Style-Aware Translation for Multi-Word Expressions in Low-Resource Languages

Antonios Dimakis[α,β], **Antonios Anastasopoulos**[α,δ]
[α]Archimedes, Athena R.C.
[β]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens
[δ]Department of Computer Science, George Mason University
andimakis@di.uoa.gr, antonis@gmu.edu

Abstract

While neural machine translation (NMT) has achieved remarkable parity with human performance in general domains, handling multi-word expressions (MWEs) remains a persistent challenge, particularly in low-resource settings. This difficulty is twofold: models need not only decode the non-compositional meaning of MWEs but also navigate the stylistic choice between idiomatic ("MWE") and literal ("plain") outputs. In this study, we introduce a new 4-way parallel dataset for Greek and English, comprising Greek and English MWE and plain sentences. We evaluate the style-transfer and style-switching capabilities of state-of-the-art Large Language Models (LLMs) across four tasks. Using traditional metrics, we demonstrate that LLMs often struggle with style-aware translation. Furthermore, we introduce StyleCOMET, a fine-tuned quality estimation metric to better evaluate stylistic nuances. Our results suggest that current limitations in MWE translation are a product of both model inability and metric-level insensitivity to idiomaticity.[1]

1 Introduction

Multi-word expressions (MWEs) are a cornerstone of natural language, accounting for a significant portion of any given lexicon (Sag et al., 2002; Candito et al., 2021). However, their inherent non-compositionality and morphosyntactically idiosyncratic behaviour (Baldwin and Kim, 2009) make processing them one of the most challenging tasks in natural language processing (NLP) (Ramisch et al., 2010). In lower-resource languages like Greek, general data scarcity can exacerbate this issue. Due to the nature of modern NLP, the need for in-context data and processing of MWEs, as opposed to handling pure lexical items with no context, has also been stressed (Dimakis et al., 2024).

When it comes to translation systems, beyond mere semantic accuracy, there is a need to offer stylistic flexibility (Havaldar et al., 2025). In our case, that would mean the ability for a model to understand the existence of MWEs in a source sentence, or lack thereof, and have the ability to include them in a translation or not, according to user desire. This task, which we define as MWE-Aware Translation, requires the model to decouple the core meaning from its stylistic encoding.

In this work, we present a comprehensive study of style transfer in the context of Greek-English MWE translation. Our contributions are as follows:

- We provide a human-validated dataset of 1,835 sentences, each containing four parallel versions: Greek MWE, Greek Plain, English MWE, and English Plain, **the first of its kind for any language pair**. An example is shown in Figure 1.
- We define MWE-Aware translation as an extension of plain MT, encompassing 4 tasks involving translation and style-switching.
- We use our tasks and dataset as a benchmark, and evaluate state-of-the-art LLMs on it.
- We analyse the shortcomings of standard COMET metrics in detecting stylistic accuracy and propose StyleCOMET, a fine-tuned version tailored for style-aware evaluation.

2 Related Work

Our work builds upon several recent advancements in evaluating and optimising style transfer, while specifically addressing the gap in cross-lingual MWE handling.

Prior research has focused on transferring stylistic attributes, though predominantly in monolingual contexts, known as Text Style Transfer (TST; Mukherjee et al., 2024, 2025; Pauli et al.,

[1]We release all code and datasets produced for this work: https://anonymous.4open.science/r/style_transfer.

Findings of the Association for Computational Linguistics: IJCNLP-AACL 2026, pages 2482–2513
November 6-10, 2026 ©2026 Association for Computational Linguistics

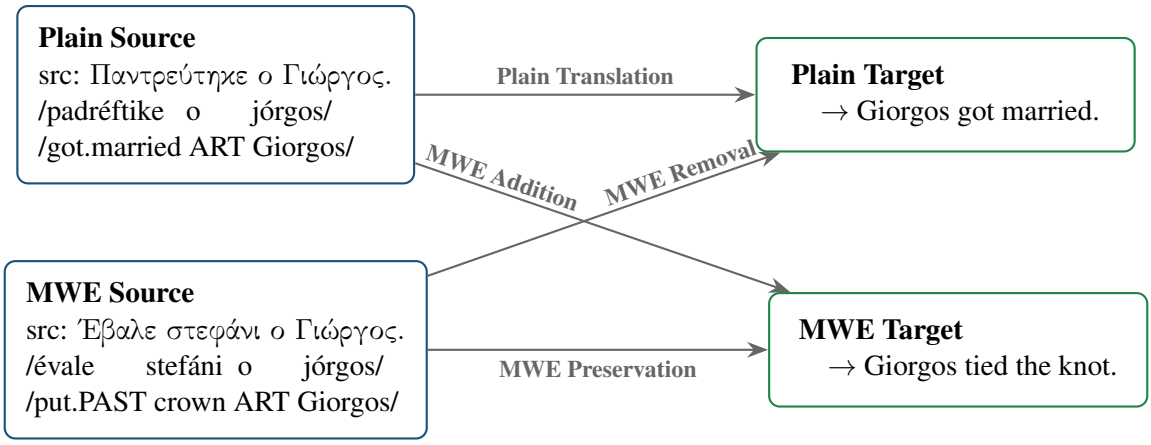


Figure 1: A 4-way parallel sentence example from our dataset and the 4 types of translation tasks we define.

2025; Liu and May, 2025). Work in style-aware cross-lingual translation (Havaldar et al., 2025) has defined style through cultural pragmatics and politeness rather than lexical or morphosyntactic idiomaticity, while also not utilising parallel datasets, precluding the use of reference-based metrics like COMET for stylistic evaluation.

Our work directly addresses the limitations identified in earlier studies of Greek MWE processing. Dimakis et al. (2024) used the IDION dataset (Markantonatou et al., 2019) to demonstrate that standard state-of-the-art MT systems, such as NLLB (NLLB Team et al., 2022), struggle significantly with Greek verbal MWEs due to MWEs' inherent scarcity and non-compositional nature. They noted that current models often revert to literal word-for-word translations, which are frequently nonsensical. While that study proposed a method to mitigate these failures, it emphasised that the lack of parallel datasets for low-resource languages like Greek remained a primary bottleneck for further research. Our work directly tackles these issues.

3 Dataset

We employed a professional translator to expand the IDION dataset to create a 4-way parallel corpus. For each original Greek sentence containing a verbal MWE, we produced:

1. **`EL-MWE`**: The original Greek idiomatic sentence
2. **`EL-Plain`**: A Greek paraphrased version using literal language
3. **`en-MWE`**: An English translation that utilises a MWE[2]
4. **`en-Plain`**: An English translation using literal language

The final dataset consists of 1835 × 4 sentences. This provides a ground-truth reference facilitating accurate translation evaluation when MWEs are involved as well as MWE-aware model training.[3]

Task	Gemini	Claude	Mistral
English → Greek			
`Plain Translation`	**0.8427**	**0.8454**	0.8384
`MWE Preservation`	**0.8221**	**0.8225**	0.8158
`MWE Removal`	0.8275	**0.8363**	0.8295
`MWE Addition`	0.8068	**0.8129**	0.8085
Greek → English			
`Plain Translation`	**0.8320**	**0.8310**	0.8282
`MWE Preservation`	**0.8199**	**0.8178**	0.8109
`MWE Removal`	**0.8181**	0.8118	0.8035
`MWE Addition`	0.7999	**0.8090**	**0.8070**

Table 1: StyleCOMET scores for LLM translation outputs evaluated against target style reference, on the test set. Statistically significantly higher score in row bolded, in case of ties both are bolded.

4 Task and Methodology

We define MWE-Aware translation as the ability to perform each of the following tasks:

1. **`Plain Translation`** (Plain → Plain): ensure that if the source does not use MWE, the hypothesis does not either

[2]The English MWE does not necessarily directly correspond to the Greek one.

[3]Two different alternatives have been provided in cases where gendered language is unavoidable in one language and unspecified in the other.

2. **`MWE Preservation`** (MWE → MWE): ensure that if the source uses a MWE, so does the output translation

3. **`MWE Addition`** (Plain → MWE): use a MWE in the translation even if the source does not employ one

4. **`MWE Removal`** (MWE → Plain): produce a translation using literal language even if the source contains a MWE

It is important to note that there may theoretically be certain cases of sentences in specific languages where it would only be possible to express the intended meaning by using a specific MWE or by not using one at all, making certain tasks impossible. However, given that we allow the insertion of a MWE at any part of the sentence for the **`MWE Preservation`** task, not just at the equivalent position to that of source sentence, this situation is rare. We have confirmed that all four tasks are well-defined for every sentence in our dataset.

4.1 Experimental Setup

We evaluate three models:[4] **Gemini 2.5 Flash** (Comanici et al., 2025), **Mistral Medium 3** (Liu et al., 2026), and **Claude Sonnet 4.5**. For every source sentence in our dataset, we prompt the models to perform the four MWE-Aware MT tasks in both directions (en↔el).

Figure 1 shows an example of these tasks with the ground truths from our dataset. The specific prompts used for each task are in Appendix A.

4.2 Evaluation Framework

We first run reference-free quality estimation on the ground-truth dataset itself, using **COMETKIWI** (`wmt22-cometkiwi-da`; Rei et al., 2022b) as a baseline for current quality estimation models using only human-generated data. We measure translations towards both directions and styles, therefore calculating 8 evaluations per instance. The full results are in Appendix C.1.1.

Turning to the LLM translations, we evaluate them against the references employing **COMET** (`wmt22-comet-da`; Rei et al., 2022a). To isolate the "style" component, each translation output was evaluated twice: once against the intended target style and once against the opposite style reference, to measure which one it was closer to, for a total of 16 evaluations per datum. The full results are in Appendix C.1.2.

StyleCOMET To address the insensitivity of standard COMET to stylistic nuances, we fine-tune it using a synthetic dataset by creating translation pairs from a train-subset of our sentence sets.

The purpose of this metric is not to replace standard COMET translation evaluation, but to provide an auxiliary measure specifically for measuring conformity to a specified target style.

We take all 8 combinations of tasks (`Plain Translation`, `MWE Preservation`, `MWE Addition`, `MWE Removal`) and translation directions (`el` → `en`, `en` → `el`) and set the reference to be either the same sentence as the hypothesis, or the equivalent but with the opposite style. We assign a high score to the former and a low score to the latter. Training details are in Appendix B.

We then use the test-subset to evaluate the LLM outputs the same way as we did with COMET. The full results can be found in Appendix C.1.3.

4.3 Statistical Testing

We perform t-statistic analyses to test several hypotheses for each evaluation framework.

For the COMETKIWI evaluations, we test whether translations across both languages are deemed better when maintaining style compared to switching it, and also whether plain-to-plain ones perform better than MWE-to-MWE.

For the COMET and StyleCOMET evaluations, for all tasks and both translation directions, we test whether a model's outputs are closer to the target style or the opposite one. We also test whether starting from the opposite style makes it harder to switch it in the translation, compared to maintaining the same style. Finally, for each of the 4 tasks per source language, we ascertain whether the Greek to English or English to Greek direction is easier for stylistic transfer.

The full results can be found in Appendix C.2.

5 Results and Analysis

Following the t-statistic analysis on the results, we observe the following:

Quality Estimation on the Dataset Evaluating COMETKIWI using the ground-truth dataset:

- Plain-to-plain translations are consistently rated higher than MWE-to-MWE.

[4] We use LLMs as dedicated MT models cannot be prompted to perform our style-aware tasks.

- Interestingly, for English-to-Greek, translations were evaluated higher when moving from MWE → Plain than when maintaining an MWE style.

The latter finding suggests that the default evaluation model struggles in processing sentences containing MWEs compared to more literal ones, even when not explicitly testing for style transfer.

Translation Task	Right Style	Wrong
[el->en] Plain Trans.	**0.7823**	0.7427
[el->en] MWE Preserv.	**0.7329**	0.6876
[el->en] MWE Removal	**0.7431**	0.7219
[el->en] MWE Addition	**0.7041**	**0.7024**
[en->el] Plain Trans.	**0.8568**	0.8057
[en->el] MWE Preserv.	**0.7975**	0.7756
[en->el] MWE Removal	**0.8269**	0.7870
[en->el] MWE Addition	0.7769	**0.7839**

Table 2: COMET evaluations on Gemini outputs, on the entire dataset, where the reference is that of the intended target style (Right Style) or the reverse (Wrong). Statistically significantly higher score in row bolded, if there is no significance both are bolded.

LLM Performance via COMET Using COMET to evaluate LLM outputs, they are generally deemed to be:

- uniformly better at English-to-Greek translation than the reverse.
- worse at switching styles during translation compared to maintaining the source style.
- particularly poor at English → Greek MWE Addition. In these instances, the outputs were consistently evaluated as being closer to the Greek-plain ground truth, regardless of intent. Relatively lower scores are also observed for Greek → English style-change tasks. An example of the results for Gemini is in Table 2.

The unexpected result where evaluations of translations towards the lower-resource ([en->el]) language are consistently higher compared to their equivalents towards the high-resource one ([el->en]) may be explained by COMET itself penalising the hypotheses more harshly for more minor differences in a language it has more training data on. Since plain sentences, and occasionally even MWEs, are able to be translated between languages using almost direct word-for-word translation, maintaining style being easier than switching was expected, and our results verify this.

Translation Task	Right Style	Wrong
[el->en] Plain Trans.	**0.8320**	0.8145
[el->en] MWE Preserv.	**0.8199**	0.7936
[el->en] MWE Removal	**0.8181**	0.8087
[el->en] MWE Addition	**0.7999**	0.7930
[en->el] Plain Trans.	**0.8427**	0.8118
[en->el] MWE Preserv.	**0.8221**	0.8072
[en->el] MWE Removal	**0.8275**	0.8057
[en->el] MWE Addition	**0.8068**	**0.8076**

Table 3: StyleCOMET evaluations on Gemini outputs, on the entire dataset, where the reference is that of the intended target style (Right Style) or the reverse (Wrong). Statistically significantly higher score in row bolded, if there is no significance both are bolded.

Since MWEs are relatively rare in written corpora, the last finding could be interpreted as difficulty in generating these relatively rare phrases in an already lower-resource language such as Greek.

LLM Performance via StyleCOMET We find that by using StyleCOMET to evaluate LLM outputs, we are much better at differentiating between idiomatic and plain variants, indicating that some of the issues identified by the COMET analysis previously lay with our evaluation metric, not with the models tested.

However even with our style-sensitive metric, the English-Plain → Greek-MWE task remains a significant failure point for LLMs, confirming the issue is rooted in the generation phase rather than just the evaluation. An example of the results for Gemini is in Table 3.

Examining a specific example of successful MWE Addition from the test set can be illuminating: "Αυτός ο Αντώνης ο Δραμινός θα με κατατροπώσει πάλι." → "This Antonis from Drama will wipe the floor with me again." The Greek source does not include any MWE, and the produced hypothesis does while also correctly conveying the meaning. But COMET incorrectly evaluated this instance as a stylistic failure, scoring it higher when using a Plain reference (0.5881) compared to using a MWE one (0.5729). StyleCOMET on the other hand scored it higher when using a MWE reference (0.7925) compared to a Plain one

(0.7756).[5]

Conversely, in an example of failed MWE Removal: "Έρχονται στιγμές που δεν βγάζουμε τσιμουδιά." → "There are moments when we don't say a word." The produced hypothesis contains a MWE ("not say a word"), while the intention is to remove it. But COMET incorrectly evaluated this instance as a stylistic success, scoring it higher when using a Plain reference (0.7830) compared to using a MWE one (0.7462), whereas StyleCOMET scored it higher when using a MWE reference (0.8277) compared to a Plain one (0.8148).[6]

A cross-LLM comparison of StyleCOMET scores on the test set, evaluated only on the intended target style per task, is in Table 1. While the task still proves challenging, Claude performs the best, with Gemini a close second.

6 Discussion and Future Work

Our experimental results (§5) show that there exists a significant blind spot in both modern LLMs and standard translation evaluation metrics. The creation of a 4-way parallel dataset allows us to somewhat disentangle translation accuracy from stylistic fidelity.

We find that while LLMs are surprisingly adept at understanding and translating MWEs into plain language, in contrast to what was indicated by prior studies testing older translation models, they lack the generative capacity to inject low-resource MWEs into a translation when starting from a literal source. Furthermore, our experiments demonstrate the danger of relying on standard metrics like COMET for stylistic tasks, as they inherently favor literal translations over valid idiomatic expressions.

Future Work This 4-way methodology could be scaled to additional language pairs to observe if this generative bottleneck exists across low-resource languages more generally, or if it is an artifact of our specific language pair.

The production of such datasets at a larger scale could also facilitate the inclusion of style-aware capabilities in future dedicated Machine Translation models, since these are not currently straightforwardly usable for our task, as they cannot be customized with a prompt where one could give specific instructions about style.

Limitations

Even though our dataset has been professionally translated and therefore we have ground truths for every data point, the LLM-provided translations have not been evaluated by humans. Therefore, this might conflate our analysis, because a low score can be either due to bad LLM translation performance or due to issues with the metric.

Our creation of StyleCOMET aims to address this limitation head-on, but naturally it is still another automatic metric that might suffer from similar or related issues. Moreover, since StyleCOMET has been trained on examples created from human-generated data, it has not been exposed to potentially flawed translations in the final fine-tuning stage. This potentially reduces the diversity of the training data compared to real-world translation scenarios. Of course, our StyleCOMET builds on the pre-trained COMET model which has indeed been exposed to all sorts of translation errors.

That said, manual inspection by the authors of a substantial sample of the translation outputs and the produced rankings by StyleCOMET makes us confident that the metric does measure what we built it to, prioritising stylistic similarity.

Ethics Statement

Due to the nature of our dataset, it being originally sourced from very informal online settings, certain sentences include harmful language or stereotypes. To ensure responsible research, we have explicitly annotated such examples within our dataset. Use of this dataset should be conducted with awareness of these data points.

We are releasing our dataset under a CC-BY-NC license. All models we use come with permissive licenses, at least when it comes to research. Their usage in this project is therefore consistent with its intended use.

Acknowledgments

This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program. Antonios Anastasopoulos is also supported by the US National Science Foundation under CAREER award 2439202.

[5]The plain English reference is "Antonis from Drama will defeat me decisively again.", while the MWE one is "Antonis from Drama will clean my clock again."

[6]The plain English reference is "There are moments when we stay completely silent.", while the MWE one is "There are moments when we don't breathe a syllable."

References

Timothy Baldwin and Su Nam Kim. 2009. Multiword expressions. In *Handbook of Natural Language Processing, Second Edition, CRC Press*, Boca Raton, USA. Nitin Indurkhya and Fred J. Damerau.

Marie Candito, Mathieu Constant, Carlos Ramisch, Agata Savary, Bruno Guillaume, Yannick Parmentier, and Silvio Cordeiro. 2021. A french corpus annotated for multiword expressions and named entities. *Journal of Language Modelling*, 8(2):415–479.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit S. Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, Luke Marris, Sam Petulla, Colin Gaffney, Asaf Aharoni, Nathan Lintz, Tiago Cardal Pais, Henrik Jacobsson, Idan Szpektor, Nan-Jiang Jiang, Krishna Haridasan, Ahmed Omran, Nikunj Saunshi, Dara Bahri, Gaurav Mishra, Eric Chu, Toby Boyd, Brad Hekman, Aaron Parisi, Chaoyi Zhang, Kornraphop Kawintiranon, Tania Bedrax-Weiss, Oliver Wang, Ya Xu, Ollie Purkiss, Uri Mendlovic, Ilai Deutel, Nam Nguyen, Adam D. Langley, Flip Korn, Lucia Rossazza, Alexandre Ram'e, Sagar M. Waghmare, Helen Miller, Vaishakh Keshava, Ying Jian, Xiaofan Zhang, Raluca Ada Popa, Kedar Dhamdhere, Blavz Bratanivc, Kyuyeun Kim, Terry Koo, Ferran Alet, Yi ting Chen, Arsha Nagrani, Hannah Muckenhirn, Zhiyuan Zhang, Corbin Quick, Filip Paveti'c, Duc Dung Nguyen, João Carreira, Michael Elabd, Ha roon Qureshi, Fabian Mentzer, Yao Yang, Danielle Eisenbud, Anmol Gulati, Ellie Talius, Eric Ni, Sahra Ghalebikesabi, Edouard Yvinec, Alaa Saade, Thatcher Ulrich, Lorenzo Blanco, Dan Andrei Calian, Muhua Huang, Aäron van den Oord, Naman Goyal, Terry Chen, Praynaa Rawlani, Christian Schallhart, S. D. Lokhande, Xianghong Luo, Jyn Shan, Ceslee Montgomery, Victoria Krakovna, Federico Piccinini, Omer Barak, Jingyu Cui, Yiling Jia, Mikhail Dektiarev, Alexey Kolganov, Shiyu Huang, Zhe Chen, Xingyu Wang, Jessica Austin, Peter de Boursac, Evgeny Sluzhaev, Feng Ding, Huijian Li, Surya Bhupatiraju, Mohit Kumar Agarwal, Slawek Kwasiborski, Paramjit Sandhu, Patrick Siegler, Ahmet Iscen, Eyal Ben-David, Shiraz Butt, Miltos Allamanis, Seth Benjamin, Róbert Istvan Busa-Fekete, Félix Hernández-Campos, Sasha Goldshtein, Matt Dibb, Weiyan Zhang, Annie Marsden, Carey Radebaugh, Stephen Roller, Abhishek Nayyar, Jacob Austin, Tayfun Terzi, Bhargav Kanagal Shamanna, Peter Shaw, Aayush Singh, Florian Luisier, Artur Mendoncca, Vaibhav Aggarwal, Larisa Markeeva, Claudio Fantacci, Sergey Brin, HyunJeong Choe, Guanyu Wang, Hartwig Adam, Avigail Dabush, T. Kiyono, E. A. Pepijn Marcus, Jeremy R. Cole, Théophane Weber, Hongrae Lee, Ronny Huang, Alex Muzio, Leandro Kieliger, Maigo Le, Courtney Biles, Long Le, Archit Sharma, Chengrun Yang, Avery Lamp, Dave Dopson, Nathan C. Hurley, Katrina Xu, Zhihao Shan, Shuang Song, Jiewen Tan, Alexandre Senges, George Fei Zhang, Chong You, Yennie Jun, David Raposo, Susanna Ricco, Xuan Yang, Weijie Chen, Prakhar Gupta, Arthur Szlam, Kevin Villela, Chun-Sung Ferng, Daniel Kasenberg, Chen Liang, Rui Zhu, Arunachalam Narayanaswamy, Florence Perot, Paul Pucciarelli, Akshay Shekhawat, Alex Stern, Rishikesh Ingale, Stefani Karp, Sanaz Bahargam, Adrian Goedeckemeyer, Jie Han, Sicheng Li, Andrea Tacchetti, Dian Yu, Abhishek Chakladar, Zhiying Zhang, Mona Mahdy, Xu Gao, Dale S. Johnson, Samrat Phatale, A. J. Piergiovanni, Hyeontaek Lim, Clément Farabet, Carl S. Lebsack, Theo Guidroz, John Blitzer, Nico Duduta, David Madras, Steve Li, Daniel von Dincklage, Xin Yue Amanda Li, Mahdis Mahdieh, George Tucker, Ganesh Jawahar, Yunxuan Owen Xiao, Daniel Tarlow, Robert Geirhos, Noam Velan, Daniel Vlasic, Kalesha Bullard, SK Park, Nishesh Gupta, Kellie Webster, Ayal Hitron, Jieming Mao, Julian Martin Eisenschlos, Laurel Prince, Nina D'souza, Kelvin Yu-Chen Zheng, Sara Nasso, Gabriela Botea, Carl Doersch, Caglar Unlu, Chris Alberti, Alexey Svyatkovskiy, Ankit Goel, Krzysztof Choromanski, Pan-Pan Jiang, Richard Nguyen, Four Flynn, Daria Ćurko, Peter Chen, Nicholas Roth, Kieran Milan, Caleb Habtegebriel, Shashi Narayan, M. Moffitt, Jake Marcus, Thomas Anthony, Brendan McMahan, Gowoon Cheon, Ruibo Liu, Megan Barnes, Lukasz Lew, Re beca Santamaria-Fernandez, Mayank Upadhyay, Arjun Akula, Arnar Mar Hrafnkelsson, A. Deniz Caceres, Andrew Bunner, Michal Sokolik, Subha Puttagunta, Lawrence Carlton Moore, Berivan Isik, Jay Hartford, Lawrence Chan, Pradeep Shenoy, Daniel Niels Holtmann-Rice, Jane Park, Fabio Viola, Alexandru Salcianu, Sujeevan Rajayogam, Ian Stewart-Binks, Zelin Wu, Richard Everett, Xi Xiong, Pierre-Antoine Manzagol, Gary Leung, Carl Saroufim, Bo Pang, Dawid Wegner, George Papamakarios, Jennimaria Palomaki, Helena Pankov, Guangda Lai, Guilherme Hideo Tubone, Shubin Zhao, Theofilos Strinopoulos, Seth Neel, Mingqiu Wang, Joe Kelley, Li Li, Ping mei Xu, Anitha Vijayakumar, Andrea D'olimpio, Omer Levy, Massimo Nicosia, Grigory Rozhdestvenskiy, Ni Lao, Sirui Xie, Yash Katariya, Jon Simon, Sanjiv Kumar, Florian Hartmann, Michael Kilgore, Jinhyuk Lee, Aroma Mahendru, Roman Ring, Tom Hennigan, Fiona Lang, Colin Cherry, David Steiner, Dawsen Hwang, Ray Smith, Pidong Wang, Jeremy Chen, Ming Yang, S. Kimberly Kwei, Philippe Schlattner, Donnie Kim, Ganesh Poomal Girirajan, Nikola Momchev, Ayushi Agarwal, Xingyi Zhou, Ilkin Safarli, Zachary Garrett, AJ Pierigiovanni, Sarthak Jauhari, Alif Raditya Rochman, Shikhar Vashishth, Quan Yuan, Christof Angermueller, Jon Blanton, Xiny ing Song, Nitesh Bharadwaj Gundavarapu, Thi Avrahami, Maxine Deines, Subhrajit Roy, Manish Gupta, Christopher Semturs, Shobha Vasudevan, Aditya Srikanth Veerubhotla, Shriya Sharma, Joshy Jacob, Zhen Yang, Andreas Terzis, Dan Karliner, Auriel Wright, Tania Rojas-Esponda, Ashley Brown, Abhijit Guha Roy, Pawan Dogra, Andrei Kapishnikov, Peter Young, Wendy Suk Fun Kan, Vinodh K. Rajendran, Maria Ivanova, Salil Deshmukh, Chia-Hua Ho, Michael Kwong, S. R. Ginzburg, Annie Louis, KP Sawhney, Slav Petrov, Jing Xie, Yun-

fei Bai, Georgi P. Stoyanov, Alex Fabrikant, Rajesh Jayaram, Yuqi Li, Joseph Heyward, Justin Gilmer, Yaqing Wang, Radu Soricut, Lu Liu, Qing ping Duan, Jamie Hayes, Maura O'Brien, Gau rav Singh Tomar, Sivan Eiger, Bahare Fatemi, Jeffrey Hui, Catarina Barros, A. Chukwuka, Alena Butryna, Saksham Thakur, Austin Huang, Zhufeng Pan, Haotian Tang, Serkan Cabi, Tulsee Doshi, Michiel A. Bakker, Sumit Bagri, Ruy Ley-Wild, Ádám Dániel Lelkes, Jennie Lees, Patrick Kane, David Greene, Shimu Wu, Jorg Bornschein, Gabriela de Castro Surita, Sarah Hodkinson, Fangtao Li, Chris Hidey, Sébastien Pereira, Sean Ammirati, Phillip Lippe, Adam Kraft, Pu Han, Sebastian Gerlach, Zifeng Wang, Liviu Panait, Feng Han, Brianne Schurott Farris, Yingying Bi, Hannah DeBalsi, Miaosen Wang, Gladys Tyen, James Cohan, Susan Zhang, Jarred Barber, Da-Woon Chung, Jaeyoung Kim, Markus Kunesch, Steven Pecht, Nami Akazawa, Abe Friesen, James Lyon, Ali Eslami, Junru Wu, Jiewen Tan, Yue Song, Ravin Kumar, Christoper A. Welty, Ilia Akolzin, Gena Gibson, Sean Augenstein, Arjun Pillai, Nancy Wang Yuen, Du Phan, Xin Wang, Iain Barr, Heiga Zen, Nan Hua, Casper Liu, Jilei Wang, Tanuj Pal Bhatia, Hao Xu, Oded Elyada, Pushmeet Kohli, Mirek Olvs'ak, Kelly Chen, Azalia Mirhoseini, Noam Shazeer, Shoshana Jakobovits, Maggie Tran, Nolan Ramsden, T Bharti, Fred Alcober, Yunjie Li, Shilpa Shetty, Jing Chen, Dmitry Kalashnikov, Megha Nawhal, Sercan Ö. Arik, Hanwen Chen, Michiel Blokzijl, Shubham Gupta, Janet Rubin, Rigel Swavely, Sophie Bridgers, Ian M. Gemp, Chenlin Su, Arun Sai Suggala, Juliette Pluto, Mary Cassin, Alain C. Vaucher, Kaiyang Ji, Jiahao Cai, Andrew Audibert, Animesh Sinha, David Tian, Efrat Farkash, Amy Hua, Jilin Chen, Duc Tran, Edward Loper, Nicole Brichtova, Lara McConnaughey, Ballie Sandhu, Robert Leland, Douglas DeCarlo, Andrew Over, James Wei chieh Huang, Xing Wu, Changshuo Fan, Eric Li, Yun-Peng Lei, Deepak Sharma, Cosmin Paduraru, Luo Yu, Matko Bovsnjak, Phuong Dao, Min Choi, Sneha Kudugunta, Jakub Adamek, Carlos Gu'ia, Ali Khodaei, Jie Feng, Wenjun Zeng, David Welling, Sandeep Tata, Christina Butterfield, Andrey Sergeevich Vlasov, Seliem El-Sayed, Swaroop Mishra, Tara N. Sainath, Shentao Yang, R. J. Skerry-Ryan, Jeremy Shar, Robert Berry, Abirami Rajendran, Arun Reddy Kandoor, Andrea Burns, Deepali Jain, Tom Stone, Wonpyo Park, Shibo Wang, Albin Cassirer, Guohui Wang, Hayato Kobayashi, Sergey Rogulenko, Vineetha Govindaraj, Mikolaj Rybi'nski, Nadav Olmert, Colin Evans, Po-Sen Huang, Kelvin Xu, Premal Shah, Terry Thurk, Caitlin Sikora, Mu Cai, Jinyu Xie, Elahe Dabir, Saloni Shah, Norbert Kalb, Carrie Zhang, Shruthi Prabhakara, Amit Sabne, Artiom Myaskovsky, Vikas Raunak, Blanca Huergo, Behnam Neyshabur, Jon Clark, Ye Zhang, Shankar Krishnan, Eden Cohen, Dinesh Tewari, James Lottes, Yumeya Yamamori, Hui Li, Mohamed Elhawaty, Ada Maksutaj Oflazer, Adrià Recasens, Sheryl Luo, Duy Nguyen, Taylor Bos, Kalyan Andra, Ana S. Salazar, Ed Huai hsin Chi, Jeongwoo Ko, Matthew L. Ginsberg, Anders Andreassen, Anian Ruoss, Todor Davchev, Elnaz Davoodi, Chenxi Liu, Min Seok Kim, Santiago Ontanon, Chi Ming To, Dawei Jia, Rosemary Ke, Jing Wang, Anna Korsun, Moran Ambar, Ilya Kornakov, Irene Giannoumis, Toni Creswell, Denny Zhou, Yi Su, Ishaan Watts, Aleksandr Zaks, Evgenii Eltyshev, Ziqiang Feng, Sidharth Mudgal, Alex Kaskasoli, J Christopher Love, Kingshuk Dasgupta, Sam Shleifer, Richard John Green, Sungyong Seo, Chansoo Lee, Dale R. Webster, Prakash Shroff, Ganna Raboshchuk, Isabel Leal, James Manyika, Sofia Erell, Daniel Murphy, Zhisheng Xiao, Anton Bulyenov, Julian Walker, Mark Collier, Matej Kastelic, Nelson George, Sushant Prakash, Sailesh Sidhwani, Alexey I. Frolov, Steven Hansen, Petko Georgiev, Tiberiu Sosea, Chris Apps, Aishwarya B Kamath, David Reid, Emma Cooney, Charlotte Magister, Oriana Riva, Alec Go, Pu-Chin Chen, Se bastian Krause, Nir Levine, Marco Fornoni, Ilya Figotin, Nick Roy, Parsa Mahmoudieh, Vladimir Magay, Mukundan Madhavan, Jin Miao, Jianmo Ni, Yasuhisa Fujii, I-Fan Chou, George Scrivener, Zak Tsai, Siobhan Mcloughlin, Jeremy Selier, Sandra Lefdal, Jeffrey Zhao, Abhijit Karmarkar, Kushal Chauhan, Shivank Goel, Zhaoyi Zhang, Vihan Jain, Parisa Haghani, Mostafa Dehghani, Jacob Scott, Erin Farnese, Anastasija Ili'c, Steven Baker, Julia Pawar, Li Zhong, Josh Camp, Yoel Zeldes, Shravya Shetty, Anand Iyer, V'it List'ik, Jiaxian Guo, Luming Tang, Mark Geller, Simon Bucher, Yifan Ding, Hongzhi Shi, Carrie Muir, Dominik Grewe, Ramy Eskander, Octavio Ponce, Boqing Gong, Derek Gasaway, Samir Khan, Umang Gupta, Angelos Filos, Weicheng Kuo, Klemen Kloboves, Jennifer Beattie, Christiana Wright, Le Li, Alicia Jin, Sandeep Mariserla, Miteyan Patel, Jens Heitkaemper, Dilip Krishnan, Vivek Kumar Sharma, David Bieber, Christian Frank, John Lambert, Paul Caron, Martin Polacek, Mai Gim'enez, Himadri Choudhury, Xing Yu, Sasan Tavakkol, Arun Ahuja, Franz Josef Och, Rodolphe Jenatton, Wojtek Skut, Bryan Richter, David Gaddy, Andy Ly, Misha Bilenko, Megh Umekar, Ethan Liang, Martin Sevenich, Mandar Joshi, Hassan Mansoor, Rebecca Lin, Sumit K. Sanghai, Abhimanyu Singh, Xiaowei Li, Sudheendra Vijayanarasimhan, Zaheer Abbas, Yonatan Bitton, Hansa Srinivasan, Manish Reddy Vuyyuru, Alexander Frommgen, Yanhua Sun, Ralph Leith, Alfonso Castaño, DJ Strouse, Le Yan, Austin Kyker, Satish Kambala, Mary Jasarevic, Thibault Sellam, Chao Jia, Alexander Pritzel, R Raghavender, Huizhong Chen, Natalie Clay, Sudeep Gandhe, Sean Kirmani, Sayna Ebrahimi, Hannah Kirkwood, Jonathan Mallinson, Chao Wang, Adnan Ozturel, Kuo-Chin Lin, Shyam Upadhyay, Vincent Cohen-Addad, Sean Purser-Haskell, Yichong Xu, Ebrahim M. Songhori, Babi Seal, Al berto Magni, Almog Gueta, Tingting Zou, Guru Guruganesh, Thais Kagohara, Hung Nguyen, Khalid Salama, Alejandro Ruiz, Justin Frye, Zhenkai Zhu, M. Lochbrunner, Simon Osindero, Wentao Yuan, Lisa Lee, Aman Prasad, Lam Nguyen Thiet, Daniele Calandriello, Victor Paromov Stone, Qixuang Feng, Han Ke, Maria Voitovich, Geta Sampemane, Lewis Chiang, Ling da Wu, Alexander Bykovsky, Matt Young, Luke Vilnis, Ishita Dasgupta, Aditya Chawla, Qin Cao, Bowen Liang, Daniel

Toyama, Szabolcs Payrits, Anca Stefanoiu, Dimitrios Vytiniotis, Ankesh Anand, Tianxiao Shen, Blagoj Mitrevski, Michael Tschannen, Sreenivas Gollapudi, S AishwaryaP, José Paulo Leal, Zhe Shen, Han Fu, Wei Wang, Arvind Kannan, Doron Kukliansky, Sergey Yaroshenko, Svetlana Grant, Umesh Telang, David Wood, Alexandra Chronopoulou, Alexandru cTifrea, Tianxiang Zhou, Tony Nguyen, Muge Ersoy, Anima Singh, Meiyan Xie, Emanuel Taropa, Woohyun Han, Eirikur Agustsson, Andrei Sozanschi, Hui Peng, Alex Chen, Yoel Drori, Efren Robles, Yang Gao, Xerxes Dotiwalla, Ying Chen, Anudhyan Boral, Alexei Bendebury, John Nham, Chris Tar, Luis Castro, Jiepu Jiang, Canoee Liu, Felix Halim, Jinoo Baek, Andy T. S. Wan, Jeremiah Liu, Yuan Cao, Shengyang Dai, Tri Dev Acharya, Ruoxi Sun, Fuzhao Xue, Saket Joshi, Morgane Lustman, Yongqin Xian, Rishabh Joshi, Deep Karkhanis, Nora Kassner, Jamie Hall, Xiangzhuo Ding, Gan Song, Gang Li, Chen Zhu, Yana Kulizhskaya, Bin Ni, Alexey Vlaskin, Solomon Demmessie, Lucio Dery, Salah Zaiem, Yanping Huang, Cindy Fan, Felix Gimeno, Ananth Balashankar, Kazuhiro Kojima, Hagai Taitelbaum, M. Meng, Dero Gharibian, Sahil Singla, Wei Chen, Ambrose Slone, Guanjie Chen, Sujeevan Rajayogam, Max Schumacher, Suyog Kotecha, Rory Blevins, Qifei Wang, Mor Hazan Taege, Alex Morris, Xin Liu, Fayaz Jamil, Richard Zhang, Pratik Joshi, Ben Ingram, Tyler Liechty, Ahmed Eleryan, Scott P. Baird, Alex Grills, Gagan Bansal, Shan Han, Kiran Yalasangi, Shawn Xu, Majd Al Merey, Isabel Gao, Felix Weissenberger, Igor Karpov, Robert Riachi, Ankit Anand, Gautam Prasad, Kay Lamerigts, Reid Hayes, J. Daniel Rogers, Mandy Guo, Ashish Shenoy, Qiong Hu, Kyle He, Yuchen Liu, Polina Zablotskaia, Sagar Gubbi, Yifan Chang, Jay Pavagadhi, Kristian Kjems, Archita Vadali, Diego Machado, Yeqing Li, Ren shen Wang, Dipankar Ghosh, Akash Mehta, Dana Alon, George Polovets, Alessio Tonioni, Nate Kushman, Joel D'sa, Lin Zhuo, Allen Wu, Rohin Shah, John Youssef, Jiayu Ye, Justin Snyder, Karel Lenc, Senaka Buthpitiya, Matthew Tung, Jichuan Chang, Tao Chen, David Saxton, Jenny Lee, Lydia Zhang, James Qin, Prabakar Radhakrishnan, Maxwell Chen, Piotr Ambroszczyk, Metin Toksoz-Exley, Yan Zhong, Nitzan Katz, Brendan O'Donoghue, Tamara von Glehn, Adi Gerzi Rosenthal, Agnieszka Swietlik, Xiaokai Zhao, Nicholas Fernando, Jinliang Wei, Jieru Mei, Sergei Vassilvitskii, Diego Cedillo, Pranjal Awasthi, Hui-Wen Zheng, Koray Kavukcuoglu, Itay Laish, Joseph Pagadora, Marc Brockschmidt, Christopher A. Choquette-Choo, Arun Byravan, Yifeng Lu, Xu Chen, Mianna Chen, Kenton Lee, Rama Pasumarthi, Sijal Bhatnagar, Aditya Shah, Qiyin Wu, Zhuoyuan Chen, Zachary Nado, Bartek Perz, Zixuan Jiang, David L. Kao, Ganesh Satish Mallya, Nino Vieillard, Lantao Mei, Sertan Girgin, Meredith Jordan, Yeongil Ko, Alekh Agarwal, Yaxin Liu, Yasemin Altun, Raoul de Liedekerke, Anastasios Kementsietsidis, Daiyi Peng, Dangyi Liu, Utku Evci, Peter C. Humphreys, Austin Tarango, Xiang Deng, Yoad Lewenberg, Kevin Aydin, Chengda Wu, Bhavishya Mittal, Tsendsuren Munkhdalai, Kleopatra Chatziprimou, Rodrigo Benenson, Uri First, Xiao Ma, Jinning Li, Armand Joulin, Hamish Tomlinson, Tingnan Zhang, Milad Nasr, Zhi Hong, Michael E. Sander, Lisa Anne Hendricks, Anuj Sharma, Andrew Bolt, Eszter V'ertes, Jiří ima, Tomer Levinboim, Olcan Sercinoglu, Divyanshu Shukla, Austin Wu, Craig Swanson, Danny Vainstein, Fan Bu, Boyu Wang, Ryan C. Julian, Charles Yoon, Sergei Lebedev, Antonious M. Girgis, Bernd Bandemer, David Du, Todd Wang, Xi Chen, Ying Xiao, Peggy Pei-Ying Lu, Natalie Ha, Vlad Ionescu, Simon Rowe, Josip Matak, Federico Lebron, Andreas Steiner, Lalit Jain, Manaal Faruqui, Nicolas Lacasse, George M. Evans, Neesha Subramaniam, D. Taylor Reich, Giulia Vezzani, Aishwarya Pandey, Joe Stanton, Tianhao Zhou, Liam McCafferty, Henry Griffiths, Verena Rieser, Soheil Hassas Yeganeh, Eleftheria Briakou, Lu Huang, Zichuan Wei, Liangchen Luo, Erik Jue, Gabby Wang, Victor Cotruta, Myriam Khan, Jongbin Park, Qi Guo, Peiran Li, Rong Rong, Diego Antognini, Anastasia Petrushkina, Chetan Tekur, Eli Collins, Parul Bhatia, Chester Kwak, Wenhu Chen, Arvind Neelakantan, Immanuel Odisho, Sheng Peng, Vincent Nallatamby, Vaibhav Tulsyan, Fabian Pedregosa, Peng Xu, Raymond Lin, Yulong Wang, Emma Wang, Sholto Douglas, Reut Tsarfaty, Elena Gribovskaya, Renga Aravamudhan, Manu Agarwal, Mara Finkelstein, Qiao Zhang, Elizabeth Cole, Phil Crone, Sarmishta Velury, Anil Das, Chris Sauer, Luyao Xu, Danfeng Qin, Chenjie Gu, Dror Marcus, CJ Zheng, Wouter Van Gansbeke, Sobhan Miryoosefi, Haitian Sun, Yaguang Li, Charlie Chen, Jaewook Yoo, Pavel Dubov, Alex Tomala, Adams Yu, Paweł Wesołowski, Alok Gunjan, Eddie Cao, Jiaming Luo, Nikhil Sethi, Arkadiusz Socala, Laura Graesser, Tomás Kociský, BC Arturo, Minmin Chen, Edward Lee, Sophie Wang, Weize Kong, Qian Xu, Nilesh Tripuraneni, Yiming Li, Xinxin Yu, Allen C. Porter, Paul Voigtlaender, Biao Zhang, Arpi Vezer, Sarah York, Qinglan Wei, Geoffrey Cideron, Mark Kurzeja, Seungyeon Kim, Benny Li, Angéline Pouget, Hyo Lee, Kaspar Daugaard, Yang Li, David Uthus, Aditya Siddhant, Paul Cavallaro, Sriram Ganapathy, Maulik Shah, Rolf Jagerman, Jeff Stanway, Piermaria Mendolicchio, Li Xiao, Kayi Lee, Tara Thompson, Shubham Milind Phal, Jason A. Chase, Sun Jae Lee, Adrian N. Reyes, Disha Shrivastava, Zhen Qin, Roykrong Sukkerd, Seth Odoom, Lior Madmoni, John Aslanides, Jonathan Herzig, Elena Pochernina, Sheng Zhang, Parker Barnes, Daisuke Ikeda, Qiujia Li, Shuo-Yiin Chang, Shakir Mohamed, Jim Sproch, Richard Powell, Bidisha Samanta, Domagoj Cevid, Anton Kovsharov, Shrestha Basu Mallick, Srinivas Tadepalli, Anne Xiangyi Zheng, Kareem W. Ayoub, Andreas Noever, Christian Reiswig, Zhuo Xu, Junhyuk Oh, Martin Matysiak, Tim Blyth, Shereen Ashraf, Julien Amelot, Boone Severson, Michele Bevilacqua, Motoki Sano, Ethan Dyer, Ofir Roval, Anu Sinha, Yi Zhong, Sagi Perel, Tea Saboli'c, Johannes Mauerer, Willi Gierke, Mauro Verzetti, Rodrigo Cabrera, Alvin Abdagic, Steven Hemingray, Austin Stone, Jong Lee, Farooq Ahmad, Karthik Raman, Lior Shani, Jonathan Lai, Orhan Firat, Nathan Waters, Eric Ge, Mo Shomrat, Himanshu Gupta, Rajeev Kumar Aggarwal, Tom Hudson, Bill Jia, Simon Baumgartner, Palak Jain, Joseph A.

Kovac, Junehyuk Jung, Ante vZuvzul, W Truong, Morteza Zadimoghaddam, Songyou Peng, M. Liang, Rachel Sterneck, Balaji Lakshminarayanan, Machel Reid, Oliver Woodman, Tong Zhou, Jianling Wang, Vincent Coriou, Arjun Narayanan, Jay A. Hoover, Yenai Ma, Apoorv Jindal, Clayton Sanford, Doug Reid, Swaroop Indra Ramaswamy, Alexey Kurakin, Roland S. Zimmermann, Yana Lunts, Dragos Dena, Zalán Borsos, Vered Cohen, Shujian Zhang, Will Grathwohl, Robert Dadashi, Morgan Redshaw, Joshua Kessinger, J. J. Odell, Silvano Bonacina, Zihang Dai, Grace Chen, Ayush Dubey, Pablo Sprechmann, Mantas Pajarskas, Wenxuan Zhou, Niharika Ahuja, Tara Elizabeth Thomas, Martin Nikoltchev, Matija Kecman, Bharath Mankalale, Andrey Ryabtsev, Jennifer She, Christian J. Walder, Jiaming Shen, Lu Li, Carolina Parada, Sheena Panthaplackel, Okwan Kwon, Matthew Lawlor, Utsav Prabhu, Yannick Schroecker, Marc'Aurelio Ranzato, Pete Blois, Iurii Kemaev, Ting Yu, Dmitry Lepikhin, Hao Xiong, Sahand Sharifzadeh, Oleaser Johnson, Jeremiah Willcock, Rui Yao, Gregory Farquhar, Sujoy Basu, Hidetoshi Shimokawa, Nina Anderson, Haiguang Li, Khiem Pham, Yizhong Liang, Sebastian Borgeaud, Alexandre Moufarek, Hideto Kazawa, Blair Kutzman, Marcin Sieniek, Sara Smoot, Ruth Wang, Natalie Axelsson, Nova Fallen, P. Kalyana Sundaram, Yuexiang Zhai, Varun Godbole, Petros Maniatis, Alek Wenjiao Wang, Ilia Shumailov, Santhosh Thangaraj, Remi Crocker, Nikita Gupta, Gang Wu, Phil Chen, G. Weisz, Céline Smith, Mojtaba Seyedhosseini, Bo Fang, Xiyang Luo, Roey Yogev, Zeynep Cankara, Andrew Straiton Hard, Helen Ran, Rahul Sukthankar, George Necula, Gael Liu, Honglong Cai, Praseem Banzal, Daniel Keysers, Sanjay Ghemawat, Connie Tao, Emma Dunleavy, Aditi Chaudhary, Wei Li, Maciej Mikuła, Chen-Yu Lee, Tiziana Refice, Krishna Somandepalli, Alexandre Fr'echette, Daniel Bahir, John E. Karro, Keith Rush, Sarah Perrin, Bill Rosgen, Xiaomeng Yang, Clara Huiyi Hu, Mah moud Alnahlawi, Justin Mao-Jones, Roopal Garg, Hoang-Phi Nguyen, Bat-Orgil Batsaikhan, Iñaki Iturrate, Anselm Levskaya, Avi Singh, Ashyana Kachra, Tony Lu, Denis Petek, Zhen Xu, Mark Graham, Lukás Zilka, Yael Karov, Marija Kostelac, Fangyu Liu, Yaohui Guo, Weiyue Wang, Bernd Bohnet, Emily Pitler, Tony Bruguier, Keisuke Kinoshita, Chrysovalantis Anastasiou, Nilpa Jha, Ting Liu, Jerome T. Connor, Phillip Wallis, Philip Pham, Eric Bailey, Shixin Li, Heng tze Cheng, Sally Ma, Haiqiong Li, Akanksha Maurya, Kate Olszewska, Manfred K. Warmuth, Christy Koh, Dominik Paulus, Siddhartha R. Jonnalagadda, Enrique Piqueras, Ali Elqursh, Geoff Brown, Hadar Shemtov, Lorenzo Maggiore, Fei Xia, Ryan Foley, Beka Westberg, George van den Driessche, Livio Baldini Soares, Arjun Kar, Michael Quinn, Siqi Zuo, Jialin Wu, Kyle Kastner, Anna Bortsova, Aijun Bai, Ales Mikhalap, Lu owei Zhou, Jenny Brennan, Vinay Venkatesh Ramasesh, Honglei Zhuang, John Maggs, Johan Schalkwyk, Yuntao Xu, Hui Huang, Andrew Howard, Sasha Brown, Lint-859 ing Xue, Gloria Shen, Brian Albert, N.K. Jha, Daniel Zheng, Varvara Krayvanova, Spurthi Amba Hombaiah, Olivier Lacombe, Gautam Vasudevan, Dan Graur, Tian Xie, Meet Gandhi, Bangju Wang, Dustin Zelle, Harman Singh, Dahun Kim, S'ebastien Cevey, Victor Ungureanu, Natasha Noy, Fei Liu, Annie Xie, Fangxi aoyu Feng, Katerina Tsihlas, Daniel Formoso, Neera Vats, Quentin Wellens, Yinan Wang, Niket Kumar Bhumihar, Samrat Ghosh, Matt Hoffman, Tom Lieber, Oran Lang, Kush Bhatia, Tom Le Paine, Aroonalok Pyne, Ronny Votel, Madeleine Elish, Benoît Schillings, Alexios Panagopoulos, Haichuan Yang, Adam Raveret, Zohar Yahav, Shuang Liu, Dalia El Badawy, Nishant Agrawal, Mohammed Badawi, Mahdi Mirzazadeh, C. Bromberg, Fan Ye, Chang Liu, Tatiana Sholokhova, George cristian Muraru, Gargi Balasubramaniam, Jonathan Malmaud, Alen Carin, Danilo Martins, Irina Jurenka, Pankil Botadra, Dave Lacey, Richa Singh, Mariano Schain, Daniel Zheng, Isabelle Guyon, Victor Lavrenko, Seungjin Lee, Xiang Zhou, Demis Hassabis, Jeshwanth Challagundla, Derek Zhiyuan Cheng, Nikhil Mehta, M Mauger, Michela Paganini, Pushkar Mishra, Katherine Lee, Zhang Li, Lexi Baugher, Ondrej Skopek, Max Chang, Amir Zait, Gaurav Menghani, Lizzetth Bellot, Guangxing Han, Jean Michel A. Sarr, Sharat Chikkerur, Himanshu Sahni, Rohan Anil, Arun Narayanan, Chandu Thekkath, Daniele Pighin, Hana Strejvcek, Marko Velic, Fred Bertsch, Manuel Tragut, Keran Rong, Alicia Parrish, Kai Bailey, Jiho Park, Isabela Albuquerque, Abhishek Bapna, Rajesh Venkataraman, Alec Kosik, Johannes Griesser, Zhiwei Deng, Alek Andreev, Qingyun Dou, Kevin Hui, Fanny Wei, Xiaobing Yu, Lei Shu, Avia Aharon, David Barker, Badih Ghazi, Sebastian Flennerhag, Chris Breaux, Yuchuan Liu, Matthew Bilotti, J Luke Woodward, Uri Alon, Stephanie Winkler, Tzu-Kuo Huang, Kostas Andriopoulos, João Gabriel Oliveira, Penporn Koanantakool, Berkin Akin, Michael B. Wunder, Cicero Nogueira dos Santos, Mohammad Hossein Bateni, Ling Yang, Dan Horgan, Beer Changpinyo, Keyvan Amiri, Min Ma, Dayeong Lee, Lihao Liang, Anirudh Baddepudi, Tejasi Latkar, Raia Hadsell, Jun Xu, Hairong Mu, Michael Han, Aedan Pope, Snchit Grover, Frank Kim, Ankit V. Bhagatwala, Guan Sun, Yamini Bansal, Amir Globerson, Ali Nazari, Samira Daruki, Hagen Soltau, Jane Labanowski, Laurent El Shafey, Matt Harvey, Yanif Ahmad, Elan Rosenfeld, William Kong, Etienne Pot, Yi-Xuan Tan, A. Wei, Victoria Langston, M. Dany Prasetya, Petar Velivckovi'c, Richard Killam, Robin Strudel, Darren Ni, Zhen-Xing Zhu, Aaron Archer, Kavya Kopparapu, Lynn Nguyen, Emilio Parisotto, Hussain Masoom, Sravanti Addepalli, Jordan Grimstad, Hexiang Hu, Joss Moore, Avinatan Hassidim, Le Hou, Mukund Raghavachari, Jared Lichtarge, Adam R. Brown, Hilal Dib, Natalia Ponomareva, Justin Fu, Yujing Zhang, Altafur Rahman, Joana Iljazi, Edouard Leurent, Gabriel Dulac-Arnold, Cosmo Du, Chulayuth Asawaroengchai, Larry Jin, Ela Gruzewska, Ziwei Ji, Benigno Uria, Daniel De Freitas, Paul Barham, Lauren Beltrone, Victor F. Campos, Jun Yan, Neel Kovelamudi, Arthur Nguyen, Elinor Davies, Zhi Wu, Zoltan Egyed, Kristina Toutanova, Nithya Attaluri, Hongliang Fei, Peter Stys, Siddhartha Brahma, Martin Izzard, S. Velusamy, Scott M. Lundberg, Vin-

cent Zhuang, Kevin Sequeira, Adam Santoro, Ehsan Amid, Ophir Aharoni, Shuai Ye, Mukund Sundararajan, Lijun Yu, Yu-Cheng Ling, Stephen Spencer, Hugo Song, Josip Djolonga, Christo Kirov, Sonal Gupta, Alessandro Bissacco, Clemens Meyer, Mukul Bhutani, Andrew M. Dai, Weiyi Wang, Siqi Liu, Ashwin Sreevatsa, Qijun Tan, Maria Wang, Lucy Kim, Yicheng Wang, Alex Irpan, Yang Xiao, Stanislav Fort, Yifan He, A. H. Gurney, Bryan Gale, Yue Ma, Monica Roy, Viorica Patraucean, Taylan Bilal, Golnaz Ghiasi, Anahita Hosseini, Melvin Johnson, Zhuowan Li, Yi Lun Tay, Benjamin Beyret, Katie Millican, Josef Broder, Mayank Lunayach, Danny Swisher, Eugen Vuvsak, David Parkinson, Mh Tessler, Adi Mayrav Gilady, Richard Song, Allan Dafoe, Yves Raimond, Masataka Yamaguchi, Itay Karo, Elizabeth Nielsen, Kevin Kilgour, Mike Dusenberry, Rajiv Mathews, Jiho Choi, Siyuan Qiao, Harsh Mehta, Sahitya Potluri, Chris Knutsen, Jialu Liu, Thay Hui Tan, Kuntal Sengupta, Keerthana Gopalakrishnan, Abodunrinwa Toki, Mencher Chiang, Mike Burrows, Grace Vesom, Zafarali Ahmed, Ilia Labzovsky, Siddharth Vashishtha, Preeti Singh, Ankur Sharma, Ada Ma, Jinyu Xie, Pranav Talluri, Hannah Forbes-Pollard, Aarush Selvan, Joel Wee, Loïc Matthey, Tom Funkhouser, Parthasarathy Gopavarapu, Lev Proleev, Cheng Li, Matt Thomas, Kashyap Babu Rao Kolipaka, Zhipeng Jia, Ashwin Kakarla, Srinivas Sunkara, Joan Puigcerver, Suraj Satishkumar Sheth, Emily A. Graves, Chen Wang, Sadhia Khan, Kai Kang, S. Buch, Fred Zhang, Omkar Savant, David Soergel, Kevin Lee, Linda Friso, Xuanyi Dong, Rahul Arya, Shreyas Chandrakaladharan, Connor Schenck, Greg Billock, Tejas Iyer, Anton Bakalov, Leslie W. Baker, Alex D. Ruiz, Angad Chandorkar, Trieu H. Trinh, Matt Miecnikowski, Yanqi Zhou, Yangsibo Huang, Jiazhong Nie, Ali Shah, Ashish V Thapliyal, Sam Haves, Lun Wang, Uri Shaham, Patrick Morris-Suzuki, Soroush Radpour, Leonard Berrada, Thomas Strohmann, Chaochao Yan, Jingwei Shen, Sonam Goenka, Tris Warkentin, Petar Devi'c, Daniel Belov, Albert Webson, Madhavi Yenugula, Puranjay Datta, Jerry Chang, Nimesh Ghelani, Aviral Kumar, Vincent Perot, Jessica Lo, Yang Song, Herman Schmit, Jianmin Chen, Vasilisa Bashlovkina, Xiaoyue Pan, Diana Mincu, Paul Roit, Isabel Edkins, Andy Davis, Yujia Li, Ben Horn, Xinjian Li, S. Pradeepkumar, Eric Doi, Wanzheng Zhu, Sriram Padmanabhan, Siddharth Verma, Jasmine Liu, Heng Chen, Mihajlo Velimirovi'c, Malcolm Reynolds, Priyanka Agrawal, N. M. Sukhanov, Abhinit Modi, Siddharth Goyal, John Palowitch, Nima Khajehnouri, Wing W. Lowe, David Klinghoffer, S Silver, Vinh Q. Tran, Candice Schumann, Francesco Tommaso Piccinno, Xi Liu, Mario Luvci'c, Xiaochen Yang, Sandeep Kumar, Ajay Kannan, Raghavendra Kotikalapudi, Mudit Bansal, Fabian B Fuchs, Mohammad Javad Hosseini, Abdelrahman Essam Abdelhamed, Dawn Bloxwich, Tianhe Yu, Ruoxin Sang, Gregory Thornton, Karan Gill, Yuchi Liu, Virat Shejwalkar, Jason Lin, Zhipeng Yan, Kehang Han, Thomas Buschmann, Michael Pliskin, Zhiqiang Xing, Susheel Tatineni, Junlin Zhang, Sissie Hsiao, Gavin Buttimore, Marcus Wu, Zefei Li, Geza Kovacs, Legg Yeung, Tao Huang, Aaron Cohen, Bethanie Brownfield, Averi Nowak, Mikel Rodriguez, Tianze Shi, H. V. Hasselt, Kevin Cen, Deepanway Ghosal, Kushal Majmundar, Weiren Yu, Warren Weilun Chen, Danila Sinopalnikov, Hao Zhang, Vlatko Galić, Di Lu, Zeyu Zheng, Mi-Hwa Song, Gary Wang, Gui Citovsky, Swapnil Gawde, Isaac R. Galatzer-Levy, David Silver, Ivana Balazevic, Dipanjan Das, Kingshuk Majumder, Yale Cong, Praneet Dutta, Dustin Tran, Hui Wan, Junwei Yuan, Daniel Eppens, Alanna Walton, Been Kim, Harry Ragan, James Cobon-Kerr, Lu Liu, Weijun Wang, Bryce Petrini, Jack W. Rae, Rakesh Shivanna, Yan Xiong, Chace Lee, Pauline Coquinot, Yiming Gu, Lisa N. Patel, Blake A. Hechtman, Aviel Boag, Orion Jankowski, Alex Wertheim, Alex X. Lee, Paul Covington, Hila Noga, Sam Sobell, Sai Vasanth, William Bono, Chirag Nagpal, Wei Fan, Xavier García, Kedar Soparkar, Aybuke Turker, Nathan Howard, Sachit Menon, Yuankai Chen, Vikas Verma, Vladimir Pchelin, Harish Rajamani, Valentin Dalibard, Ana Ramalho, Yang Guo, Kartikeya Badola, Seojin Bang, Nathalie Rauschmayr, Julia Proskurnia, Sudeep Dasari, Xinyun Chen, Mikhail Sushkov, Anja Hauth, Pauline Sho, Abhinavkumar Singh, Bilva Chandra, Allie Culp, Max Dylla, Olivier Bachem, James Besley, He Zhao, Timothy P. Lillicrap, Wei Wei, Wael Al Jishi, Ning Niu, Alban Rrustemi, Raphael Lopez Kaufman, Ryan Poplin, Jewel Zhao, Minh Truong, Shikhar Bharadwaj, Ester Hlavnova, Eli Stickgold, Cordelia Schmid, Georgi Stephanov, Zhaoqi Leng, Frederick Liu, L'eonard Hussenot, Shenil Dodhia, Juliana Franco, Lesley Katzen, Abhanshu Sharma, Sarah Cogan, Zuguang Yang, Aniket Ray, Sergi Caelles, Shen Yan, Ravin Kumar, Daniel Gillick, Renee Wong, Joshua Ainslie, Jonathan Hoech, Sébastien M. R. Arnold, Daniel A. Abolafia, Anca Dragan, Ben Hora, Grace Hu, Alexey Guseynov, Yang Lu, Chas Leichner, Jinmeng Rao, Abhimanyu Goyal, Nagabhushan Baddi, Daniel Hernandez Diaz, Timothy McConnell, Max Bain, Jake Abernethy, Qiqi Yan, Rylan Schaeffer, Paul Vicol, Will Thompson, Montse Gonzalez Arenas, Mathias M.J. Bellaiche, Pablo Barrio, Stefan Zinke, Riccardo Patana, Pulkit Mehta, J. Kearns, Avraham Ruderman, Scott Pollom, David B. D'Ambrosio, Catherine Anne Hope, Yang Yu, Andrea Gesmundo, Kuang-Huei Lee, Aviv Rosenberg, Yiqian Zhou, Yaoyiran Li, Drew Garmon, Yonghui Wu, Safeen Huda, Gil Fidel, Martin Baeuml, Jian Li, Phoebe Kirk, Rhys May, Tao Tu, Sara Marie Mc Carthy, Toshiyuki Fukuzawa, Miranda Aperghis, Chih-Kuan Yeh, Tetsuma Yoshino, Bo Li, Austin Myers, Kaisheng Yao, Ben Limonchik, Changwan Ryu, Rohun Saxena, Alex Goldin, Ruizhe Zhao, Rocky Rhodes, Tao Zhu, Divya Tyam, Heidi Howard, Nathan Byrd, Hongxu Ma, Yan Wu, Ryan Mullins, Qingze Wang, Aida Amini, Sebastien Baur, Yiran Mao, Subhashini Venugopalan, Will Song, Wen Ding, Paul Collins, Sashank J. Reddi, Megan Shum, Andrei A. Rusu, Luisa M. Zintgraf, Kelvin C. K. Chan, Sheela Goenka, Mathieu Blondel, Michael Collins, Renke Pan, Marissa Giustina, Nikolai Chinaev, Christian Schuler, Ce Zheng, Jonas Valfridsson, Alyssa Marie Loo, Alexey Yakubovich, Jamie Smith,

Tao Jiang, Rich Munoz, Gabriel Barcik, Rishabh Bansal, Ming Yang, Yilun Du, Pablo Andrés Maya Duque, Mary Phuong, Alexandra M. Belias, Kunal Lad, Zeyu Leo Liu, Tal Schuster, Karthik Duddu, Jieru Hu, P. Kunkle, Matthew Watson, Jackson Tolins, Joshua R. Smith, Denis Teplyashin, Garrett Bingham, Marvin Ritter, Marco Andreetto, Divya Pitta, Mohak Patel, S. Viswanadha, Trevor Strohman, Catalin Ionescu, Jincheng Luo, Yogesh Kalley, Jeremy Wiesner, Daniel Deutsch, Derek Lockhart, Peter Choy, Rumen Dangovski, Chawin Sitawarin, Cat Graves, Tanya Lando, Joost R. van Amersfoort, Ndidi Elue, Zhouyuan Huo, Pooya Moradi, Jean Tarbouriech, Henryk Michalewski, Wenting Ye, Eunyoung Kim, Alex Druinsky, Florent Altch'e, Xinyi Chen, Artur Dwornik, Da-Cheng Juan, Rivka Moroshko, Horia Toma, Jarrod Kahn, Hai Qian, Maximilian Sieb, Irene Cai, Roman Goldenberg, Praneeth Netrapalli, Sindhu Raghuram, Yuan Gong, Lijie Fan, Evan Palmer, Yossi Matias, Valentin Gabeur, Shreya Pathak, Tom Ouyang, Donald Metzler, Geoff Bacon, Srinivasan Venkatachary, Sridhar Thiagarajan, Alex A. Cullum, Eran. O. Ofek, Vytenis akėnas, Mohamed Hammad, Cesar Ilharco Magalhaes, Mayank Daswani, Os car Chang, Ashok Popat, Ruichao Li, Komal Jalan, Yanhan Hou, Josh Lipschultz, Antoine He, Wenhao Jia, Pier Giuseppe Sessa, Prateek Kolhar, William Wong, Sumeet Singh, Lukas Haas, Jay Whang, Hanna Klimczak-Pluci'nska, Georges Rotival, Grace Yuna Chung, Yiqing Hua, Anfal Siddiqui, Nicolás Serrano, Dongkai Chen, Billy Porter, Libin Bai, Keshav Shivam, Sho Arora, Partha Talukdar, Tom Cobley, Sangnie Bhardwaj, Evgeny Gladchenko, Simon Peter Green, Kelvin Guu, Felix Fischer, Xiao Yue Wu, Eric Wang, Achintya Singhal, Tatiana Matejovicova, James Martens, Hongji Li, Roma Patel, Elizabeth A. Kemp, Jiaqi Pan, Lily Wang, Blake JianHang Chen, Jean-Baptiste Alayrac, Navneet Potti, Erika Gemzer, Eugene Ie, Kay McKinney, Takaaki Saeki, Edward Chou, Pascal Lamblin, SQ Mah, Zachary Fisher, Martin Chadwick, Jon Stritar, Obaid Sarvana, Andrew Hogue, A. S. Shtefan, Hadi Hashemi, Yang Xu, Jindong Gu, Sharad Vikram, Chung-Ching Chang, Sabela Ramos, Logan Kilpatrick, Wei Xi, Jenny Brennan, Yinghao Sun, Ab hishek Jindal, Ionel Gog, Dawn Chen, Felix Wu, Jason Lee, Sudhindra Kopalle, Srinadh Bhojanapalli, Oriol Vinyals, Natan Potikha, Burcu Karagol Ayan, Yuan Yuan, Michael Riley, Piotr Stańczyk, Sergey Kishchenko, Bing Wang, Dan Garrette, Antoine Yang, Vladimir Feinberg, Cj Carey, Javad Azizi, Viral R. Shah, Erica Moreira, Chong-Rong Shi, Joshua Feldman, Elizabeth Salesky, Thomas Lampe, Aneesh Pappu, Duhyeong Kim, Jonas Adler, Avi Caciularu, Brian J. Walker, Yunhan Xu, Yochai Blau, Dylan Scandinaro, Terry Huang, Sam El-Husseini, Abhishek Sinha, Lijie Ren, Taylor Tobin, Patrik Sundberg, Tae Sung Sohn, Vikas Yadav, Mimi Ly, Emily Xue, Jing Xiong, Afzal Shama Soudagar, Sneha Mondal, Nikhil Khadke, Qing Chun Ren, Ben Vargas, Stanley M. Bileschi, Sarah Chakera, Cindy Wang, Boyu Wang, Yoni Halpern, Joe Wenjie Jiang, Vikas Sindhwani, Petr Petrov, Pranavaraj Ponnuramu, Sanket Vaibhav Mehta, Yuichi Watanabe, Betty Chan, Matheus Augusto Wisniewski, Trang Pham, Jingwei Zhang, Conglong Li, Dario de Cesare, Art Khurshudov, Alex Vasiloff, Melissa Tan, Zoe Ashwood, Bobak Shahriari, Maryam Majzoubi, Garrett Tanzer, Olga Kozlova, Robin Alazard, James Lee-Thorp, Nguyet Minh Phu, Isaac Y. Tian, Junwhan Ahn, Andy Crawford, Lauren Lax, Shangguan Yuan, Iftekhar Naim, David A. Ross, Oleksandr Ferludin, Tongfei Guo, Andrea Banino, Hubert Soyer, Xiaoen Ju, Dominika Rogozi'nska, Ishaan Malhi, Marcella Valentine, Daniel Balle, Apoorv Kulshreshtha, Maciej Kula, Yiwen Song, Sophia Austin, John Schultz, Roy Hirsch, Arthur Douillard, Apoorv Reddy, Michael Fink, Summer Yue, Khyatti Gupta, Ada Zhang, Norman Alexander Rink, Daniel McDuff, Lei Meng, Andr'as Gyorgy, Yasaman Razeghi, Ricky Liang, Kazuki Osawa, Aviel Atias, Matan Eyal, Tyrone Hill, Nikolai Grigorev, Zhengdong Wang, Nitish Kulkarni, Rachel Soh, Ivan Lobov, Zachary Charles, Sid Lall, Kazuma Hashimoto, Ido Kessler, Victor do Nascimento Gomes, Zelda E. Mariet, Danny Driess, Alessandro Agostini, Canfer Akbulut, Jing Hu, Marissa Ikonomidis, Emily Caveness, Kartik Audhkhasi, Saurabh Agrawal, Ioana Bica, Evan Senter, Jayaram Mudigonda, Kelly Chen, Jingchen Ye, Xuanhui Wang, James Svensson, Philipp Franken, Joshua Newlan, Li Lao, Eva Schnider, Sami Alabed, Joseph Kready, Jesse Emond, Afief Halumi, Tim Zaman, Chengxi Ye, Naina Raisinghani, Vilobh Meshram, Bo-Cian Chang, Ankit Singh Rawat, Axel Stjerngren, Sergey Levi, Rui Wang, Xiang Long, Mitchelle Rasquinha, Steven Hand, Aditi Mavalankar, Lauren Agubuzu, Sudeshna Roy, Junquan Chen, Jarek Wilkiewicz, Hao Zhou, Michał K. Jastrzębski, Agustin Dal Lago, Ramya Sree Boppana, Wei-Jen Ko, Jennifer Prendki, Yao Su, Zhi Li, Eliza Rutherford, Girish Ramchandra Rao, Ramona Comanescu, Adrià Puigdomènech, Qihang Chen, Dessie Petrova, Christine Chan, Vedrana Milutinovic, Felipe Tiengo Ferreira, Chin-Yi Cheng, Ming Zhang, T. K. Dey, Sherry Yang, Ramesh Sampath, Quoc V. Le, Howard Zhou, Chu cheng Lin, Hoi Lam, Christine Kaeser-Chen, Kai Hui, Dean Hirsch, Tom Eccles, Basil Mustafa, Shruti Rijhwani, Morgane Rivière, Yuanzhong Xu, Junjie Wang, Xinyang Geng, Xi ance Si, Arjun Khare, Cheolmin Kim, Vahab S. Mirrokni, Ka-Wei Lee, Khuslen Baatarsukh, Nathan Braun, Lisa Wang, LV Pallavi, Richard Tanburn, Yuvein Zhu, Fang fang Li, Setareh Ariafar, Daniel W. Goldberg, Ken Burke, Daniil Mirylenka, Meiqi Guo, Olaf Ronneberger, Hadas Vogel, Liqun Cheng, Nishita Shetty, Johnson Jia, Thomas Jimma, Corey Fry, Ted Xiao, Martin Sundermeyer, Ryan Burnell, Yannis Assael, Mario Pinto, JD Chen, Rohit Sathyanarayana, Donghyun Cho, Jin Lu, Rishabh Agarwal, Sugato Basu, Lucas Gonzalez, Dhruv Shah, Meng Wei, Dre Mahaarachchi, Rohan Agrawal, Tero Rissa, Yani Donchev, Ramiro Leal-Cavazos, Adrian Hutter, M.-T. Mircea, Alon Jacovi, Faruk Ahmed, Jiageng Zhang, Shuguang Hu, Bo-Juen Chen, Jonni Kanerva, Guillaume Desjardins, Andrew Lee, Nikos Parotsidis, Asier Mujika, Tobias Weyand, Jasper Snoek, J. M. Chick, Kai Chen, Paul Chang, Ethan Mahintorabi, Zi Wang, Tolly Powell, Orgad Keller,

Abhirut Gupta, Claire Sha, Kanav Garg, Nico-351 las Heess, 'Agoston Weisz, Cassidy Hardin, Bartek P. Wydrowski, Benjamin Coleman, Karina Zainullina, Pankaj Joshi, Alessandro Epasto, Terry Spitz, Binbin Xiong, Kai Zhao, Arseniy Klimovskiy, Ivy Zheng, Johan Ferret, Itay Yona, Waleed Khawaja, Jean-Baptiste Lespiau, Maxim Krikun, Siamak Shakeri, Timothée Cour, Bonnie Li, Igor Krivokon, Daniel Suh, Alex Hofer, Jad Al Abdallah, Nikita Putikhin, Oscar Akerlund, Silvio Lattanzi, Anurag Kumar, Sean O. Settle, Himanshu Srivastava, Folawiyo Campbell-Ajala, Edouard Rosseel, Mihai Istin, Nishanth Dikkala, Anand Rao, Nick Young, Kate Lin, Dhruva Bhaswar, Yiming Wang, Jaume Sanchez Elias, K. Muralidharan, James Keeling, Dayou Du, Siddharth Gopal, Gregory Dibb, Charles Blundell, Manolis Delakis, Jacky Liang, Marco Túlio Ribeiro, Georgi E. Karadzhov, Guillermo Garrido, Ankur Bapna, Jiawei Cao, Adam Sadovsky, Pouya Dehghani Tafti, Arthur Guez, Coline Devin, Yixian Di, Jinwei Xing, Chuqiao Xu, Hanzhao Lin, Chun-Te Chu, Sameera S. Ponda, Wesley Helmholz, Fan Yang, Yue Gao, Sara Javanmardi, Wael Farhan, Alex Ramírez, Ricardo Figueira, Khe Chai Sim, Yuval Bahat, Ashwin Vaswani, Liangzhe Yuan, Gufeng Zhang, Leland Rechis, Hanjun Dai, Tayo Oguntebi, Alexandra Cordell, Eug'enie Rives, Kaan Tekelioglu, Naveen Kumar, Bing Zhang, Aurick Zhou, Nikolay Savinov, Andrew Leach, Alex Tudor, Sanjay Ganapathy, Yanyan Zheng, Mirko Rossini, Vera Axelrod, Arnaud Autef, Yukun Zhu, Zheng Zheng, Mingda Zhang, Baochen Sun, Jie Ren, Nenad Tomaev, Nithish Kannen, Amer Sinha, Charlie Chen, Louis O'Bryan, Alex Pak, Aditya Kusupati, Wee Jun Yang., Deepak Ramachandran, Patrick Griffin, Seokhwan Kim, Philipp Neubeck, Craig Schiff, Tammo Spalink, Mingyang Ling, Arun Nair, Ga-Young Joung, Linda Deng, Avishkar Bhoopchand, Lora Aroyo, Tom Duerig, Jordan Griffith, Gabriel Barth-Maron, Jake Ades, Alex Haig, Ankur Taly, Yunting Song, Paul Michel, Dave Orr, Dean Weesner, Corentin Tallec, Carrie Grimes Bostock, Paul Niemczyk, Andrew Twigg, Mudit Verma, Rohith Vallu, Henry Wang, Marco Gelmi, Kiranbir Sodhia, Aleksandr Chuklin, Omer Goldman, Jasmine George, Liang Bai, Ke Zhang, Petar Sirkovic, Efrat Nehoran, Golan Pundak, Jiaqi Mu, Alice Xuelian Chen, Alex C. Greve, Paulo Zacchello, David Amos, Heming Ge, Eric Noland, Colton Bishop, Jeffrey M. Dudek, Youhei Namiki, Elena Buchatskaya, Jing Li, Dorsa Sadigh, Masha Samsikova, Dan Malkin, Damien Vincent, Robert David, Rob Willoughby, Phoenix Meadowlark, Shawn Gao, Yan Li, Rajas Apte, Amit Jhindal, S. Lin, Alex Polozov, Zhicheng Wang, Tomas Mery, G. Anirudh, Varun Yerram, S. Ashley Stevens, Tianqi Liu, Noah Fiedel, Charles Sutton, Matthew Johnson, Xiaodan Song, Kate Baumli, Nir Shabat, Muqthar Mohammad, Hao Liu, Marco Selvi, Yichao Zhou, Mehdi Hafezi Manshadi, Chu-Ling Ko, Anthony Chen, Michael Bendersky, Jorge Gonzalez Mendez, Nisarg Ashish Kothari, Amir Zandieh, Yiling Huang, Daniel Andor, Ellie Pavlick, Idan Brusilovsky, Jitendra K. Harlalka, Sally Goldman, Andrew K. Lampinen, Guowang Li, Asahi Ushio, Somit Gupta, Lei Zhang, Chuyuan Kelly Fu, Madhavi Sewak, Timo Denk, Jed Borovik, Brendan Jou, Avital Zipori, Prateek Jain, Junwen Bai, Thang Luong, Jonathan Tompson, Alice Li, Li Liu, George Powell, Jiajun Shen, Alex Feng, Grishma Chole, Dahai Yu, Yinlam Chow, Tongxin Yin, Eric Malmi, Kefan Xiao, Yash Pande, Shachi Paul, Niccolò Dal Santo, Adil Dostmohamed, Sergio Guadarrama, Aaron B. Phillips, Thanumalayan Sankaranarayana Pillai, G. Yona, Amin Ghafouri, Preethi Lahoti, Benjamin Lee, Dhruv Madeka, Eren Sezener, Simon Tokumine, Adrian Collister, Nicola De Cao, Richard Shin, Uday Kalra, Parker Beak, Emily Nottage, Ryo Nakashima, Ivan Jurin, Vikash Sehwag, Meenu Gaba, Junhao Zeng, Kevin R. McKee, Fernando Pereira, Tamar Yakar, Amayika Panda, Arka Dhar, Peilin Zhong, Daniel Sohn, Mark Brand, Lars Lowe Sjoesund, Viral Carpenter, Sharon Lin, Shantanu Thakoor, Marcus Wainwright, Ashwin Chaugule, Pranesh Srinivasan, Muye Zhu, Bernett Orlando, Jack Weber, Ayzaan Wahid, Gilles Baechler, Apurv Suman, Jovana Mitrovi'c, Gabe Taubman, Honglin Yu, Helen King, Josh Dillon, Cathy Yip, D. Surya Chandra Varma, Tomás Izo, Levent Bolelli, Borja de Balle Pigem, Julia Di Trapani, Fotis Iliopoulos, Adam Paszke, Nishant Ranka, Joe Zou, Francesco Pongetti, Jed N McGiffin, Alexander Siegman, Rich Galt, Ross Hemsley, Goran vZuvzi'c, Victor Carbune, Tao Li, Myle Ott, Félix de Chaumont Quitry, David Vilar Torres, Yuri Chervonyi, Tomy Tsai, Prem Eruvbetine, Samuel J. Yang, Matthew Denton, Jacob Walker, Slavica Andavci'c, Idan Heimlich Shtacher, Vittal Premachandran, Harshal Tushar Lehri, Cip Baetu, Damion Yates, Lampros Lamprou, Mariko Iinuma, Ioana Mihailescu, Ben Albrecht, Shachi Dave, S. Sargsyan, Bryan Perozzi, Lucas Manning, Chiyuan Zhang, Denis Vnukov, Igor Mordatch, Raia Hadsell Wolfgang Macherey, Ryan D. Kappedal, Jim Stephan, Aditya Tripathi, Klaus Macherey, Jincheng Qian, Abhishek Bhowmick, Shekoofeh Azizi, Rémi Leblond, Shivani Garlapati, Timothy Knight, Matt Wiethoff, Wei-Chih Hung, Anelia Angelova, Georgios Evangelopoulos, Paweł Janus, Dimitris Paparas, Matthew Rahtz, Ken Caluwaerts, Vivek Sampathkumar, Daniel Jarrett, Shadi Noghabi, Antoine Miech, Chak Ha Yeung, Geoff Clark, Henry Prior, Fei Zheng, Jean Pouget-Abadie, Indro Bhattacharya, Kalpesh Krishna, William Bishop, Zhe Yuan, Yunxiao Deng, Ashutosh Sathe, Kacper Krasowiak, Ciprian Chelba, Cho-Jui Hsieh, Kiran Vodrahalli, Buhua Liu, Thomas Koppe, Amr Khalifa, Lubo Litchev, Pichi Charoenpanit, Ruben Roberts, Sachin Yadav, Yasumasa Onoe, Desislav Ivanov, Megha Mohabey, Vighnesh Birodkar, Nemanja Raki'cevi'c, Pierre Sermanet, Vaibhav Mehta, Krishan Subudhi, Travis Choma, Willa Ng, Luheng He, Kathie Wang, Tasos Kementsietsidis, Shane Gu, Mansi Gupta, Andrew Nystrom, Mehran Kazemi, Timothy Chung, N.F. Cano, Nikhil Dhawan, Yufei Wang, Jiawei Xia, Trevor Yacovone, Eric Jia, Mingqing Chen, Simeon Emilov Ivanov, Ashrith Sheshan, Sid Dalmia, Paweł Stradomski, Pengcheng Yin, Salem Haykal, Congchao Wang, Dennis Duan, Neslihan Bulut, Greg Kochanski, Liam MacDermed, Nam-

rata Godbole, Shitao Weng, Jingjing Chen, Rachana Fellinger, Ramin Mehran, Dan Suo, Hisham Husain, Tong He, Kaushal Patel, Joshua Howland, Randall E. Parker, Kelvin Nguyen, Sharath Maddineni, Christopher Rawles, Mina Khan, Shlomi Cohen-Ganor, Amol Mandhane, Xinyi Wu, Chenkai Kuang, Iulia Comcsa, R. Ganeshan, Hanie Sedghi, Adam E. Bloniarz, Nuo Wang Pierse, Anton Briukhov, Petr Mitrichev, Anita Gergely, Serena Zhan, Allan Zhou, Nikita Saxena, Eva Lu, Josef Dean, Ashish Gupta, Nicolas Perez-Nieves, Renjie Wu, Cory Y. McLean, Wei Liang, Disha Jindal, Anton Tsitsulin, Wenhao Yu, Kaiz Alarakyia, Tom Schaul, Piyush A. Patil, Peter Sung, Eli Peake, Hongkun Yu, Feryal M. P. Behbahani, JD Co-Reyes, Alan Ansell, Sean Sun, Clara Barbu, Jonathan Lee, Seb Noury, James Urquhart Allingham, Bilal Piot, Mohit Sharma, Christo pher Yew, Ivan Korotkov, Bibo Xu, Deborah Brady, Goran Stanajic Petrovic, Shibl Mourad, Claire Cui, Aditya Gupta, Parker Schuh, Saarthak Khanna, Anna Goldie, Abhinav Arora, Vadim Zubov, Amy Lynette Stuart, Mark Epstein, Yun Zhu, Jianqiao Liu, Yury Stuken, Ziyue Wang, Karolis Misiunas, Dee Guo, Akshay Gill, Ale Jakse Hartman, Zaid Nabulsi, Au rko Roy, Aleksandra Faust, Jason Riesa, Ben Withbroe, Mengchao Wang, Marco Tagliasacchi, Andreea Marzoca, James Noraky, Sergey Toropov, Malika Mehrotra, Bahram Raad, Sanja Deur, Steve Xu, Marianne Monteiro, Zhongru Wu, Yi Luan, Sam Ritter, Nick Li, Haavard Garnes, Yanzhang He, Martin Zlocha, Jifan Zhu, Matteo Hessel, Will Bernard Wu, Spandana Raj Babbula, Chizuko Kawamoto, Yuanzhen Li, Mehadi Hassen, Yan Wang, Brian Wieder, Jim Freedman, Yin Zhang, Xinyi Bai, Tianli Yu, David Reitter, XiangHai Sheng, Mateo Wirth, Aditya Kini, Dima Damen, Mingcen Gao, Rachel Hornung, Michael Voznesensky, Brian Roark, Adhi Kuncoro, Yuxiang Zhou, Rushin Shah, Anthony Brohan, Kuan-Chia Chen, James G Wendt, David Rim, Paul Kishan Rubenstein, Jonathan J. Halcrow, Michelle Liu, Ty Geri, Yunhsuan Sung, J. Shapiro, Shaan Bijwadia, Chris Duvarney, Christina Sorokin, Paul Natsev, R. Reeve Ingle, Pramod Gupta, Youngill Maeng, Ndaba Ndebele, Kexin Zhu, Valentin Anklin, Katherine Lee, Yuan Liu, Yaroslav Akulov, Shaleen Gupta, Guolong Su, Flavien Prost, Tianlin Liu, Vitaly Kovalev, Pol Moreno, Matthias Scholz, Sam Redmond, Zongwei Zhou, Alex Castro-Ros, André Susano Pinto, Dia Kharrat, Michal Yarom, Rachel Saputro, Jannis Bulian, Ben Caine, Ji Liu, Abbas Abdolmaleki, Shariq Iqbal, Tautvydas Misiunas, Mikhail Sirotenko, Shefali Garg, Guy Bensky, Huan Gui, Xuezhi Wang, Raphael Koster, Mike Bernico, Da Huang, Romal Thoppilan, Trevor Cohn, Ben Golan, Wenlei Zhou, Andrew Rosenberg, Markus Freitag, Tynan Gangwani, Vincent Tsang, Anand Shukla, Xiaoqi Ren, Minh Giang, China Zou, Andre Elisseeff, Charline Le Lan, Dheeru Dua, Shuba Lall, Pranav Shyam, Frankie Garcia, Sarah Nguyen, Michael Guzmán, AJ Maschinot, Marcello Maggioni, Ming-Wei Chang, Karol Gregor, Lotte Weerts, K.G.S. Venkatesan, Bogdan Damoc, Leo Liu, Jan Wassenberg, Lewis Ho, Becca Roelofs, Majid Hadian, Franccois-Xavier Aubet, Yu Liang, Sami Lachgar, Danny Karmon, Yong Cheng, Amelio V'azquez-Reina, Angie Chen, Zhuyun Dai, Andy Brock, Shubham Agrawal, Chenxi Pang, Peter Garst, Mariella Sanchez-Vargas, Ivor Rendulic, Aditya Ayyar, Andrija Ravznatovi'c, Olivia Ma, Roopali Vij, Navjyoti Sharma, Ashwin Balakrishna, Bing Liu, Ian Mackinnon, Sorin Baltateanu, Petra Poklukar, Gabriel Ibagon, Colin Ji, Hongyan Jiao, Isaac Noble, Wojciech Stokowiec, Zhihao Li, Jeffrey Dean, David Lindner, Mark Omernick, Kristen Chiafullo, Mathilde Dimarco, Vitor Hugo Serravalle Reis Rodrigues, Vittorio Selo, Garrett Honke, Xintian Wu, Wei He, A. Hillier, Anhad Mohananey, Vihari Piratla, Chang Ye, Chase Malik, Sebastian Riedel, Samuel Albanie, Zi Yang, Kenny K. Vassigh, Maria Bauzá, Sheng Li, Yiqing Tao, Nevan Wichers, Andrii Maksai, Abe Ittycheriah, Ross Mcilroy, Bryan Seybold, Noah Goodman, Romina Datta, Steven Hernandez, Tianze Shi, Yony Kochinski, Anna Bulanova, Ken Franko, Mikita Sazanovich, Nicholas Fitzgerald, Praneeth Kacham, Shubha Raghvendra, Vincent J. Hellendoorn, Alexander Grushetsky, Julian Salazar, Angeliki Lazaridou, Jason Chang, Jan-Thorsten Peter, Sushant Kafle, Yann Dauphin, Abhishek Rao, Filippo Graziano, Izhak Shafran, Yuguo Liao, Tianli Ding, Geng Yan, Grace Chu, Zheming Fu, Vincent Roulet, Gabriel Rasskin, Duncan Williams, Shahar Drath, Alexander Mossin, Raphael Hoffmann, Jordi Orbay, Francesco Bertolini, Hila Sheftel, Justin Chiu, Siyan Xue, Yuheng Kuang, Ferjad Naeem, Swaroop Nath, Nana Barimah Nti, Phil Culliton, Kashyap Krishnakumar, Michael Isard, Pei Sun, Ayan Chakrabarti, Nathan Clement, Regev Cohen, Arissa Wongpanich, Gs Oh, Ashwin Murthy, Hao Zheng, Jessica B. Hamrick, Oskar Bunyan, Suhas Ganesh, Nitish Gupta, Roy Frostig, J. Michael Wieting, Yury Malkov, Pierre Marcenac, Zhixin Lai, Xiaodan Tang, Mohammad Saleh, Fedir Zubach, Chinmay Kulkarni, Huanjie Zhou, Vicky Zayats, Nan Ding, Anshuman Tripathi, Arijit Pramanik, Patrik Zochbauer, Harish Ganapathy, Vedant Misra, Zach Behrman, Hugo Vallet, Mingyang Zhang, Mukund Sridhar, Ye Jin, Samuel Gehman, Mohammad Babaeizadeh, Siim Põder, Megha Goel, Divya Jain, Tajwar Nasir, Shubham Mittal, Tim Dozat, Diego Ardila, Aliaksei Severyn, Fabio Pardo, Sammy Jerome, Siyang Qin, Louis Rouillard, Amir Yazdanbakhsh, Zizhao Zhang, Shivani Agrawal, Kaushik Shivakumar, Caden Lu, Praveen Kallakuri, Rachita Chhaparia, Kanishka Rao, Charles Kwong, Anastasiia Fadeeva, Shitij Nigam, Yan Virin, Yuan Zhang, Balaji Venkatraman, Beliz Gunel, Marc Wilson, Huiyu Wang, Abhinav Gupta, Abhinav Gupta, Adrien Ali Taiga, Kareem Mohamed, Doug Fritz, Daniel Rodríguez, Zoubin Ghahramani, Harry Askham, Lior Belenki, James Xu Zhao, Rahul Gupta, Krzysztof Jastrzkebski, Takahiro Kosakai, Kaan Katircioglu, Jon Schneider, Rina Panigrahy, Konstantinos Bousmalis, Peter Grabowski, Prajit Ramachandran, Chaitra Hegde, Mihaela Roșca, Angelo Scorza Scarpati, K. Axiotis, Ying Xu, Zach Gleicher, Assaf Hurwitz Michaely, Mandar Sharma, Sanil Jain, Christoph Hirnschall, Tal Marian, Xuhui Jia, K Mather, Kilol Gupta, Linhai Qiu, Nigamaa Nayakanti, Lucian Ionita, Steven Zheng, Lucia Lo-

her, Kurt Shuster, Igor Petrovski, Roshan Sharma, Rahma Chaabouni, Angel Yeh, James An, Arushi Gupta, Steven L. Schwarcz, Shannon Ellis, Sam Conway-Rahman, Javier Snaider, Alex Zhai, James Atwood, Daniel Golovin, Liqian Peng, Ivanova Te, Vivian Xia, Salvatore Scellato, Mahan Malihi, Arthur Bravzinskas, Vlăducă Ion, Younghoon Jun, James Swirhun, Soroosh Mariooryad, Jiao Sun, Steve A. Chien, Rey Coaguila, A. J. Brand, Yi Gao, Tom Kwiatkowski, Roee Aharoni, Cheng-Chun Lee, Mislav vZani'c, Yichi Zhang, Danielle M Ethier, Vitaly Nikolaev, Pranav Ajit Nair, Yoav Ben Shalom, Hen Fitoussi, Jai Prakash Gupta, Hongbin Liu, Dee Cattle, Tolga Bolukbasi, Ben Murdoch, Fantine Huot, Yin Li, Chris Hahn, Urvashi Khandelwal, Frederik Benzing, Arthur Conmy, Andrey E. Simanovsky, Franccoise Beaufays, Eugene Weinstein, Tongzhou Chen, Luke Leonhard, and Bhuvana Ramabhadran. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *ArXiv*, abs/2507.06261.

Antonios Dimakis, Stella Markantonatou, and Antonios Anastasopoulos. 2024. Dictionary-aided translation for handling multi-word expressions in low-resource languages. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2588–2595, Bangkok, Thailand. Association for Computational Linguistics.

Shreya Havaldar, Adam Stein, Eric Wong, and Lyle Ungar. 2025. Towards style alignment in cross-cultural translation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 32213–32230, Vienna, Austria. Association for Computational Linguistics.

Alexander H. Liu, Kartik Khandelwal, Sandeep Subramanian, Victor Jouault, Abhinav Rastogi, Adrien Sadé, Alan Jeffares, Albert Jiang, Alexandre Cahill, Alexandre Gavaudan, Alexandre Sablayrolles, Amélie Héliou, Amos You, Andy Ehrenberg, Andy Lo, Anton Eliseev, Antonia Calvi, Avinash Sooriyarachchi, Baptiste Bout, Baptiste Rozière, Baudouin De Monicault, Clémence Lanfranchi, Corentin Barreau, Cyprien Courtot, Daniele Grattarola, Darius Dabert, Diego de las Casas, Elliot Chane-Sane, Faruk Ahmed, Gabrielle Berrada, Gaëtan Ecrepont, Gauthier Guinet, Georgii Novikov, Guillaume Kunsch, Guillaume Lample, Guillaume Martin, Gunshi Gupta, Jan Ludziejewski, Jason Rute, Joachim Studnia, Jonas Amar, Joséphine Delas, Josselin Somerville Roberts, Karmesh Yadav, Khyathi Chandu, Kush Jain, Laurence Aitchison, Laurent Fainsin, Léonard Blier, Lingxiao Zhao, Louis Martin, Lucile Saulnier, Luyu Gao, Maarten Buyl, Margaret Jennings, Marie Pellat, Mark Prins, Mathieu Poirée, Mathilde Guillaumin, Matthieu Dinot, Matthieu Futeral, Maxime Darrin, Maximilian Augustin, Mia Chiquier, Michel Schimpf, Nathan Grinsztajn, Neha Gupta, Nikhil Raghuraman, Olivier Bousquet, Olivier Duchenne, Patricia Wang, Patrick von Platen, Paul Jacob, Paul Wambergue, Paula Kurylowicz, Pavankumar Reddy Muddireddy, Philomène Chagniot, Pierre Stock, Pravesh Agrawal, Quentin Torroba, Romain Sauvestre, Roman Soletskyi, Rupert Menneer, Sagar Vaze, Samuel Barry, Sanchit Gandhi, Siddhant Waghjale, Siddharth Gandhi, Soham Ghosh, Srijan Mishra, Sumukh Aithal, Szymon Antoniak, Teven Le Scao, Théo Cachet, Theo Simon Sorg, Thibaut Lavril, Thiziri Nait Saada, Thomas Chabal, Thomas Foubert, Thomas Robert, Thomas Wang, Tim Lawson, Tom Bewley, Tom Bewley, Tom Edwards, Umar Jamil, Umberto Tomasini, Valeriia Nemychnikova, Van Phung, Vincent Maladière, Virgile Richard, Wassim Bouaziz, Wen-Ding Li, William Marshall, Xinghui Li, Xinyu Yang, Yassine El Ouahidi, Yihan Wang, Yunhao Tang, and Zaccharie Ramzi. 2026. Ministral 3.

Shuai Liu and Jonathan May. 2025. Style transfer with multi-iteration preference optimization. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2663–2681, Albuquerque, New Mexico. Association for Computational Linguistics.

Stella Markantonatou, Panagiotis Minos, George Zakis, Vassiliki Moutzouri, and Maria Chantou. 2019. Idion: A database for modern greek multiword expressions. pages 130–134.

Sourabrata Mukherjee, Atul Kr. Ojha, Akanksha Bansal, Deepak Alok, John P. McCrae, and Ondřej Dušek. 2024. Multilingual text style transfer: Datasets & models for Indian languages. In *Proceedings of the 17th International Natural Language Generation Conference*, pages 494–522, Tokyo, Japan. Association for Computational Linguistics.

Sourabrata Mukherjee, Atul Kr. Ojha, John P. McCrae, and Ondřej Dušek. 2025. Evaluating text style transfer evaluation: Are there any reliable metrics? In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)*, pages 418–434, Albuquerque, USA. Association for Computational Linguistics.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation.

Amalie Brogaard Pauli, Isabelle Augenstein, and Ira Assent. 2025. Mind the style gap: Meta-evaluation of

style and attribute transfer metrics. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 21550–21564, Suzhou, China. Association for Computational Linguistics.

Carlos Ramisch, Aline Villavicencio, and Christian Boitet. 2010. Multiword expressions in the wild? the mwetoolkit comes in handy. In *Coling 2010: Demonstrations*, pages 57–60, Beijing, China. Coling 2010 Organizing Committee.

Ricardo Rei, José G. C. de Souza, Duarte Alves, Chrysoula Zerva, Ana C Farinha, Taisiya Glushkova, Alon Lavie, Luisa Coheur, and André F. T. Martins. 2022a. COMET-22: Unbabel-IST 2022 submission for the metrics shared task. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 578–585, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Ricardo Rei, Marcos Treviso, Nuno M. Guerreiro, Chrysoula Zerva, Ana C Farinha, Christine Maroti, José G. C. de Souza, Taisiya Glushkova, Duarte Alves, Luisa Coheur, Alon Lavie, and André F. T. Martins. 2022b. CometKiwi: IST-unbabel 2022 submission for the quality estimation shared task. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 634–645, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann A. Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for nlp. In *Conference on Intelligent Text Processing and Computational Linguistics*.

A Prompts used

A.1 Greek → English, MWE → MWE

'Below is a Greek sentence using one or multiple multi-word expressions (MWE). Translate it to English keeping the meaning and style as close as possible, while replacing the MWEs with equivalent English MWEs whenever possible. Reply with the sentence and nothing else:<sentence>'

A.2 Greek → English, MWE → Plain

'Below is a Greek sentence using one or multiple multi-word expressions (MWE). Translate it to English, keeping the intended meaning but removing the MWE. Reply with the sentence and nothing else:<sentence>'

A.3 Greek → English, Plain → MWE

'Below is a literal Greek sentence. Translate it to English, keeping the same meaning, but try to use one or multiple multi-word expressions (MWE). Reply with the sentence and nothing else:<sentence>'

A.4 Greek → English, Plain → Plain

'Below is a Greek sentence. Translate it to English keeping the meaning and style as close as possible. Reply with the sentence and nothing else:<sentence>'

A.5 English → Greek, MWE → MWE

'Below is an English sentence using one or multiple multi-word expressions (MWE). Translate it to Greek keeping the meaning and style as close as possible, while replacing the MWEs with equivalent Greek MWEs whenever possible. Reply with the sentence and nothing else:<sentence>'

A.6 English → Greek, MWE → Plain

'Below is an English sentence using one or multiple multi-word expressions (MWE). Translate it to Greek, keeping the intended meaning but removing the MWE. Reply with the sentence and nothing else:<sentence>'

A.7 English → Greek, Plain → MWE

'Below is a literal English sentence. Translate it to Greek, keeping the same meaning, but try to use one or multiple multi-word expressions (MWE). Reply with the sentence and nothing else:<sentence>'

A.8 English → Greek, Plain → Plain

'Below is an English sentence. Translate it to Greek keeping the meaning and style as close as possible. Reply with the sentence and nothing else:<sentence>'

B StyleCOMET training details

We randomly selected 400 rows from the dataset to create the test set, and used 90% of the rest (1292 sentences) for training and 10% (143 sentences) for validation.

The "high" score given to instances where the hypothesis and reference styles matched was 0.98. The "low" score was 0.8.

The hyperparameters used for training were the following:

- learning rate: 5.0e-6
- layerwise decay: 0.95
- dropout: 0.1
- pool: avg
- batch size: 16
- max epochs: 2
- monitor: val_pearson

We also experimented with freezing the embedding layers and fine-tuning just the head, but observed worse results.

Our model contains approximately 580 million parameters in total and was trained using around 0.7 GPU hours of computation.

C Full results

C.1 Quality evaluation

C.1.1 COMETKIWI

Task	COMETKIWI
[en->el] MWE Preservation	0.6239
[en->el] MWE Removal	0.6646
[en->el] MWE Addition	0.5962
[en->el] Plain Translation	0.7238
[el->en] MWE Preservation	0.6497
[el->en] MWE Removal	0.6415
[el->en] MWE Addition	0.6587
[el->en] Plain Translation	0.7333

Table 4: Quality estimation (COMETKIWI) scores across translation tasks on ground-truth data. Plain-to-plain translations tend to be rated higher than ones containing MWEs.

C.1.2 COMET

Task	COMET
[en->el] MWE Preservation (MWE reference)	0.7975
[en->el] MWE Preservation (Plain reference)	0.7756
[en->el] MWE Removal (MWE reference)	0.7870
[en->el] MWE Removal (Plain reference)	0.8269
[en->el] MWE Addition (MWE reference)	0.7769
[en->el] MWE Addition (Plain reference)	0.7839
[en->el] Plain Translation (MWE reference)	0.8057
[en->el] Plain Translation (Plain reference)	0.8568
[el->en] MWE Preservation (MWE reference)	0.7329
[el->en] MWE Preservation (Plain reference)	0.6876
[el->en] MWE Removal (MWE reference)	0.7219
[el->en] MWE Removal (Plain reference)	0.7431
[el->en] MWE Addition (MWE reference)	0.7041
[el->en] MWE Addition (Plain reference)	0.7024
[el->en] Plain Translation (MWE reference)	0.7427
[el->en] Plain Translation (Plain reference)	0.7823

Table 5: COMET scores for Gemini translation outputs evaluated against target and alternative style references, on the entire dataset

Task	COMET
[en->el] MWE Preservation (MWE reference)	0.7987
[en->el] MWE Preservation (Plain reference)	0.7830
[en->el] MWE Removal (MWE reference)	0.7926
[en->el] MWE Removal (Plain reference)	0.8394
[en->el] MWE Addition (MWE reference)	0.7935
[en->el] MWE Addition (Plain reference)	0.8177
[en->el] Plain Translation (MWE reference)	0.8073
[en->el] Plain Translation (Plain reference)	0.8611
[el->en] MWE Preservation (MWE reference)	0.7261
[el->en] MWE Preservation (Plain reference)	0.6784
[el->en] MWE Removal (MWE reference)	0.7315
[el->en] MWE Removal (Plain reference)	0.7353
[el->en] MWE Addition (MWE reference)	0.7230
[el->en] MWE Addition (Plain reference)	0.7194
[el->en] Plain Translation (MWE reference)	0.7449
[el->en] Plain Translation (Plain reference)	0.7849

Table 6: COMET scores for Claude translation outputs evaluated against target and alternative style references, on the entire dataset

Task	COMET
[en->el] MWE Preservation (MWE reference)	0.7872
[en->el] MWE Preservation (Plain reference)	0.7779
[en->el] MWE Removal (MWE reference)	0.7879
[en->el] MWE Removal (Plain reference)	0.8241
[en->el] MWE Addition (MWE reference)	0.7928
[en->el] MWE Addition (Plain reference)	0.8333
[en->el] Plain Translation (MWE reference)	0.7988
[en->el] Plain Translation (Plain reference)	0.8526
[el->en] MWE Preservation (MWE reference)	0.7158
[el->en] MWE Preservation (Plain reference)	0.6727
[el->en] MWE Removal (MWE reference)	0.7241
[el->en] MWE Removal (Plain reference)	0.7196
[el->en] MWE Addition (MWE reference)	0.7238
[el->en] MWE Addition (Plain reference)	0.7320
[el->en] Plain Translation (MWE reference)	0.7408
[el->en] Plain Translation (Plain reference)	0.7770

Table 7: COMET scores for Mistral translation outputs evaluated against target and alternative style references, on the entire dataset

Task	COMET
[en->el] MWE Preservation (MWE reference)	0.7901
[en->el] MWE Preservation (Plain reference)	0.7751
[en->el] MWE Removal (MWE reference)	0.7811
[en->el] MWE Removal (Plain reference)	0.8231
[en->el] MWE Addition (MWE reference)	0.7757
[en->el] MWE Addition (Plain reference)	0.7857
[en->el] Plain Translation (MWE reference)	0.8006
[en->el] Plain Translation (Plain reference)	0.8564
[el->en] MWE Preservation (MWE reference)	0.7307
[el->en] MWE Preservation (Plain reference)	0.6854
[el->en] MWE Removal (MWE reference)	0.7227
[el->en] MWE Removal (Plain reference)	0.7466
[el->en] MWE Addition (MWE reference)	0.7053
[el->en] MWE Addition (Plain reference)	0.7009
[el->en] Plain Translation (MWE reference)	0.7439
[el->en] Plain Translation (Plain reference)	0.7825

Table 8: COMET scores for Gemini translation outputs evaluated against target and alternative style references, on the test set

Task	COMET
[en->el] MWE Preservation (MWE reference)	0.7937
[en->el] MWE Preservation (Plain reference)	0.7808
[en->el] MWE Removal (MWE reference)	0.7882
[en->el] MWE Removal (Plain reference)	0.8389
[en->el] MWE Addition (MWE reference)	0.7928
[en->el] MWE Addition (Plain reference)	0.8224
[en->el] Plain Translation (MWE reference)	0.8049
[en->el] Plain Translation (Plain reference)	0.8634
[el->en] MWE Preservation (MWE reference)	0.7253
[el->en] MWE Preservation (Plain reference)	0.6753
[el->en] MWE Removal (MWE reference)	0.7250
[el->en] MWE Removal (Plain reference)	0.7318
[el->en] MWE Addition (MWE reference)	0.7217
[el->en] MWE Addition (Plain reference)	0.7174
[el->en] Plain Translation (MWE reference)	0.7436
[el->en] Plain Translation (Plain reference)	0.7825

Table 9: COMET scores for Claude translation outputs evaluated against target and alternative style references, on the test set

Task	COMET
[en->el] MWE Preservation (MWE reference)	0.7851
[en->el] MWE Preservation (Plain reference)	0.7784
[en->el] MWE Removal (MWE reference)	0.7831
[en->el] MWE Removal (Plain reference)	0.8256
[en->el] MWE Addition (MWE reference)	0.7883
[en->el] MWE Addition (Plain reference)	0.8338
[en->el] Plain Translation (MWE reference)	0.7912
[en->el] Plain Translation (Plain reference)	0.8499
[el->en] MWE Preservation (MWE reference)	0.7105
[el->en] MWE Preservation (Plain reference)	0.6663
[el->en] MWE Removal (MWE reference)	0.7156
[el->en] MWE Removal (Plain reference)	0.7092
[el->en] MWE Addition (MWE reference)	0.7225
[el->en] MWE Addition (Plain reference)	0.7274
[el->en] Plain Translation (MWE reference)	0.7378
[el->en] Plain Translation (Plain reference)	0.7748

Table 10: COMET scores for Mistral translation outputs evaluated against target and alternative style references, on the test set

C.1.3 StyleCOMET

Task	StyleCOMET
[en->el] MWE Preservation (MWE reference)	0.8221
[en->el] MWE Preservation (Plain reference)	0.8072
[en->el] MWE Removal (MWE reference)	0.8057
[en->el] MWE Removal (Plain reference)	0.8275
[en->el] MWE Addition (MWE reference)	0.8068
[en->el] MWE Addition (Plain reference)	0.8076
[en->el] Plain Translation (MWE reference)	0.8118
[en->el] Plain Translation (Plain reference)	0.8427
[el->en] MWE Preservation (MWE reference)	0.8199
[el->en] MWE Preservation (Plain reference)	0.7936
[el->en] MWE Removal (MWE reference)	0.8087
[el->en] MWE Removal (Plain reference)	0.8181
[el->en] MWE Addition (MWE reference)	0.7999
[el->en] MWE Addition (Plain reference)	0.7930
[el->en] Plain Translation (MWE reference)	0.8145
[el->en] Plain Translation (Plain reference)	0.8320

Table 11: StyleCOMET scores for Gemini translation outputs evaluated against target and alternative style references, on the test set

Task	StyleCOMET
[en->el] MWE Preservation (MWE reference)	0.8225
[en->el] MWE Preservation (Plain reference)	0.8106
[en->el] MWE Removal (MWE reference)	0.8090
[en->el] MWE Removal (Plain reference)	0.8363
[en->el] MWE Addition (MWE reference)	0.8129
[en->el] MWE Addition (Plain reference)	0.8255
[en->el] Plain Translation (MWE reference)	0.8123
[en->el] Plain Translation (Plain reference)	0.8454
[el->en] MWE Preservation (MWE reference)	0.8178
[el->en] MWE Preservation (Plain reference)	0.7889
[el->en] MWE Removal (MWE reference)	0.8116
[el->en] MWE Removal (Plain reference)	0.8118
[el->en] MWE Addition (MWE reference)	0.8090
[el->en] MWE Addition (Plain reference)	0.8034
[el->en] Plain Translation (MWE reference)	0.8137
[el->en] Plain Translation (Plain reference)	0.8310

Table 12: StyleCOMET scores for Claude translation outputs evaluated against target and alternative style references, on the test set

Task	StyleCOMET
[en->el] MWE Preservation (MWE reference)	0.8158
[en->el] MWE Preservation (Plain reference)	0.8086
[en->el] MWE Removal (MWE reference)	0.8074
[en->el] MWE Removal (Plain reference)	0.8295
[en->el] MWE Addition (MWE reference)	0.8085
[en->el] MWE Addition (Plain reference)	0.8310
[en->el] Plain Translation (MWE reference)	0.8068
[en->el] Plain Translation (Plain reference)	0.8384
[el->en] MWE Preservation (MWE reference)	0.8109
[el->en] MWE Preservation (Plain reference)	0.7857
[el->en] MWE Removal (MWE reference)	0.8094
[el->en] MWE Removal (Plain reference)	0.8035
[el->en] MWE Addition (MWE reference)	0.8070
[el->en] MWE Addition (Plain reference)	0.8054
[el->en] Plain Translation (MWE reference)	0.8114
[el->en] Plain Translation (Plain reference)	0.8282

Table 13: StyleCOMET scores for Mistral translation outputs evaluated against target and alternative style references, on the test set

C.2 t-tests

“Supported" means a hypothesis was true with statistical significance. “Reverse" means the reverse of the hypothesis was true with statistical significance.

C.2.1 Style dataset comparisons

Using the human-generated sentences, each hypothesis tests whether a translation’s quality is estimated to be significantly higher if its style matches that of the sources compared to using the reverse one, and also whether there is a significant bias in favour of matched plain translations as compared to matched MWE translations.

Hypothesis Tested	p-value	Result
`[el->en] MWE Preservation`	1.1×10^{-3}	Supported
`[el->en] Plain Preservation`	4.0×10^{-119}	Supported
`[el->en] Plain over MWE`	2.5×10^{-124}	Supported
`[en->el] MWE Preservation`	4.3×10^{-40}	Reversed
`[en->el] Plain Preservation`	4.2×10^{-236}	Supported
`[en->el] Plain over MWE`	5.1×10^{-133}	Supported

Table 14: Paired t-test hypotheses on quality estimation (COMETKIWI) evaluations on the full ground-truth data

C.2.2 Style reference comparisons

Each hypothesis tests whether the given evaluation model scores the given LLM's translations on a specific task higher when the reference given has the same style as was the intention, compared to when the reference has the opposite style. In effect, it tests success on the given task style-wise.

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	7.0×10^{-80}	Supported
[el->en] MWE Preservation	1.0×10^{-87}	Supported
[el->en] MWE Removal	2.1×10^{-24}	Supported
[el->en] MWE Addition	3.7×10^{-01}	Insignificant
[en->el] Plain Translation	1.2×10^{-149}	Supported
[en->el] MWE Preservation	4.1×10^{-27}	Supported
[en->el] MWE Removal	2.3×10^{-95}	Supported
[en->el] MWE Addition	9.0×10^{-05}	Reversed

Table 15: Paired t-test hypotheses on COMET evaluations on Gemini outputs, on the entire dataset

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	8.3×10^{-81}	Supported
[el->en] MWE Preservation	3.0×10^{-99}	Supported
[el->en] MWE Removal	6.4×10^{-02}	Insignificant
[el->en] MWE Addition	8.3×10^{-02}	Insignificant
[en->el] Plain Translation	7.8×10^{-171}	Supported
[en->el] MWE Preservation	9.8×10^{-16}	Supported
[en->el] MWE Removal	3.6×10^{-126}	Supported
[en->el] MWE Addition	5.7×10^{-41}	Reversed

Table 16: Paired t-test hypotheses on COMET evaluations on Claude outputs, on the entire dataset

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	1.3×10^{-67}	Supported
[el->en] MWE Preservation	2.1×10^{-91}	Supported
[el->en] MWE Removal	2.7×10^{-02}	Reversed
[el->en] MWE Addition	2.5×10^{-05}	Reversed
[en->el] Plain Translation	6.9×10^{-168}	Supported
[en->el] MWE Preservation	2.9×10^{-07}	Supported
[en->el] MWE Removal	2.0×10^{-78}	Supported
[en->el] MWE Addition	2.0×10^{-99}	Reversed

Table 17: Paired t-test hypotheses on COMET evaluations on Mistral outputs, the entire dataset

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	7.7×10^{-18}	Supported
[el->en] MWE Preservation	3.8×10^{-19}	Supported
[el->en] MWE Removal	1.1×10^{-07}	Supported
[el->en] MWE Addition	2.8×10^{-01}	Insignificant
[en->el] Plain Translation	1.1×10^{-39}	Supported
[en->el] MWE Preservation	4.8×10^{-04}	Supported
[en->el] MWE Removal	8.7×10^{-23}	Supported
[en->el] MWE Addition	7.8×10^{-03}	Reversed

Table 18: Paired t-test hypotheses on COMET evaluations on Gemini outputs on the test data

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	1.3×10^{-18}	Supported
[el->en] MWE Preservation	1.1×10^{-26}	Supported
[el->en] MWE Removal	1.3×10^{-01}	Insignificant
[el->en] MWE Addition	3.5×10^{-01}	Insignificant
[en->el] Plain Translation	2.0×10^{-44}	Supported
[en->el] MWE Preservation	2.9×10^{-03}	Supported
[en->el] MWE Removal	2.3×10^{-30}	Supported
[en->el] MWE Addition	4.0×10^{-14}	Reversed

Table 19: Paired t-test hypotheses on COMET evaluations on Claude outputs on the test data

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	1.6×10^{-16}	Supported
[el->en] MWE Preservation	1.5×10^{-23}	Supported
[el->en] MWE Removal	1.4×10^{-01}	Insignificant
[el->en] MWE Addition	2.6×10^{-01}	Insignificant
[en->el] Plain Translation	1.3×10^{-44}	Supported
[en->el] MWE Preservation	8.6×10^{-02}	Insignificant
[en->el] MWE Removal	2.2×10^{-22}	Supported
[en->el] MWE Addition	8.9×10^{-29}	Reversed

Table 20: Paired t-test hypotheses on COMET evaluations on Mistral outputs on the test data

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	2.7×10^{-15}	Supported
[el->en] MWE Preservation	7.5×10^{-25}	Supported
[el->en] MWE Removal	9.8×10^{-06}	Supported
[el->en] MWE Addition	3.6×10^{-04}	Supported
[en->el] Plain Translation	5.0×10^{-40}	Supported
[en->el] MWE Preservation	1.5×10^{-09}	Supported
[en->el] MWE Removal	2.9×10^{-20}	Supported
[en->el] MWE Addition	6.6×10^{-01}	Insignificant

Table 21: Paired t-test hypotheses on StyleCOMET evaluations on Gemini outputs on the test data

Hypothesis Tested	p-value	Result
[el->en] Plain Translation	3.7×10^{-15}	Supported
[el->en] MWE Preservation	1.2×10^{-32}	Supported
[el->en] MWE Removal	9.4×10^{-01}	Insignificant
[el->en] MWE Addition	1.4×10^{-02}	Supported
[en->el] Plain Translation	3.7×10^{-44}	Supported
[en->el] MWE Preservation	7.6×10^{-07}	Supported
[en->el] MWE Removal	1.3×10^{-28}	Supported
[en->el] MWE Addition	2.5×10^{-09}	Reversed

Table 22: Paired t-test hypotheses on StyleCOMET evaluations on Claude outputs on the test data

Hypothesis Tested	p**-value**	**Result**
`[el->en] Plain Translation`	1.8×10^{-13}	Supported
`[el->en] MWE Preservation`	2.1×10^{-27}	Supported
`[el->en] MWE Removal`	7.7×10^{-03}	Reversed
`[el->en] MWE Addition`	4.5×10^{-01}	Insignificant
`[en->el] Plain Translation`	3.1×10^{-44}	Supported
`[en->el] MWE Preservation`	3.1×10^{-04}	Supported
`[en->el] MWE Removal`	1.4×10^{-21}	Supported
`[en->el] MWE Addition`	8.9×10^{-25}	Reversed

Table 23: Paired t-test hypotheses on StyleCOMET evaluations on Mistral outputs on the test data

C.2.3 Style source comparisons

Each hypothesis tests whether the given evaluation model scores the given LLM's translations towards a desired style higher when the source style is the same compared to when it is the reverse, the reference always being aligned in style with the intention. In effect, it tests whether performance is rated higher in tasks where style is preserved compared to ones where it is reversed.

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	1.5×10^{-63}	Supported
`[el->en] To MWE`	2.6×10^{-28}	Supported
`[en->el] To Plain`	7.7×10^{-45}	Supported
`[en->el] To MWE`	2.9×10^{-18}	Supported

Table 24: Paired t-test hypotheses on COMET evaluations on Gemini outputs, on the entire dataset

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	3.1×10^{-92}	Supported
`[el->en] To MWE`	2.2×10^{-01}	Insignificant
`[en->el] To Plain`	1.7×10^{-35}	Supported
`[en->el] To MWE`	9.7×10^{-03}	Supported

Table 25: Paired t-test hypotheses on COMET evaluations on Claude outputs, on the entire dataset

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	2.1×10^{-106}	Supported
`[el->en] To MWE`	8.7×10^{-04}	Reversed
`[en->el] To Plain`	2.1×10^{-47}	Supported
`[en->el] To MWE`	5.7×10^{-03}	Reversed

Table 26: Paired t-test hypotheses on COMET evaluations on Mistral outputs, the entire dataset

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	1.6×10^{-14}	Supported
`[el->en] To MWE`	4.0×10^{-07}	Supported
`[en->el] To Plain`	2.3×10^{-10}	Supported
`[en->el] To MWE`	4.3×10^{-03}	Supported

Table 27: Paired t-test hypotheses on COMET evaluations on the Gemini outputs on the test data

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	2.5×10^{-22}	Supported
`[el->en] To MWE`	5.1×10^{-01}	Insignificant
`[en->el] To Plain`	2.9×10^{-10}	Supported
`[en->el] To MWE`	8.5×10^{-01}	Insignificant

Table 28: Paired t-test hypotheses on COMET evaluations on Claude outputs on the test data

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	1.3×10^{-29}	Supported
`[el->en] To MWE`	1.9×10^{-02}	Reversed
`[en->el] To Plain`	3.5×10^{-07}	Supported
`[en->el] To MWE`	4.9×10^{-01}	Insignificant

Table 29: Paired t-test hypotheses on COMET evaluations on Mistral outputs on the test data

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	2.9×10^{-11}	Supported
`[el->en] To MWE`	1.9×10^{-17}	Supported
`[en->el] To Plain`	5.6×10^{-08}	Supported
`[en->el] To MWE`	7.9×10^{-11}	Supported

Table 30: Paired t-test hypotheses on StyleCOMET evaluations on Gemini outputs on the test data

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	3.7×10^{-18}	Supported
`[el->en] To MWE`	2.1×10^{-04}	Supported
`[en->el] To Plain`	5.1×10^{-05}	Supported
`[en->el] To MWE`	1.8×10^{-06}	Supported

Table 31: Paired t-test hypotheses on StyleCOMET evaluations on Claude outputs on the test data

Hypothesis Tested	p-value	Result
`[el->en] To Plain`	5.0×10^{-24}	Supported
`[el->en] To MWE`	8.9×10^{-02}	Insignificant
`[en->el] To Plain`	8.2×10^{-05}	Supported
`[en->el] To MWE`	1.5×10^{-04}	Supported

Table 32: Paired t-test hypotheses on StyleCOMET evaluations on Mistral outputs on the test data

C.2.4 Language comparisons

Ground Truth Pair	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7333	0.7238	1.2×10^{-11}	el → en
MWE Preservation	0.6497	0.6239	4.9×10^{-62}	el → en

Table 33: Cross-lingual directionality comparison on quality estimation (COMETKIWI) evaluations on parallel ground-truth data

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7770	0.8526	1.4×10^{-233}	en → el
MWE Preservation	0.7158	0.7872	6.0×10^{-164}	en → el
MWE Removal	0.7196	0.8241	2.1×10^{-234}	en → el
MWE Addition	0.7238	0.7928	5.1×10^{-137}	en → el

Table 34: Cross-lingual directionality comparison on COMET evaluations on Gemini outputs, on the entire dataset

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7849	0.8611	1.3×10^{-301}	en → el
MWE Preservation	0.7261	0.7987	6.7×10^{-182}	en → el
MWE Removal	0.7353	0.8394	1.5×10^{-249}	en → el
MWE Addition	0.7230	0.7935	5.2×10^{-136}	en → el

Table 35: Cross-lingual directionality comparison on COMET evaluations on Claude outputs, on the entire dataset

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7770	0.8526	1.4×10^{-233}	en → el
MWE Preservation	0.7158	0.7872	6.0×10^{-164}	en → el
MWE Removal	0.7196	0.8241	2.1×10^{-234}	en → el
MWE Addition	0.7238	0.7928	5.1×10^{-137}	en → el

Table 36: Cross-lingual directionality comparison on COMET evaluations on Mistral outputs, the entire dataset

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7825	0.8564	3.6×10^{-46}	en → el
MWE Preservation	0.7307	0.7901	1.2×10^{-26}	en → el
MWE Removal	0.7466	0.8231	7.5×10^{-30}	en → el
MWE Addition	0.7053	0.7757	5.2×10^{-29}	en → el

Table 37: Cross-lingual directionality comparison on COMET evaluations on Gemini outputs, on the test set

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7825	0.8634	9.6×10^{-75}	en → el
MWE Preservation	0.7253	0.7937	1.7×10^{-39}	en → el
MWE Removal	0.7318	0.8389	3.8×10^{-61}	en → el
MWE Addition	0.7217	0.7928	2.4×10^{-31}	en → el

Table 38: Cross-lingual directionality comparison on COMET evaluations on Claude outputs, on the test set

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.7748	0.8499	8.6×10^{-50}	en → el
MWE Preservation	0.7105	0.7851	4.1×10^{-39}	en → el
MWE Removal	0.7092	0.8256	6.2×10^{-56}	en → el
MWE Addition	0.7225	0.7883	1.2×10^{-26}	en → el

Table 39: Cross-lingual directionality comparison on COMET evaluations on Mistral outputs, on the test set

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.8320	0.8427	1.2×10^{-06}	en → el
MWE Preservation	0.8199	0.8221	3.6×10^{-01}	Tie (Insignificant)
MWE Removal	0.8181	0.8275	2.8×10^{-03}	en → el
MWE Addition	0.7999	0.8068	3.1×10^{-03}	en → el

Table 40: Cross-lingual directionality comparison on StyleCOMET evaluations on Gemini outputs, on the test set

Task Evaluated	el → en Mean	en → el Mean	p-value	Winner
Plain Translation	0.8310	0.8454	1.2×10^{-14}	en → el
MWE Preservation	0.8178	0.8225	1.5×10^{-02}	en → el
MWE Removal	0.8118	0.8363	1.4×10^{-18}	en → el
MWE Addition	0.8090	0.8129	8.8×10^{-02}	Tie (Insignificant)

Table 41: Cross-lingual directionality comparison on StyleCOMET evaluations on Claude outputs, on the test set

Task Evaluated	**el → en Mean**	**en → el Mean**	p**-value**	**Winner**
Plain Translation	0.8282	0.8384	4.7×10^{-07}	en → el
MWE Preservation	0.8109	0.8158	2.3×10^{-02}	en → el
MWE Removal	0.8035	0.8295	1.3×10^{-19}	en → el
MWE Addition	0.8070	0.8085	5.2×10^{-01}	Tie (Insignificant)

Table 42: Cross-lingual directionality comparison on StyleCOMET evaluations on Mistral outputs, on the test set