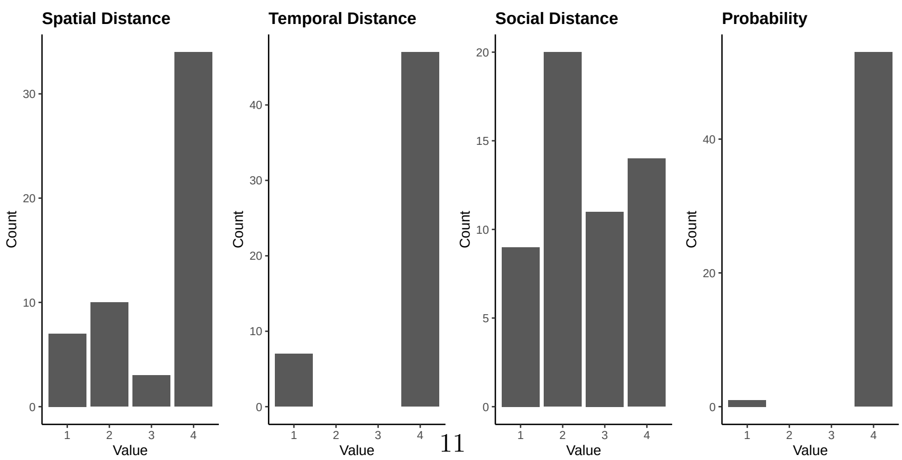
However, I applied post-processing to reduce the overall number of classes from four to two, by merging any 1s and 2s ('Disagree' and 'Slightly disagree') into one category and 3s and 4s ('Slightly agree' and 'Agree') into a separate category. This leads to a sizeable bump in performance metrics, as seen in the third and fourth columns of Table 1. The performance of both generative language models is now very strong, suggesting that they capture overall valence well. Their primary point of disagreement with the human annotator, prior to post-processing, is magnitude.



(a) Human annotator response distribution



(b) GPT-3.5 model response distribution



(c) GPT-4 model response distribution

Figure 3: Comparison of 4-point Likert-scale response distributions for human annotator, GPT-3.5, and GPT-4

The literature indicates that dimensions of psychological distance to a large extent capture the same underlying construct and should exhibit a strong relationship (Liberman & Trope, 2014; Spence et al., 2012). Figure 4 shows this to be the case for our dataset.

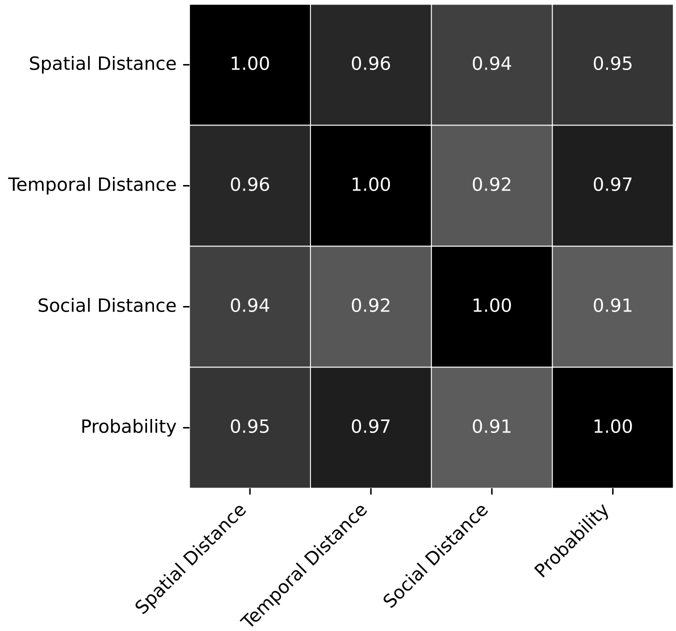


Figure 4: Correlation among dimensions of psychological distance in synthetic training data

5.2 Bidirectional encoder model performance

Table 2 shows the performance of the DistilBERT model on 220 items of human-annotated data after being fine-tuned exclusively on the 2,000 items of synthetically-annotated training data.

Table 2: DistilBERT model performance after 5 epochs of fine-tuning

Annotation	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Climate Impact	0.891	0.838	0.848	0.842
Spatial Distance	0.873	0.810	0.734	0.762
Temporal Distance	0.900	0.842	0.829	0.835
Social Distance	0.868	0.867	0.657	0.699
Probability	0.900	0.844	0.865	0.853

Performance across all dimensions appears strong. In fact, referring back to Table 1, DistilBERT performance actually exceeds that of GPT-4 in the case of temporal distance and probability. Performance for social distance lagged slightly behind other metrics.

6 Discussion and future work

This study has demonstrated the utility of generative language models to synthesise training data to fine-tune bidirectional encoder models. Rephrasing binary questions as statements with binary response options significantly improved model performance, indicating that minor adjustments in prompt style can have substantial impacts on outcomes. The analysis also showed that while the models struggled with the original 4-point Likert scale, they were successful in capturing the valence of responses accurately. This suggests that future approaches to training data synthesis could benefit from focusing on binary or simpler categorical annotations to maintain model accuracy while minimising complexity. The strong performance across all dimensions in terms of F1 scores, post-merging of classes, suggests that generative models can effectively approximate human-like understanding of text under certain conditions. This is promising for applications in political content analysis, even where the subject of inquiry requires a nuanced understanding of context.

Looking ahead, employing an ensemble of generative language models might further improve the quality of synthetic training data. Establishing a minimum threshold for agreement among models could potentially increase the reliability of the generated annotations. Fine-tuning open-source generative language models for measuring psychological distance could retain the benefits of cost-efficiency and replicability while possibly providing superior performance over bidirectional encoder models, due to their larger parameter sets and more extensive training data. Additionally, to ensure the validation data is representative and robust, there are plans to engage more human annotators in future phases of the research.

Future substantive work will examine variation in the framing of climate impacts in the European Parliament using this novel measure of psychological distance. Investigating the relationship between the use of emotive language and psychological distance could reveal whether emotional intensity helps bridge perceived distances in political speech. Moralised language may serve a similar function. This could further our understanding of how politicians might strategically use moral-emotional appeals to make distant concepts feel more immediate and urgent. Analysing whether MEPs whose constituencies have recently experienced climate shocks such as flooding or wildfires alter their use of psychological distance could provide insights into how personal and regional experiences influence political rhetoric. Tracing changes in discourse over time will also shed light on whether significant packages of climate legislation lead to shifts in psychological distancing in political communication.

This research has laid a foundational framework for using advanced NLP techniques to analyse psychological distance in political speech about climate change. These methods establish the groundwork for substantive research to not only enhance our understanding of current discourse strategies but also open up new avenues for effectively communicating about climate change in a way that might better engage the public and prompt action. As climate action becomes ever more urgent, refining our communication strategies is increasingly vital.

References

Banerjee, S., & Lavie, A. (2005, June). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In J. Goldstein, A. Lavie, C.-Y. Lin, & C. Voss (Eds.), *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65–72). Association for Computational Linguistics. Retrieved April 24, 2024, from https://aclanthology.org/W05-0909

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. Retrieved April 28, 2024, from https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Brügger, A. (2020). Understanding the psychological distance of climate change: The limitations of construal level theory and suggestions for alternative theoretical perspectives. *Global Environmental Change*, *60*, 102023. https://doi.org/10.1016/j.gloenvcha.2019.102023

Brügger, A., Dessai, S., Devine-Wright, P., Morton, T. A., & Pidgeon, N. F. (2015). Psychological responses to the proximity of climate change [Publisher: Nature Publishing Group]. *Nature Climate Change*, *5*(12), 1031–1037. https://doi.org/10.1038/nclimate2760

Brügger, A., Morton, T. A., & Dessai, S. (2016). “Proximising” climate change reconsidered: A construal level theory perspective. *Journal of Environmental Psychology*, *46*, 125–142. https://doi.org/10.1016/j.jenvp.2016.04.004

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Chu, H., & Yang, J. Z. (2018). Taking climate change here and now – mitigating ideological polarization with psychological distance. *Global Environmental Change*, *53*, 174–181. https://doi.org/10.1016/j.gloenvcha.2018.09.013

Chu, H., & Yang, J. Z. (2019). Emotion and the Psychological Distance of Climate Change [Publisher: SAGE Publications Inc]. *Science Communication*, *41*(6), 761–789. https://doi.org/10.1177/1075547019889637

Davis, J. I., Gross, J. J., & Ochsner, K. N. (2011). Psychological distance and emotional experience: What you see is what you get: Emotion [Publisher: American Psychological Association]. *Emotion*, *11*(2), 438–444. https://doi.org/10.1037/a0021783

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *North American Chapter of the Association for Computational Linguistics*. https://api.semanticscholar.org/CorpusID:52967399

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram cooccurrence statistics. *Proceedings of the second international conference on Human Language Technology Research*, 138–145.

Ejelöv, E., Hansla, A., Bergquist, M., & Nilsson, A. (2018). Regulating Emotional Responses to Climate Change – A Construal Level Perspective [Publisher: Frontiers]. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00629

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Johnson-Grey, K. M., Boghrati, R., Wakslak, C. J., & Dehghani, M. (2020). Measuring Abstract Mind-Sets Through Syntax: Automating the Linguistic Category Model [Publisher: SAGE

Publications Inc]. *Social Psychological and Personality Science*, *11*(2), 217–225. https://doi.org/10.1177/1948550619848004
Jones, C., Hine, D. W., & Marks, A. D. G. (2017). The Future is Now: Reducing Psychological Distance to Increase Public Engagement with Climate Change. *Risk Analysis*, *37*(2), 331–341. https://doi.org/10.1111/risa.12601
Keller, E., Marsh, J. E., Richardson, B. H., & Ball, L. J. (2022). A systematic review of the psychological distance of climate change: Towards the development of an evidence-based construct. *Journal of Environmental Psychology*, *81*, 101822. https://doi.org/10.1016/j.jenvp.2022.101822
Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical Phrase-based Translation. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, 48–54. https://doi.org/10.3115/1073445.1073462
Leviston, Z., Price, J., & Bishop, B. (2014). Imagining climate change: The role of implicit associations and affective psychological distancing in climate change responses. *European Journal of Social Psychology*, *44*(5), 441–454. https://doi.org/10.1002/ejsp.2050
Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences*, *18*(7), 364–369. https://doi.org/10.1016/j.tics.2014.03.001
Liberman, N., Trope, Y., McCrea, S. M., & Sherman, S. J. (2007). The effect of level of construal on the temporal distance of activity enactment. *Journal of Experimental Social Psychology*, *43*(1), 143–149. https://doi.org/10.1016/j.jesp.2005.12.009
Liberman, N., Trope, Y., Stephan, E., et al. (2007). Psychological distance. In *Social psychology: Handbook of basic principles* (2nd ed., pp. 353–383, Vol. 2).
Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. *Text Summarization Branches Out*, 74–81. Retrieved April 24, 2024, from https://aclanthology.org/W04-1013
Loy, L. S., & Spence, A. (2020). Reducing, and bridging, the psychological distance of climate change. *Journal of Environmental Psychology*, *67*, 101388. https://doi.org/10.1016/j.jenvp.2020.101388
Luo, Y., & Zhao, J. (2019). Motivated Attention in Climate Change Perception and Action [Publisher: Frontiers]. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01541
Maglio, S. J., Trope, Y., & Liberman, N. (2013a). The Common Currency of Psychological Distance [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *22*(4), 278–282. https://doi.org/10.1177/0963721413480172
Maglio, S. J., Trope, Y., & Liberman, N. (2013b). Distance from a distance: Psychological distance reduces sensitivity to any further psychological distance: Journal of Experimental Psychology: General [Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, *142*(3), 644–657. https://doi.org/10.1037/a0030258
Meng, Y., Huang, J., Zhang, Y., & Han, J. (2022). Generating Training Data with Language Models: Towards Zero-Shot Language Understanding. *Advances in Neural Information Processing Systems*, *35*, 462–477. Retrieved April 27, 2024, from https://proceedings.neurips.cc/paper_files/paper/2022/hash/0346c148ba1c21c6b4780a961ea141dc-Abstract-Conference.html
Møller, A. G., Pera, A., Dalsgaard, J., & Aiello, L. (2024, March). The Parrot Dilemma: Human-Labeled vs. LLM-augmented Data in Classification Tasks. In Y. Graham & M. Purver (Eds.), *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 179–192). Association for Computational Linguistics. Retrieved April 27, 2024, from https://aclanthology.org/2024.eacl-short.17

O'Neill, S., & Nicholson-Cole, S. (2009). "Fear Won't Do It": Promoting Positive Engagement With Climate Change Through Visual and Iconic Representations [Publisher: SAGE Publications Inc]. *Science Communication*, *30*(3), 355–379. https://doi.org/10.1177/1075547008329201

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024, March). GPT-4 Technical Report [arXiv:2303.08774 [cs]]. https://doi.org/10.48550/arXiv.2303.08774

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). BLEU: A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, 311. https://doi.org/10.3115/1073083.1073135

Post, M. (2018, September). A Call for Clarity in Reporting BLEU Scores [arXiv:1804.08771 [cs]]. https://doi.org/10.48550/arXiv.1804.08771

Rim, S., Hansen, J., & Trope, Y. (2013). What happens why? Psychological distance and focusing on causes versus consequences of events: Journal of Personality and Social Psychology [Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *104*(3), 457–472. https://doi.org/10.1037/a0031024

Sacchi, S., Riva, P., & Aceto, A. (2016). Myopic about climate change: Cognitive style, psychological distance, and environmentalism. *Journal of Experimental Social Psychology*, *65*, 68–73. https://doi.org/10.1016/j.jesp.2016.03.006

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020, February). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter [arXiv:1910.01108 [cs]]. https://doi.org/10.48550/arXiv.1910.01108

Schuldt, J. P., Rickard, L. N., & Yang, Z. J. (2018). Does reduced psychological distance increase climate engagement? On the limits of localizing climate change. *Journal of Environmental Psychology*, *55*, 147–153. https://doi.org/10.1016/j.jenvp.2018.02.001

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language: Journal of Personality and Social Psychology [Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *54*(4), 558–568. https://doi.org/10.1037/0022-3514.54.4.558

Singh, A. S., Zwickle, A., Bruskotter, J. T., & Wilson, R. (2017). The perceived psychological distance of climate change impacts and its influence on support for adaptation policy. *Environmental Science & Policy*, *73*, 93–99. https://doi.org/10.1016/j.envsci.2017.04.011

Snefjella, B., & Kuperman, V. (2015). Concreteness and Psychological Distance in Natural Language Use [Publisher: SAGE Publications Inc]. *Psychological Science*, *26*(9), 1449–1460. https://doi.org/10.1177/0956797615591771

Spence, A., Poortinga, W., & Pidgeon, N. (2012). The Psychological Distance of Climate Change. *Risk Analysis*, *32*(6), 957–972. https://doi.org/10.1111/j.1539-6924.2011.01695.x

Sylvester, C., Ershova, A., Khokhlova, A., Yordanova, N., & Greene, Z. (2023, February). ParlEE plenary speeches V2 data set: Annotated full-text of 15.1 million sentence-level plenary speeches of six EU legislative chambers. https://doi.org/10.7910/DVN/VOPK0E

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance [Place: US Publisher: American Psychological Association]. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Valkengoed, A. M. v., Steg, L., & Perlaviciute, G. (2023). The psychological distance of climate change is overestimated [Publisher: Elsevier]. *One Earth*, *6*(4), 362–391. https://doi.org/10.1016/j.oneear.2023.03.006

Van Boven, L., Kane, J., McGraw, A. P., & Dale, J. (2010). Feeling close: Emotional intensity reduces perceived psychological distance: Journal of Personality and Social Psychology [Publisher:

American Psychological Association]. *Journal of Personality and Social Psychology*, *98*(6), 872–885. https://doi.org/10.1037/a0019262
Wang, S., Liu, Y., Xu, Y., Zhu, C., & Zeng, M. (2021, August). Want To Reduce Labeling Cost? GPT-3 Can Help [arXiv:2108.13487 [cs]]. https://doi.org/10.48550/arXiv.2108.13487
Wang, S., Hurlstone, M. J., Leviston, Z., Walker, I., & Lawrence, C. (2019). Climate Change From a Distance: An Analysis of Construal Level and Psychological Distance From Climate Change [Publisher: Frontiers]. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00230
Williams, L. E., Stein, R., & Galguera, L. (2014). The Distinct Affective Consequences of Psychological Distance and Construal Level [Publisher: Oxford University Press]. *Journal of Consumer Research*, *40*(6), 1123–1138. https://doi.org/10.1086/674212
Yeomans, M. (2021). A concrete example of construct construction in natural language. *Organizational Behavior and Human Decision Processes*, *162*, 81–94. https://doi.org/10.1016/j.obhdp.2020.10.008
Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020, February). BERTScore: Evaluating Text Generation with BERT [arXiv:1904.09675 [cs]]. https://doi.org/10.48550/arXiv.1904.09675

A Translation performance by source language

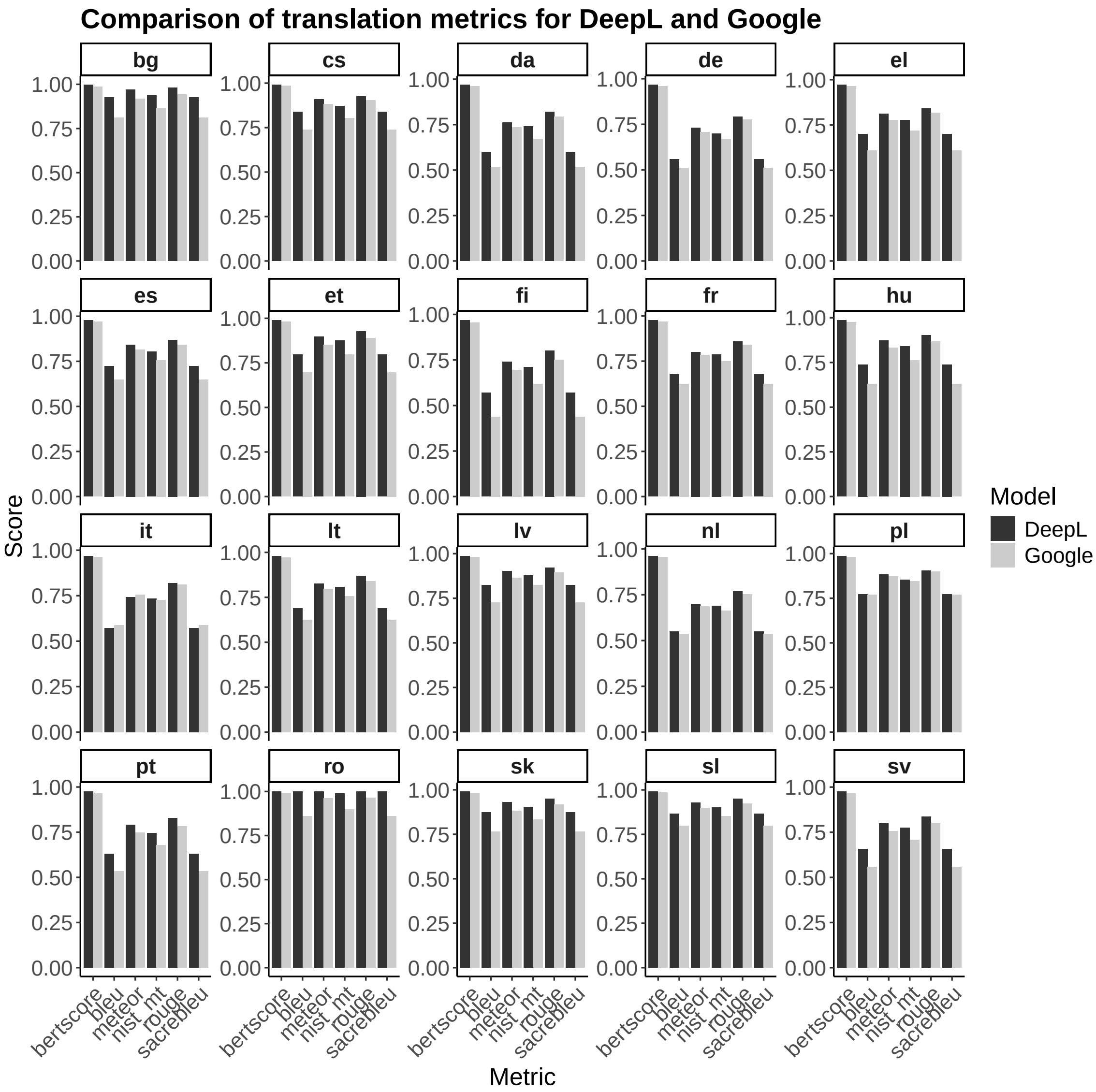


Figure 5: Machine translation performance by service for each language pair