# LREC-Stylist-Title Generation-Manual Evaluation

Input: abstract of a research paper; output: title of a research paper 20 tokens approx.

For each title, choose a rating from 1(poor) to 5(excellent) to describe how suitable you think the title is.

You should consider whether the title adequately describes the main focus of the paper and follows the stated style (e.g., rhyming).

\* Indicates required question

1.	Evaluator Name *

# QP1

Multimodal machine translation (MMT) is a challenging task that seeks to improve translation quality by incorporating visual information. However, recent studies have indicated that the visual information provided by existing MMT datasets is insufficient, causing models to disregard it and overestimate their capabilities. This issue presents a significant obstacle to the development of MMT research. This paper presents a novel solution to this issue by introducing 3AM, an ambiguity-aware MMT dataset comprising 26,000 parallel sentence pairs in English and Chinese, each with corresponding images. Our dataset is specifically designed to include more ambiguity and a greater variety of both captions and images than other MMT datasets. We utilize a word sense disambiguation model to select ambiguous data from vision-and-language datasets, resulting in a more challenging dataset. We further benchmark several state-of-the-art MMT models on our proposed dataset. Experimental results show that MMT models trained on our dataset exhibit a greater ability to exploit visual information than those trained on other MMT datasets. Our work provides a valuable resource for researchers in the field of multimodal learning and encourages further exploration in this area. The data, code and scripts are freely available at https://github.com/MaxyLee/3AM.

Dataset	AM: An A	3 ,			Widoriirio	
Catchy: 3AM: En Richer Visual Cue	_	Ambiguity	-Aware M	lultimodal	Machine	Translation with
Rhyming: Visual (	Clarity for	Translati	onal Verit	y: Introdu	cing 3AM	l for Multimodal
Funny: Lost in Tra	anslation:	When Im	ages Spe	eak Ambig	guity	
Comprehensive: 3 Aware Visual Data Plain: 3AM: Ambig	asets					
Mark only one oval p	per row.					
	ner row.	2	3	4	5	_
Author-Written:  Catchy:		2	3	4	5	_
Author-Written:		2	3	4	5	_
Author-Written: Catchy:			3	4	5	
Author-Written: Catchy: Rhyming:			3	4	5	

Natural Language Processing (NLP) models tend to inherit and amplify stereotypical biases present in their training data, leading to harmful societal consequences. Current efforts to rectify these biases typically revolve around making models oblivious to bias, which is at odds with the idea that humans require increased awareness to tackle these biases better. This prompts a fundamental research question: are bias-oblivious models the only viable solution to combat stereotypical biases? This paper answers this question by proposing the Agency-BeLiefs Embedding (ABLE) model, a novel approach that actively encodes stereotypical biases into the embedding space. ABLE draws upon social psychological theory to acquire and represent stereotypical biases in the form of agency and belief scores rather than directly representing stereotyped groups. Our experimental results showcase ABLE's effectiveness in learning agency and belief stereotypes while preserving the language model's proficiency. Furthermore, we underscore the practical significance of incorporating stereotypes within the ABLE model by demonstrating its utility in various downstream tasks. Our approach exemplifies the potential benefits of addressing bias through awareness, as opposed to the prevailing approach of mitigating bias through obliviousness.

		LREC-S	Stylist-Title Ger	neration-Manu	al Evaluation	
Author-Written: Althrough Awarenes	Ŭ	•		dding to A	ddress S	tereotypical Bias
Catchy: ABLE Mo	del: Addr	essing Bi	as in NLF	Through	Awarene	ess, Not Oblivion
Rhyming: Data to	Combat	the Hate	Debate: A	ABLE's Ro	ole in Bias	s Weighing Rate
Funny: Bias Be G	one: The	Awkward	I Truth Ab	out NLP I	Models	
Comprehensive: A	ABLE: A N	Novel App	roach to	Encoding	Stereoty	pical Biases in
NLP Models						
Plain: ABLE Mode	el: Addres	ssing Ster	eotypical	Biases in	NLP	
Mark only one oval រុ	per row.					
	1	2	3	4	5	_
Author-Written:						_
Catchy:						_
Rhyming:						_
Funny:						_
Comprehensive:						_
Plain						
						_

the given ones if you were to retitle the paper.

Any Comments:		

Previous stance detection studies typically concentrate on evaluating stances within individual instances, thereby exhibiting limitations in effectively modeling multi-party discussions concerning the same specific topic, as naturally transpire in authentic social media interactions. This constraint arises primarily due to the scarcity of datasets that authentically replicate real social media contexts, hindering the research progress of conversational stance detection. In this paper, we introduce a new multi-turn conversation stance detection dataset (called MT-CSD), which encompasses multiple targets for conversational stance detection. To derive stances from this challenging dataset, we propose a global-local attention network (GLAN) to address both long and short-range dependencies inherent in conversational data. Notably, even state-of-the-art stance detection methods, exemplified by GLAN, exhibit an accuracy of only 50.47%, highlighting the persistent challenges in conversational stance detection. Furthermore, our MT-CSD dataset serves as a valuable resource to catalyze advancements in cross-domain stance detection, where a classifier is adapted from a different yet related target. We believe that MT-CSD will contribute to advancing real-world applications of stance detection research. Our source code, data, and models are available at https://github.com/nfq729/MT-CSD.

Author-\N/ritten. A		LINEO-S	stylist-Title Gel	neration-Manu	ai Evaluation				
Author-Written: A Challenge Dataset and Effective Models for Conversational Stance Detection									
Catchy: MT-CSD: Attention Network		ng Conve	rsational	Stance D	etection w	vith Global-Lo			
Rhyming: Multi-Tu Way	ırn Data 1	for Stance	e to Relay	/: Advanci	ng Conve	ersations the			
Funny: Stance Wa	ars: The (	Quest for	Accurate	Social Me	edia Opini	ions			
Comprehensive: No Detection in Conv			urn Datas	set and Gl	LAN for E	nhanced Sta			
Plain: MT-CSD: M Model	lulti-Turn	Conversa	ation Stan	nce Detec	tion Datas	set and GLAN			
Mark only one oval p	per row.								
	1	2	3	4	5				
Author-Written:	1	2	3	4	5	_			
Author-Written: Catchy:	1	2	3	4	5	_			
		2	3	4	5				
Catchy:			3	4	5				
Catchy: Rhyming:			3	4	5				
Catchy:  Rhyming:  Funny:					5				

Empathy is critical for effective communication and mental health support, and in many online health communities people anonymously engage in conversations to seek and provide empathetic support. The ability to automatically recognize and detect empathy contributes to the understanding of human emotions expressed in text, therefore advancing natural language understanding across various domains. Existing empathy and mental health-related corpora focus on broader contexts and lack domain specificity, but similarly to other tasks (e.g., learning distinct patterns associated with COVID-19 versus skin allergies in clinical notes), observing empathy within different domains is crucial to providing tailored support. To address this need, we introduce AcnEmpathize, a dataset that captures empathy expressed in acne-related discussions from forum posts focused on its emotional and psychological effects. We find that transformer-based models trained on our dataset demonstrate excellent performance at empathy classification. Our dataset is publicly released to facilitate analysis of domain-specific empathy in online conversations and advance research in this challenging and intriguing domain.

	ions	ппаш-оре	∍ome ⊑m¢	Janry Dele	5011011 III <i>F</i>	Acne-Relate
Rhyming: Empath Fail	ny Unveile	ed in Acne	e's Tale: A	∖cnEmpat	hize for S	Support With
Funny: Acne and	Empathy	: Spotting	Support	in Skin De	eep Conv	ersations!
Comprehensive: A	•			omain-Sp	ecific Em	pathy in Ac
Discussions for E	nhanced	NLP Sup	port			
Plain: AcnEmpath	nize: Dom	ain-Spec	ific Empa	thy Detec	tion in Or	nline Health
Communities						
Mark onlv one oval เ	per row.					
Mark only one oval เ	per row.	2	3	4	5	
Mark only one oval p		2	3	4	5	_
		2	3	4	5	_
Author-Written:		2	3	4	5	
Author-Written: Catchy:			3	4	5	
Author-Written: Catchy: Rhyming			3	4	5	_
Author-Written: Catchy: Rhyming Funny:			3	4	5	

Sentence representation learning is a fundamental task in NLP. Existing methods use contrastive learning (CL) to learn effective sentence representations, which benefit from high-quality contrastive data but require extensive human annotation. Large language models (LLMs) like ChatGPT and GPT4 can automatically generate such data. However, this alternative strategy also encounters challenges: 1) obtaining high-quality generated data from small-parameter LLMs is difficult, and 2) inefficient utilization of the generated data. To address these challenges, we propose a novel adaptive reinforcement tuning (ART) framework. Specifically, to address the first challenge, we introduce a reinforcement learning approach for fine-tuning small-parameter LLMs, enabling the generation of highquality hard contrastive data without human feedback. To address the second challenge, we propose an adaptive iterative framework to guide the small-parameter LLMs to generate progressively harder samples through multiple iterations, thereby maximizing the utility of generated data. Experiments conducted on seven semantic text similarity tasks demonstrate that the sentence representation models trained using the synthetic data generated by our proposed method achieve state-of-the-art performance. Our code is available at https://github.com/WuNein/AdaptCL.

A 41 \ \ \ / - : 44 \ A .	d	_		eration-Manual		la an Hand
Author-Written: Ad Generators for Se				ng Langua	age Mode	els as Hard
Catchy: ART: Enh Reinforcement Tu	•		•		th Adaptiv	/e
Rhyming: Data to	Elevate:	Contrasti	ve Learni	ing's New	Fate with	n ART We C
Funny: Teaching <sup>-</sup>	Tiny Mod	els to Pla	y Hardbal	ll with Ser	ntence Sm	narts!
Comprehensive: A	•			•	nhanced S	Sentence
Representation Le	earning U	Jsing LLM	l-Generat	ed Data		
Plain: Adaptive Re	einforcen	nent Tunir	ng for Effe	ective Sen	tence Re	presentatio
Learning			.9 .0			p. 5 5 5
· ·						
Mark only one oval រុ	per row.					
Mark only one oval រុ	per row. 1	2	3	4	5	_
Mark only one oval p		2	3	4	5	_
		2	3	4	5	
Author-Written:		2	3	4	5	
Author-Written: Catchy:			3	4	5	- - -
Author-Written: Catchy: Rhyming			3	4	5	- - -
Author-Written: Catchy: Rhyming Funny:			3		5	
Author-Written: Catchy: Rhyming Funny: Comprehensive:			3		5	- - -
Author-Written: Catchy: Rhyming Funny: Comprehensive:			3		5	- - -

 $https://docs.google.com/forms/d/1YdrFDhxbBa\_agRQFxGXWxzF1BCLhFtMFo4PZAvF3t98/editable for the control of the$ 

Traditional non-simultaneous Sign Language Translation (SLT) methods, while effective for pre-recorded videos, face challenges in real-time scenarios due to inherent inference delays. The emerging field of simultaneous SLT aims to address this issue by progressively translating incrementally received sign video. However, the sole existing work in simultaneous SLT adopts a fixed gloss-based policy, which suffer from limitations in boundary prediction and contextual comprehension. In this paper, we delve deeper into this area and propose an adaptive policy for simultaneous SLT. Our approach introduces the concept of "confident translation length, denoting maximum accurate translation achievable from current input. An estimator measures this length for streaming sign video, enabling the model to make informed decisions on whether to wait for more input or proceed with translation. To train the estimator, we construct a training data of confident translation length based on the longest common prefix between translations of partial and complete inputs. Furthermore, we incorporate adaptive training, utilizing pseudo prefix pairs, to refine the offline translation model for optimal performance in simultaneous scenarios. Experimental results on PHOENIX2014T and CSL-Daily demonstrate the superiority of our adaptive policy over existing methods, particularly excelling in situations requiring extremely low latency.

12. Author-Written: Adaptive Simultaneous Sign Language Translation with Confident Translation Length Estimation

Catchy: Adaptive Policy for Low-Latency Simultaneous Sign Language Translation with Confident Translation Length

Rhyming: Adapting to Signs: Lengths of Translation Align for Real-Time Design

Funny: Translating Signs at Lightning Speed: No More Delays, Just Displays!

Comprehensive: Adaptive Policy for Real-Time Sign Language Translation Using Confident Translation Length Estimation

Plain: Adaptive Policy for Simultaneous Sign Language Translation

Mark only one oval per row.

1	2	3	4	5

13.	In the COMMENT box, mention the title (e.g., Rhyming) you would choose
	among the given ones if you were to retitle the paper.

With an auxiliary corpus (non-target speaker corpus) for model pre-training, Text-to-Speech (TTS) methods can generate high-quality speech with a limited target speaker corpus. However, this approach comes with expensive training costs. To overcome the challenge, a high-quality TTS method is proposed, significantly reducing training costs while maintaining the naturalness of synthesized speech. In this paper, we propose an auxiliary corpus compression algorithm that reduces the training cost while the naturalness of the synthesized speech is not significantly degraded. We then use the compressed corpus to pre-train the proposed TTS model CMDTTS, which fuses phoneme and word multi-level prosody modeling components and denoises the generated melspectrograms using denoising diffusion probabilistic models (DDPMs). In addition, a fine-tuning step that the conditional generative adversarial network (cGAN) is introduced to embed the target speaker feature and improve speech quality using the target speaker corpus. Experiments are conducted on Chinese and English single speakers corpora, and the results show that the method effectively balances the model training speed and the synthesized speech quality and outperforms the current models.

14. Author-Written: A Fast and High-quality Text-to-Speech Method with Compressed Auxiliary Corpus and Limited Target Speaker Corpus

Catchy: CMDTTS: Efficient TTS with Compressed Auxiliary Corpus and Multi-Level Prosody Modeling

Rhyming: Compressing the Source for Speech with No Force: CMDTTS Redefines the TTS Course

Funny: Talk Is Cheap: Cutting Costs While Keeping Voices Smooth!

Comprehensive: Cost-Effective Text-to-Speech Synthesis with Compressed Auxiliary Corpus and Enhanced Naturalness

Plain: Efficient TTS Training with Auxiliary Corpus Compression

Mark only one oval per row.

1	2	3	4	5

15.	In the COMMENT box, mention the title (e.g., Rhyming) you would choose
	among the given ones if you were to retitle the paper.

,y	Commonto.

The task of financial analysis primarily encompasses two key areas: stock trend prediction and the corresponding financial question answering. Currently, machine learning and deep learning algorithms (ML&DL) have been widely applied for stock trend predictions, leading to significant progress. However, these methods fail to provide reasons for predictions, lacking interpretability and reasoning processes. Also, they can not integrate textual information such as financial news or reports. Meanwhile, large language models (LLM) have remarkable textual understanding and generation ability. But due to the scarcity of financial training datasets and limited integration with real-time knowledge, LLM still suffer from hallucinations and unable to keep up with the latest information. To tackle these challenges, we first release AlphaFin datasets, combining traditional research datasets, real-time financial data, and handwritten chain-of-thought (CoT) data. It has positive impact on training LLM for completing financial analysis. We then use AlphaFin datasets to benchmark a state-of-the-art method, called Stock-Chain, for effectively tackling the financial analysis task, which integrates retrieval-augmented generation (RAG) techniques. Extensive experiments are conducted to demonstrate the effectiveness of our framework on financial analysis.

16. Author-Written : AlphaFin: Benchmarking Financial Analysis with Retrieval-Augmented Stock-Chain Framework

Catchy: Stock-Chain: Enhancing Financial Analysis with AlphaFin Dataset and Retrieval-Augmented Generation

Rhyming: AlphaFin Unleashed: A Dataset to Teach Financial Trends and Insights That Reach

Funny: Money Talks: Decoding Stocks Without Losing Our Minds!

Comprehensive: AlphaFin: A Comprehensive Dataset and Framework for Interpretable Financial Analysis Using LLMs

Plain: AlphaFin: Integrating LLM for Financial Analysis

Mark only one oval per row.

	1	2	3	4	5
Author-Written:					
Catchy:					
Rhyming:					
Funny:					
Comprehensive:					
Plain:					

17.	In the COMMENT box, mention the title (e.g., Rhyming) you would choose
	among the given ones if you were to retitle the paper.

Any Comments:		

Due to the lack of parallel data, the mainstream fine-tuning-based domain adaptation methods have the overfitting problem in the translation of low-resource domains, and it is difficult for the model to learn the in-domain generalization knowledge. To address the above issue, in this work, we propose a novel Reinforcement Learning Domain Adaptation method for Neural Machine Translation (RLDA-NMT) in the low-resource domain. RLDA-NMT utilizes in-domain source monolingual data to make up for the lack of parallel data, and reinforces domain features learning to make the translation model learn the domainspecific knowledge more fully. Specifically, we first train a ranking-based model with a small-scale in-domain parallel corpus, and then adopt it as the reward model to select higher-quality generated translations for reinforcement when fine-tuning pre-trained NMT model using in-domain source monolingual data. We conduct experiments on Education, Laws, Thesis, and Patent domains of Chinese⇔English translation tasks. Experimental results demonstrate that RLDA-NMT can alleviate overfitting and reinforce the NMT model to learn domain-specific knowledge. Additionally, the results also show that RLDA-NMT and back-translation (BT) are nicely complementary to each other, where combining RLDA-NMT with BT can further improve translation quality.

	ten: A Reinfor anslation Lev		• .	•	Improve L	₋ow-Resourd
	DA-NMT: Rei			_		ain Adaptatio
Rhyming: R Resource F	einforcement rustration	for Transla	ation's Ne	w Station:	RLDA-NI	MT Tackles L
Funny: Lost	in Translatio	n? Not Any	more—R	einforcem	ent to the	Rescue!
Comprehen	sive: RLDA-N	NMT: Enhar	ncing Low	-Resource	e Neural M	/lachine
Translation <sup>*</sup>	Through Reir	nforcement	Learning	and Mond	olingual Da	ata
	orcement Lea	arning Dom	ain Adapt	ation for L	₋ow-Resoı	urce NMT
IVIAIR OIIIY OIIE	e ovar per row.					
	1	2	3	4	5	
Author-Writt		2	3	4	5	-
Author-Writt		2	3	4	5	-
		2	3	4	5	-
Catchy:			3	4	5	-
Catchy: Rhyming:	ten:		3	4	5	-
Catchy: Rhyming: Funny:	ten:				5	

Comparing annotations is a constant and necessary step in corpus analysis. Although the nature of these annotations is normally research-specific, the tools used for this purpose do not have to be. Here, we present a tool for extracting and comparing annotations from ELAN, despite their idiosyncrasies. The intention behind this tool is to provide a handy way to analyze ELAN annotated files, by comparing tiers to a reference unit. Using the presented tool, it is possible to see how tiers overlap (even if they are of symbolic type), to which ratio, and the displacement regarding a reference unit. We present an example of multimodal corpus analysis, regarding the coordination between speech and gesture units based on a pragmatic reference. We argue that looking into overlap ratios can be more informative of the association between speech and gestures, and that considering a time buffer between speech and gestural events can be misleading.

20. Author-Written: A Tool for Determining Distances and Overlaps between Multimodal Annotations

Catchy: Analyzing ELAN Annotations: A Tool for Comparing Multimodal Overlaps in Corpus Analysis

Rhyming: Comparative Insights in ELAN's Light: A Tool for Annotations That Delivers Clear Sight

Funny: Annotation Adventures: Finding Overlaps and Other Mischief!

Comprehensive: Tool for Analyzing ELAN Annotations: Comparing Speech and Gesture Overlap in Multimodal Corpora

Plain: Tool for Comparing ELAN Annotations in Multimodal Corpus Analysis

Mark only one oval per row.

	1	2	3	4	5
Author-Written:					
Catchy:					
Rhyming:					
Funny:					
Comprehensive:					
Plain:					
Plain:					

21.

а	among the given ones if you were to retitle the paper.
A	Any Comments:
_	

In the COMMENT box, mention the title (e.g., Rhyming) you would choose

This content is neither created nor endorsed by Google.

Google Forms