Event Extraction as Question Generation and Answering

Di Lu Shihao Ran Joel Tetreault Alejandro Jaimes
Dataminr Inc.
{dlu,sran,jtetreault,ajaimes}@dataminr.com

Abstract

Recent work on Event Extraction has reframed the task as Question Answering (QA), with promising results. The advantage of this approach is that it addresses the error propagation issue found in traditional token-based classification approaches by directly predicting event arguments without extracting candidates first. However, the questions are typically based on fixed templates and they rarely leverage contextual information such as relevant arguments. In addition, prior QA-based approaches have difficulty handling cases where there are multiple arguments for the same role. In this paper, we propose QGA-EE, which enables a Question Generation (QG) model to generate questions that incorporate rich contextual information instead of using fixed templates. We also propose dynamic templates to assist the training of QG model. Experiments show that QGA-EE outperforms all prior single-task-based models on the ACE05 English dataset.[1]

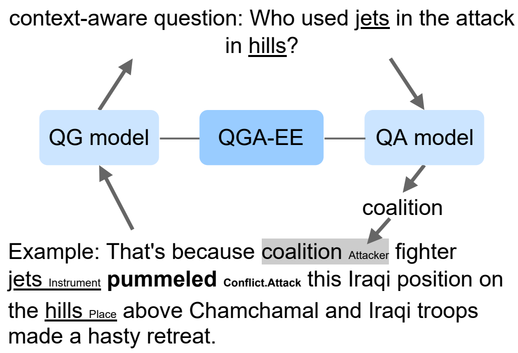

Figure 1: An event mention example from ACE. An ACE `Conflict.Attack` event with *pummeled* as trigger word and three event arguments: *coalition* (`Attacker`), *jets* (`Instrument`) and *hills* (`Place`).

1 Introduction

Event Extraction (EE) aims to extract core information elements (e.g. who, what, where, when) from text, and is a very important task in Natural Language Processing (NLP). It provides inputs to downstream applications such as Summarization (Filatova and Hatzivassiloglou, 2004), Knowledge Base Population (Ji and Grishman, 2011), and Recommendation (Lu et al., 2016).

Previous work (Li et al., 2013; Nguyen et al., 2016; Sha et al., 2018) is typically based on a pipeline approach, which first identifies the event trigger word/phrase and argument candidates, and then applies a classifier to the pair-wise features to classify the roles of the candidates. Unfortunately, errors tend to propagate down the pipeline. Recently, some approaches have formulated EE as a Question Answering (QA) problem (Du and Cardie, 2020; Li et al., 2020; Lyu et al., 2021) to mitigate the issue, in which questions for each argument role are manually defined by templates. For example, extracting the `Attack` argument from the `Conflict.Attack` event in the sentence *'That's because coalition fighter jets pummeled this Iraqi position on the hills above Chamchamal and Iraqi troops made a hasty retreat.'* is reframed as answering the question *'Who was the attacking agent?'* These approaches have shown promising results, but template-based questions are limiting: since the templates are built manually, they are fixed and rarely include contextual information (i.e., specific to the inputs), except for trigger words in some work (Du and Cardie, 2020). Formulating good questions, however, has been shown to improve performance for standard QA tasks (Rajpurkar et al., 2018). For QA-based EE, a question that incorporates richer contextual information such as other event arguments could yield better results (e.g. *'Who used jets in the attack in hills?'* in Figure 1).

In this paper, we propose QGA-EE, which consists of 1) a QG model for generating a context-aware question conditioned on a target argument

[1]Our code is available at `https://github.com/dataminr-ai/Event-Extraction-as-Question-Generation-and-Answering` for research purposes.

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 2: Short Papers, pages 1666–1688
July 9-14, 2023 ©2023 Association for Computational Linguistics

role and 2) a QA model for answering the context-aware question to extract the event argument. We also design dynamic templates to generate the gold context-aware questions for QG model training.

To the best of our knowledge, this is the first QA-based EE work that utilizes dynamic templates and focuses on generating context-aware questions. Li et al. (2020) also propose a model to generate questions that incorporate contextual information for both event trigger and arguments. However, our work has two main advantages. First, in Li et al. (2020) the question only incorporates the contextual information at the ontology level (e.g. argument role, event type). In our work, the generated questions incorporate contextual information at an event mention-level. For example, the question generated by our model includes the real event argument rather than just the argument role (e.g. 'hills' vs 'Place'). Second, the questions in their work are generated by filling in the templates, but our templates are dynamic and used to train the QG model which can automatically generate the optimal question given a specific event mention and the concerned argument role.

Experimental results show that QGA-EE outperforms all of the single-task-based models on the Automatic Content Extraction (ACE) 2005 English dataset (Doddington et al., 2004) and even achieves competitive performance with state-of-the-art joint IE models.

2 Model

Figure 1 shows the overall framework of QGA-EE. It focuses on Event Argument Extraction (EAE) only, but can be paired with any event trigger tagger to perform end-to-end EE. In Section 4, we pair it with a standard sequence labeling trigger tagger to evaluate its end-to-end EE performance.

2.1 Question Generation Model

Previous QA-based EE work (Du and Cardie, 2020) fills in pre-designed templates with trigger information to generate the input questions to the QA model. However missing contextual information in the questions is a bottleneck for the performance of the QA model.

QGA-EE uses a QG model to generate context-aware questions conditioned on the input sentence and target role, which is based on a sequence-to-sequence architecture (e.g. BART(Lewis et al., 2020), T5(Raffel et al., 2020)). In order to train the QG model, we design **Dynamic Templates** for each role in the ACE ontology.[2] We design multiple templates for each role, and each of them includes different combinations of other argument roles.

Who was the attacking agent?
Who attacked [Target]?
Who used [Instrument] in the attack?
Who made the attack in [Place]?
Who attacked [Target] using [Instrument]?
Who attacked [Target] in [Place]?
Who used [Instrument] in the attack in [Place]?
Who attacked [Target] using [Instrument] in [Place]?

Table 1: Dynamic templates for Attacker role in Conflict.Attack event with different combinations of known argument roles based on ACE ontology.

For example, the Conflict.Attack event in ACE has four predefined argument roles: Attacker, Target, Instrument and Place.[3] For the Attacker role, we exhaustively design eight templates using all of the possible combinations of the other roles included in the question (Table 1). When the model fills in the templates given a specific event mention, it is common that some of the predefined argument roles do not exist in the event mention. Thus the model only keeps the templates that contain the slots for argument roles appearing in the event mention. For the example in Figure 1, the Target role is not mentioned. So we ignore all of the templates that contain the [Target] slot, and we obtain four candidate questions for the Attacker role with corresponding arguments filled in: (1)*Who was the attacking agent?* (2) *Who used jets in the attack?* (3) *Who made the attack in hills?* (4) *Who used jets in the attack in hills?*

To train a QG model to generate the questions that cover as many contextual information as possible, we use the question that contains the most contextual arguments as the ground truth. For the example in Figure 1, we choose the question *'Who used jets in the attack in hills?'*, because it contains two arguments: *'jets'* and *'hills'*, the other three candidate questions listed above contain one or zero arguments. If more than one candidate question contains the most contextual arguments, we then pick the first one. The input and output examples for the QG model are as follows:

[2] https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-events-guidelines-v5.4.3.pdf

[3] We follow the experimental setting of prior work, which excludes all the Value and Timex. Thus the argument roles such as Time are not included.

Input: role: attacker context: That's because coalition fighter jets * pummeled * this Iraqi position on the hills above Chamchamal and Iraqi troops made a hasty retreat.
Output: Who used jets in the attack in hills?

2.2 Question Answering Model

Different from prior QA-based EE work that adapt an encoder-only architecture and predict the offsets of the event arguments (Chen et al., 2019; Du and Cardie, 2020; Li et al., 2020; Liu et al., 2020; Feng et al., 2020; Lyu et al., 2021; Zhou et al., 2021), our QA model is based on a sequence-to-sequence architecture (e.g. BART, T5), and generates the answer string directly. This enables prediction of multiple event arguments that are associated with the same role. Li et al. (2021) also adapts a generation model, but the input template is fixed. The examples of input and output are as follows:

Input: *question: Who was harmed in * injured * event? context: Injured Russian diplomats and a convoy of America's Kurdish comrades in arms were among unintended victims caught in crossfire and friendly fire Sunday.*
Output: *diplomats; convoy; victims < /s >*

Post-processing We split the output into a list of candidates (by ';'), and retrieve the arguments with offsets by exactly matching against the original sentence. We dynamically change the start position for searching to preserve the order of the retrieved event arguments. If an argument candidate cannot be matched with the original sentence, we discard it. Unlike the QG model, we use all of the possible questions as inputs during training for data augmentation purposes, and the size of the training data increases from 15,426 to 20,681. But in the testing phase, we use the single question generated by the QG model for each argument role.

3 Experimental Setup

3.1 Dataset and Evaluation Metrics

We conduct the experiments on the ACE 2005 English corpora, which has 33 event types and 22 argument roles. It contains 599 documents collected from newswire, weblogs, broadcast conversations, and broadcast news. More specifically, we follow the pre-processing steps in Wadden et al. (2019),[4] and evaluate our models on the resulting ACE05-E dataset.

For evaluation, we use the same criteria as prior work (Li et al., 2013): An **event trigger** is correctly identified if its offsets exactly match a reference. It is correctly classified if both its offsets and event type match a reference. An **event argument** is correctly identified (Arg-I) if its offsets and event type match a reference in the ground truth. It is correctly classified (Arg-C) if all of its offsets, event type, and argument role match a reference.

3.2 Compared Baselines

Model Variants. To evaluate the generalizability of our approach, we evaluate two QGA-EE variants: **QGA-EE**$_{BART}$ and **QGA-EE**$_{T5}$, which use BART and T5 as backbones respectively.

We compare the proposed models against SOTA EE models. **BERT QA** (Du and Cardie, 2020) use BERT as the encoder and predict the positions of the argument directly with role-driven questions. **TANL** (Paolini et al., 2021) transfers input sentences into augmented natural language sentences for structured prediction. **TEXT2EVENT** (Lu et al., 2021) is a sequence-to-structure network for event extraction.[5] Ma et al. (2020) utilizes dependency parses as additional features. **BART-Gen** (Li et al., 2021) is a BART-based generation model proposed for document-level event extraction.

We also compare with joint IE models trained on all of the ACE annotations which include entities, relations, and events. They benefit from additional information from other tasks and usually achieve better performance than the models trained on a single task. It is not fair to directly compare our model with the joint models since they incorporate more information beyond the standard EE training sets, but we still list their scores as a reference. **DYGIE++** (Wadden et al., 2019) is a BERT-based model that models span representations with within-sentence and cross-sentence context. **ONEIE** (Lin et al., 2020) leverages global features. **FourIE** (Nguyen et al., 2021) and **GraphIE** (Van Nguyen et al., 2022) are Graph Convolutional Networks-based models and **AMR-IE** (Zhang and Ji, 2021) utilizes AMR (Banarescu et al., 2013) parser.

3.3 Implementation Details

We conduct all of the experiments on a single V100 GPU. For finetuning, we use the Adafactor (Shazeer and Stern, 2018) optimizer with a

[4] https://github.com/dwadden/dygiepp

[5] DEGREE (Hsu et al., 2022) is not included because it is not evaluated on all of the argument roles.

learning rate of $1 * 10^{-4}$, weight decay of $1 * 10^{-5}$, and clip threshold of 1.0. We train the model for 20 epochs. Further details such as hyperparameters and data statics for model training and evaluation are in Appendix C.

4 Results

4.1 Event Argument Extraction Performance

	Arg-I	Arg-C
BERT_QA (Du and Cardie, 2020)	68.2	65.4
TANL^{+} (Paolini et al., 2021)	65.9	61.0
Ma et al. (2020)	-	62.1
BART-Gen (Li et al., 2021)	69.9	66.7
DYGIE++$^{*+}$ (Wadden et al., 2019)	66.2	60.7
ONEIE^{*+} (Lin et al., 2020)	73.2	69.3
QGA-EE$_{BART}$ (ours)	72.4	70.3
QGA-EE$_{T5}$ (ours)	**75.0**	**72.8**

Table 2: Event Extraction Results on ACE05-E test data (F1, %) with gold triggers. * models are trained with additional entity and relation data. $^{+}$ numbers are reported from Hsu et al. (2022), and others are from the original papers.

Table 2 shows the performance of QGA-EE models on ACE05-E test set with gold triggers.[6] Both QGA-EE variants outperform all other approaches, and using T5 as backbone provides an improvement of 2.5% over BART. The improvement over the prior QA-based models BERT_QA shows that generation-based QA models are more effective than position-based QA models for EE. QGA-EE$_{BART}$ outperforms the BART-based baseline BART-Gen and QGA-EE$_{T5}$ outperforms the T5-based baseline TANL, which demonstrates the effectiveness of our models with different backbones. Our models even outperform the joint IE models DYGIE++ and ONEIE, which leverage additional information from entities and relations.

4.2 Event Extraction Performance

We also evaluate our models on ACE05-E in a more "real world" fashion with *predicted* triggers extracted by an ALBERT-based (Lan et al., 2019) sequence-labeling model (Table 3).[7] Similar to the performance on gold triggers, QGA-EE benefits more from the T5 backbone on predicted triggers. Both QGA-EE variants outperform all the EE-task-centered baselines by more than 1% on Arg-C.

[6]Performance of FourIE, AMR-IE and GraphIE in gold triggers are not available in their original papers.

[7]The model is trained on ACE05-E and the F1 score on test set is 72.96%. More details in Appendix.

	Arg-I	Arg-C
BERT_QA (Du and Cardie, 2020)	54.1	53.1
TANL (Paolini et al., 2021)	50.1	47.6
TEXT2EVENT (Lu et al., 2021)	-	53.8
Ma et al. (2020)	56.7	54.3
BART-Gen (Li et al., 2021)	-	53.7
DYGIE++* (Wadden et al., 2019)	54.1	51.4
ONEIE* (Lin et al., 2020)	59.2	56.8
FourIE* (Nguyen et al., 2021)	60.7	58.0
AMR-IE* (Zhang and Ji, 2021)	**60.9**	58.6
GraphIE* (Van Nguyen et al., 2022)	**-**	**59.4**
QGA-EE$_{BART}$ (ours)	57.1	55.6
QGA-EE$_{T5}$ (ours)	59.8	57.9

Table 3: Event Extraction Results on ACE05-E test data (F1, %) with predicted triggers. * models are trained with additional entity and relation data. All numbers of baselines are reported from the original papers.

We also include the scores from SOTA joint IE models, DYGIE++, ONEIE, FourIE, AMR-IE and GraphIE, as reference. But, as stated earlier, it is not fair to compare our models directly with them, as they benefit from being trained with all of the annotations from entities, relations, and events. Also it should be noted that their trigger labeling models have more complicated architectures and thus perform significantly better than the sequence-labeling based tagger we use (F1 75.4% from FourIE and F1 74.7% from OneIE). This further boosts the end-to-end EE performance.

4.3 Ablation Study

Table 4 shows the ablation study of the QGA-EE$_{T5}$ model on the ACE05 test set with gold triggers. By replacing the QG model with simple context-unaware templates, the F1 score decreases by 1.65%. It demonstrates that the context-aware questions generated by our QG component enhance the end-to-end event argument extraction performance. Additionally, the generation-based QA model deals with multi-argument situations better and provides an improvement of 4.24%.

	Arg-I	Arg-C
QGA-EE$_{T5}$	75.04	72.78
- w/o pretrained QG	73.57	71.13
- w/o pretrained QG & mutli-arg support	69.61	66.89

Table 4: Ablation study with gold triggers on ACE05-E test set (F1, %).

4.4 Impact of Data Augmentation

As we mentioned in Section 2.2, the size of the training data increases from 15,426 to 20,681 as a benefit of our proposed dynamic templates. To evaluate the contribution of the data augmentation, we evaluate the performance of QGA-EE on ACE05 test data with partial training data (with gold triggers). With 40% of the training examples after data augmentation (8,272), QGA-EE achieves a F1 score of 71.42% on ACE05-E test set with gold triggers. It outperforms all of the baselines in Table 2, which demonstrates the effectiveness of our proposed model.

	Arg-I	Arg-C
QGA-EE$_{T5}$ with 100% training data	75.04	72.78
QGA-EE$_{T5}$ with 80% training data	73.86	71.64
QGA-EE$_{T5}$ with 60% training data	73.15	71.63
QGA-EE$_{T5}$ with 40% training data	73.47	71.42
QGA-EE$_{T5}$ with 20% training data	71.15	69.13

Table 5: Performance of QGA-EE on ACE05 test data (F1, %) with gold triggers with partial training data. Training data is randomly sampled.

4.5 Analysis and Discussion

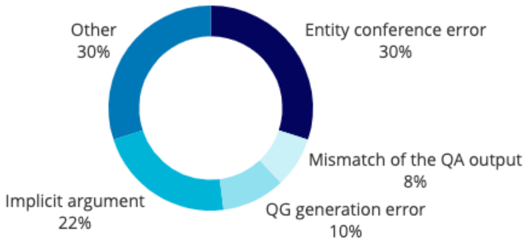

Figure 2: The portion of each category of error based on 50 error examples in test set.

The average length of the questions generated by QGA-EE$_{T5}$ is 10.5 tokens, compared with 6.7 in Du and Cardie (2020). They contain more context. For example, QGA-EE generates *'Who was attacked by mob in state?'* for the `Target` role in *'At least three members of a family in Indias northeastern state of Tripura were **[hacked$_{Conflict.Attack}$]** to death by a tribal mob for allegedly practicing witchcraft, police said Thursday.'* It incorporates `Attacker` ('mob') and `Place` ('state') information.

We categorize the errors into four groups:

1. Bad question generated by the QG model.

   For example, QGA-EE generates *'What did state buy in * sell * event?'* for the `Artifact` role in *'... that the Stalinist state had developed nuclear weapons and hinted it may sell or use them, depending on US actions.'*. It should have been *'What did state sell in * sell * event?'* and this introduces an error to the QA model.

2. Errors resulting from a mismatch of the QA output result. QGA-EE may retrieve wrong offsets if a target candidate matches with multiple text strings in the original sentence. For example, QGA-EE matches the candidate *'Welch'* with the first mention in *'He also wants to subpoena all documents maintained in Jane Beasley Welch's personnel file by Shearman; Sterling, a prestigious corporate law firm where she worked before she **[married$_{Life.Marry}$]** Welch.'*, where the correct one is the second mention.

3. Errors resulting from missing entity conference. For example, QGA-EE identifies *'Jacques Chirac'* as the `Entity` of the `Contact.Phone-Write` event in *'French President Jacques Chirac received only a reserved response when he tried to mend fences by placing a telephone call Tuesday to Bush.'*. But *'he'* is the ground truth and refers to *'Jacques Chirac'*.

4. Predictions not explicitly mentioned. For example, in *'Kelly, the US assistant secretary for East Asia and Pacific Affairs, arrived in Seoul from Beijing Friday to brief Yoon, the foreign minister.'*, QGA-EE infers *'Seoul'* as the `Place` of the `Contact.Meet` event, but it is not explicitly mentioned in the context, thus not covered by the gold annotations.

We manually analyzed a subset of the errors from the test set (50 examples), and show the portion of each category of error in Figure 2.

5 Conclusion

In this paper, we present QGA-EE, a novel sequence-to-sequence based framework for EE, which utilizes a QG model to generate context-aware questions as inputs to a QA model for EAE. Our model naturally supports the cases in which multiple event arguments play the same role within a specific event mention. We conduct experiments on the ACE05-E dataset and the proposed model outperforms all of the single-task-based models and achieves competitive results with state-of-the-art joint IE models. In the future, we plan to utilize the extensibility of the QA framework to incorporate knowledge from semi-structured event-relevant data such as Wikipedia Infoboxes. We also plan to extend our approach to multilingual EE and joint IE.

Limitations

The design of the dynamic templates requires knowledge of the event ontology and is time-consuming. The authors of the paper spent 30 hours designing the exclusive templates that cover all of the possible argument combinations for each argument role in ACE ontology. With a more complicated ontology, a much larger amount of time is required.

Another limitation of our approach is the offset retrieval method. If one sentence contains multiple mentions of the same entities, or even multiple text strings that have the same spellings but refer to different entities, the QGA-EE model always retrieves the position where the mention appears for the first time in the sentence as the offset of the extracted target. It may be improved by asking the model to generate contextual text as a position reference.

Acknowledgements

We thank our colleague Aoife Cahill and the anonymous reviewers for their constructive comments and suggestions.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.

Yunmo Chen, Tongfei Chen, Seth Ebner, Aaron Steven White, and Benjamin Van Durme. 2019. Reading the manual: Event extraction as definition comprehension. *arXiv preprint arXiv:1912.01586*.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*.

Xinya Du and Claire Cardie. 2020. Event extraction by answering (almost) natural questions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Rui Feng, Jie Yuan, and Chao Zhang. 2020. Probing and fine-tuning reading comprehension models for few-shot event extraction. *arXiv preprint arXiv:2010.11325*.

Elena Filatova and Vasileios Hatzivassiloglou. 2004. Event-based extractive summarization. In *Text Summarization Branches Out*.

I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. 2022. DEGREE: A data-efficient generation-based event extraction model. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Heng Ji and Ralph Grishman. 2011. Knowledge base population: Successful approaches and challenges. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu, and Yong Zhu. 2020. Event extraction as multi-turn question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2020*.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Sha Li, Heng Ji, and Jiawei Han. 2021. Document-level event argument extraction by conditional generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Di Lu, Clare Voss, Fangbo Tao, Xiang Ren, Rachel Guan, Rostyslav Korolov, Tongtao Zhang, Dongang Wang, Hongzhi Li, Taylor Cassidy, Heng Ji, Shih-fu Chang, Jiawei Han, William Wallace, James Hendler, Mei Si, and Lance Kaplan. 2016. Cross-media event extraction and recommendation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*.

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*.

Qing Lyu, Hongming Zhang, Elior Sulem, and Dan Roth. 2021. Zero-shot event extraction via transfer learning: Challenges and insights. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*.

Jie Ma, Shuai Wang, Rishita Anubhai, Miguel Ballesteros, and Yaser Al-Onaizan. 2020. Resource-enhanced neural model for event argument extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*.

Minh Van Nguyen, Viet Dac Lai, and Thien Huu Nguyen. 2021. Cross-task instance representation interactions and label dependencies for joint information extraction with graph convolutional networks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cicero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. 2021. Structured prediction as translation between augmented natural languages. In *9th International Conference on Learning Representations, ICLR 2021*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know what you don't know: Unanswerable questions for SQuAD. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.

Lei Sha, Feng Qian, Baobao Chang, and Zhifang Sui. 2018. Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Noam Shazeer and Mitchell Stern. 2018. Adafactor: Adaptive learning rates with sublinear memory cost. In *International Conference on Machine Learning*, pages 4596–4604. PMLR.

Minh Van Nguyen, Bonan Min, Franck Dernoncourt, and Thien Nguyen. 2022. Joint extraction of entities, relations, and events via modeling inter-instance and inter-label dependencies. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4363–4374.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

Zixuan Zhang and Heng Ji. 2021. Abstract meaning representation guided graph encoding and decoding for joint information extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 39–49.

Yang Zhou, Yubo Chen, Jun Zhao, Yin Wu, Jiexin Xu, and Jinlong Li. 2021. What the role is vs. what plays the role: Semi-supervised event argument extraction via dual question answering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 14638–14646.

A ACE05-E Data Preprocessing

We follow the preprocessing steps in Wadden et al. (2019) to preprocess ACE2005 corpora. More specifically, we use the preprocessing script at `https://github.com/dwadden/dygiepp`. In addition, we retrieve the character positions of the event triggers and arguments, because T5 uses a

SentencePiece tokenizer. Table 6 shows the statistics of the ACE05-E dataset.

Split	#Sents	#Events	#Arguments
Train	17,172	4,202	4,859
Dev	923	450	605
Test	832	403	576

Table 6: Data statistics of the ACE05-E dataset.

B Complete Dynamic Templates for ACE ontology

Table 12 shows the complete list of templates with different combinations of known argument roles for each ACE event argument role.

C Implementation Details

We use Huggingface Transformers library (Wolf et al., 2020) to load the model checkpoints.

C.1 Event Trigger Labeling Model

Hyperparamter	Value
Learning rate	3e-5
Learning rate decay	1e-5
Epoch	20
Batch size	4
Gradient accumulation steps	4

Table 7: Hyperparameter for Event Trigger Labeling Model training.

We implemented an ALBERT-based sequence labeling model for event trigger detection. We simply apply Softmax on top of the ALBERT encoder to predict the BIO schema based event label. We finetune the `albert-xxlarge-v2` checkpoint provided by Huggingface during training. [8]. We use the Adam optimizer with clip threshold of 1.0 and warmup proportion of 0.1. Table 7 shows the hyperparameter to train the Event Trigger Labeling Model.

C.2 QG model

When generating the groundtruth for QG model training, we use the basic template (e.g. 'Who was the attacking agent?') without incorporating any arguments if the target event role does not exist in the event mention. And we do not restrict the QG model to generate verbs that only appear in the templates. They are preserved for training the QA model.

Hyperparamter	Value
Learning rate	1e-4
Learning rate decay	1e-5
Epoch	20
Batch size	2
Gradient accumulation steps	32
Number of beam	4
Length penalty	0.0

Table 8: Hyperparameter for QG Model training.

We finetune the `T5-large` checkpoint provided by Huggingface during training. [9] with the Adafactor optimizer with clip threshold of 1.0 and warmup proportion of 0.1. Table 8 shows the hyperparameter to train the QG Model. And Table 9 shows the numbers of examples used to train and evaluate the QG model.

	Train	Dev	Test
QG Model	15,785	1,767	1,434
QA Model	20,681	1,713	1,391

Table 9: Number of examples used to train and evaluate the QG and QA models.

C.3 QA model

Hyperparamter	Value
Learning rate	2e-4
Learning rate decay	1e-5
Epoch	20
Batch size	2
Gradient accumulation steps	32
Number of beam	4
Length penalty	-2.5

Table 10: Hyperparameter for QA Model training.

For the QA model training, we use the Adafactor optimizer with a learning rate of 2e-4, and weight decay of 1e-5, and clip threshold as 1.0. We set all of the relative_step, scale_parameter, and warmup_init parameters to `False`. For optimizer scheduler, we set the warmup proportion to 0.1.

If there are no event arguments for the argument role, the output is empty, as the following example. We include them to train the QA model. Table 9

[8] https://huggingface.co/albert-xxlarge-v2

[9] https://huggingface.co/t5-large

shows the numbers of examples used to train and evaluate the QA model.

Input: *question: What device was used to inflict the harm in * injured * event? context: Injured Russian diplomats and a convoy of America's Kurdish comrades in arms were among unintended victims caught in crossfire and friendly fire Sunday.*
Output: < /s >

In postprocessing, we dynamically change the start position for searching to keep the order of the retrieved event arguments.

D Experiment Details

For all of the scores reported in the paper, the numbers are based on a single run with a fixed random seed 42.

D.1 Event Trigger Labeling Model

Table 11 shows the performance of the Event Trigger Labeling model on ACE05-E test set.

Trigger Identification			Trigger Classification		
P	R	F1	P	R	F1
72.52	79.9	76.03	69.59	76.67	72.96

Table 11: Performance of our event trigger labeling model on ACE05-E test data (%).

D.2 QG Model

We use Rouge (Lin, 2004) score (ROUGE-1) as the evaluation metric for QG model training, and the score on ACE05-E test set is 0.892.

Life.Be-Born	Person	- Place	Who was born? Who was born in [Place]?
	Place	- Person	Where did the birth take place? Where was [Person] born?
Life.Marry	Person	- Place	Who was married? Who was married in [Place]?
	Place	- Person	Where did the marriage take place? Where was [Person] married?
Life.Divorce	Person	- Place	Who was divorced? Who was divorced in [Place]?
	Place	- Person	Where did the divorce take place? Where was [Person] divorced?
Life.Injure	Agent	- Victim Instrument Place Victim, Instrument Victim, Place Instrument, Place Victim, Instrument, Place	Who enacted the harm? Who harmed [Victim]? Who enacted the harm using [Instrument]? Who enacted the harm in [Place]? Who harmed [Victim] using [Instrument]? Who harmed [Victim] in [Place]? Who enacted the harm using [Instrument] in [Place]? Who harmed [Victim] using [Instrument] in [Place]?
	Victim	- Agent Instrument Place Agent, Instrument Agent, Place Instrument, Place Agent, Instrument, Place	Who was harmed? Who was harmed by [Agent]? Who was harmed with [Instrument]? Who was harmed in [Place]? Who was harmed by [Agent] with [Instrument]? Who was harmed by [Agent] in [Place]? Who was harmed with [Instrument] in [Place]? Who was harmed by [Agent] with [Instrument] in [Place]?
	Instrument	- Agent Victim Place Agent, Victim Agent, Place Victim, Place Agent, Victim, Place	What device was used to inflict the harm? What device was used by [Agent] to inflict the harm? What device was used to harm [Victim]? What device was used to inflict the harm in [Place]? What device was used by [Agent] to harm [Victim]? What device was used by [Agent] to inflict the harm in [Place]? What device was used to harm [Victim] in [Place]? What device was used by [Agent] to harm [Victim] in [Place]?
	Place	- Agent Victim Instrument Agent, Victim Agent, Instrument Victim, Instrument Agent, Victim, Instrument	Where did the injuring take place? Where did [Agent] enact the harm? Where was [Victim] harmed? Where was [Instrument] used to inflict the harm? Where did [Agent] harm [Victim]? Where did [Agent] enact the harm with [Instrument]? Where was [Victim] harmed with [Instrument]? Where did [Agent] harm [Victim] with [Instrument]?
Life.Die	Agent	- Victim Instrument Place Victim, Instrument Victim, Place Instrument, Place Victim, Instrument, Place	Who was the killer? Who killed [Victim]? Who killed others using [Instrument]? Who killed others in [Place]? Who killed [Victim] using [Instrument]? Who killed [Victim] in [Place]? Who killed others using [Instrument] in [Place]? Who killed [Victim] using [Instrument] in [Place]?
	Victim	- Agent Instrument Place Agent, Instrument Agent, Place Instrument, Place Agent, Instrument, Place	Who was killed? Who was killed by [Agent]? Who was killed with [Instrument]? Who was killed in [Place]? Who was killed by [Agent] with [Instrument]? Who was killed by [Agent] in [Place]? Who was killed with [Instrument] in [Place]? Who was killed by [Agent] with [Instrument] in [Place]?

	Instrument	-	What device was used to kill?
		Agent	What device did [Agent] use to kill others?
		Victim	What device was used to kill [Victim]?
		Place	What device was used to kill others in [Place]?
		Agent, Victim	What device did [Agent] use to kill [Victim]?
		Agent, Place	What device did [Agent] use to kill others in [Place]?
		Victim, Place	What device was used to kill [Victim] in [Place]?
		Agent, Victim, Place	What device did [Agent] use to kill [Victim] in [Place]?
	Place	-	Where did the death take place?
		Agent	Where did [Agent] kill others?
		Victim	Where was [Victim] killed?
		Instrument	Where were people killed with [Instrument]?
		Agent, Victim	Where did [Agent] kill [Victim]?
		Agent, Instrument	Where did [Agent] kill others with [Instrument]?
		Victim, Instrument	Where was [Victim] killed with [Instrument]?
		Agent, Victim, Instrument	Where did [Agent] kill [Victim] with [Instrument]?
	Agent	-	Who is responsible for the transport event?
		Artifact	Who transported [Artifact]?
		Vehicle	Who transported artifact using [Vehicle]?
		Origin	Who transported artifact from [Origin]?
		Destination	Who transported artifact to [Destination]?
		Artifact, Vehicle	Who transported [Artifact] using [Vehicle]?
		Artifact, Origin	Who transported [Artifact] from [Origin]?
		Artifact, Destination	Who transported [Artifact] to [Destination]?
		Vehicle, Origin	Who transported artifact from [Origin] using [Vehicle]?
		Vehicle, Destination	Who transported artifact to [Destination] using [Vehicle]?
		Origin, Destination	Who transported artifact from [Origin] to [Destination]?
		Artifact, Vehicle, Origin	Who transported [Artifact] from [Origin] using [Vehicle]?
		Artifact, Vehicle, Destination	Who transported [Artifact] to [Destination] using [Vehicle]?
		Artifact, Origin, Destination	Who transported [Artifact] from [Origin] to [Destination]?
		Vehicle, Origin, Destination	Who transported artifact from [Origin] to [Destination] using [Vehicle]?
		Artifact, Vehicle, Origin, Destination	Who transported [Artifact] from [Origin] to [Destination] using [Vehicle]?
	Artifact	-	Who was transported?
		Agent	Who was transported by [Agent]?
		Vehicle	Who was transported with [Vehicle]?
		Origin	Who was transported from [Origin]?
		Destination	Who was transported to [Destination]?
		Agent, Vehicle	Who was transported by [Agent] with [Vehicle]?
		Agent, Origin	Who was transported from [Origin] by [Agent]?
		Agent, Destination	Who was transported to [Destination] by [Agent]?
		Vehicle, Origin	Who was transported from [Origin] with [Vehicle]?
		Vehicle, Destination	Who was transported to [Destination] with [Vehicle]?
		Origin, Destination	Who was transported from [Origin] to [Destination]?
		Agent, Vehicle, Origin	Who was transported from [Origin] by [Agent] with [Vehicle]?
		Agent, Vehicle, Destination	Who was transported to [Destination] by [Agent] with [Vehicle]?
		Agent, Origin, Destination	Who was transported from [Origin] to [Destination] by [Agent]?
		Vehicle, Origin, Destination	Who was transported from [Origin] to [Destination] with [Vehicle]?
		Agent, Vehicle, Origin, Destination	Who was transported from [Origin] to [Destination] by [Agent] with [Vehicle]?

Movement. Transport	Vehicle	-	What vehicle was used for transporting?
		Agent	What vehicle did [Agent] use for transporting?
		Artifact	What vehicle was used for transporting [Artifact]?
		Origin	What vehicle was used for transporting from [Origin]?
		Destination	What vehicle was used for transporting to [Destination]?
		Agent, Artifact	What vehicle did [Agent] use for transporting [Artifact]?
		Agent, Origin	What vehicle did [Agent] use for transporting from [Origin]?
		Agent, Destination	What vehicle did [Agent] use for transporting to [Destination]?
		Artifact, Origin	What vehicle was used for transporting [Artifact] from [Origin]?
		Artifact, Destination	What vehicle was used for transporting [Artifact] to [Destination]?
		Origin, Destination	What vehicle was used for transporting from [Origin] to [Destination]?
		Agent, Artifact, Origin	What vehicle did [Agent] use for transporting [Artifact] from [Origin]?
		Agent, Artifact, Destination	What vehicle did [Agent] use for transporting [Artifact] to [Destination]?
		Agent, Origin, Destination	What vehicle did [Agent] use for transporting from [Origin] to [Destination]?
		Artifact, Origin, Destination	What vehicle was used for transporting [Artifact] from [Origin] to [Destination]?
		Agent, Artifact, Origin, Destination	What vehicle did [Agent] use for transporting [Artifact] from [Origin] to [Destination]?
	Origin	-	Where did the transporting originate?
		Agent	Where did [Agent] transport artifact from?
		Artifact	Where was [Artifact] transported from?
		Vehicle	Where was artifact transported from with [Vehicle]?
		Destination	Where was artifact transported from to [Destination]?
		Agent, Artifact	Where did [Agent] transport [Artifact] from?
		Agent, Vehicle	Where did [Agent] transport artifact from with [Vehicle]?
		Agent, Destination	Where did [Agent] transport artifact from to [Destination]?
		Artifact, Vehicle	Where was [Artifact] transported from with [Vehicle]?
		Artifact, Destination	Where was [Artifact] transported from to [Destination]?
		Vehicle, Destination	Where was artifact transported from to [Destination] with [Vehicle]?
		Agent, Artifact, Vehicle	Where did [Agent] transport [Artifact] from with [Vehicle]?
		Agent, Artifact, Destination	Where did [Agent] transport [Artifact] from to [Destination]?
		Agent, Vehicle, Destination	Where did [Agent] transport artifact from to [Destination] with [Vehicle]?
		Artifact, Vehicle, Destination	Where was [Artifact] transported from to [Destination] with [Vehicle]?
		Agent, Artifact, Vehicle, Destination	Where did [Agent] transport [Artifact] from to [Destination] with [Vehicle]?
		-	Where was the transporting directed?
		Agent	Where did [Agent] transport artifact to?
		Artifact	Where was [Artifact] transported to?
		Vehicle	Where was artifact transported to with [Vehicle]?
		Origin	Where was artifact transported to from [Origin]?
		Agent, Artifact	Where did [Agent] transport [Artifact] to?
		Agent, Vehicle	Where did [Agent] transport artifact to with [Vehicle]?
		Agent, Origin	Where did [Agent] transport artifact to from [Origin]?

	Destination	Artifact, Vehicle	Where was [Artifact] transported to with [Vehicle]?
		Artifact, Origin	Where was [Artifact] transported to from [Origin]?
		Vehicle, Origin	Where was artifact transported to from [Origin] with [Vehicle]?
		Agent, Artifact, Vehicle	Where did [Agent] transport [Artifact] to with [Vehicle]?
		Agent, Artifact, Origin	Where did [Agent] transport [Artifact] to from [Origin]?
		Agent, Vehicle, Origin	Where did [Agent] transport artifact to from [Origin] with [Vehicle]?
		Artifact, Vehicle, Origin	Where was [Artifact] transported to from [Origin] with [Vehicle]?
		Agent, Artifact, Vehicle, Origin	Where did [Agent] transport [Artifact] to from [Origin] with [Vehicle]?
Transaction. Transfer -Ownership	Buyer	-	Who is the buying agent?
		Seller	Who bought things from [Seller]?
		Beneficiary	Who bought things for [Beneficiary]?
		Artifact	Who bought [Artifact]?
		Place	Who bought things in [Place]?
		Seller, Beneficiary	Who bought things from [Seller] for [Beneficiary]?
		Seller, Artifact	Who bought [Artifact] from [Seller]?
		Seller, Place	Who bought things from [Seller] in [Place]?
		Beneficiary, Artifact	Who bought [Artifact] for [Beneficiary]?
		Beneficiary, Place	Who bought things for [Beneficiary] in [Place]?
		Artifact, Place	Who bought [Artifact] in [Place]?
		Seller, Beneficiary, Artifact	Who bought [Artifact] from [Seller] for [Beneficiary]?
		Seller, Beneficiary, Place	Who bought things from [Seller] for [Beneficiary] in [Place]?
		Seller, Artifact, Place	Who bought [Artifact] from [Seller] in [Place]?
		Beneficiary, Artifact, Place	Who bought [Artifact] for [Beneficiary] in [Place]?
		Seller, Beneficiary, Artifact, Place	Who bought [Artifact] from [Seller] for [Beneficiary] in [Place]?
	Seller	-	Who is the selling agent?
		Buyer	Who sold things to [Buyer]?
		Beneficiary	Who did buyer buy things from for [Beneficiary]?
		Artifact	Who sold [Artifact]?
		Place	Who sold things in [Place]?
		Buyer, Beneficiary	Who did [Buyer] buy things from for [Beneficiary]?
		Buyer, Artifact	Who sold [Artifact] to [Buyer]?
		Buyer, Place	Who sold things to [Buyer] in [Place]?
		Beneficiary, Artifact	Who did buyer buy [Artifact] from for [Beneficiary]?
		Beneficiary, Place	Who did buyer buy things from for [Beneficiary] in [Place]?
		Artifact, Place	Who sold [Artifact] in [Place]?
		Buyer, Beneficiary, Artifact	Who did [Buyer] buy [Artifact] from for [Beneficiary]?
		Buyer, Beneficiary, Place	Who did [Buyer] buy things from for [Beneficiary] in [Place]?
		Buyer, Artifact, Place	Who sold [Artifact] to [Buyer] in [Place]?
		Beneficiary, Artifact, Place	Who did buyer buy [Artifact] from for [Beneficiary] in [Place]?
		Buyer, Beneficiary, Artifact, Place	Who did [Buyer] buy [Artifact] from for [Beneficiary] in [Place]?

Beneficiary	-	Who benefits from the transaction?
	Buyer	Who did [Buyer] buy things for?
	Seller	Who did buyer buy things from [Seller] for?
	Artifact	Who did buyer buy [Artifact] for?
	Place	Who did buyer buy things for in [Place]?
	Buyer, Seller	Who did [Buyer] buy things from [Seller] for?
	Buyer, Artifact	Who did [Buyer] buy [Artifact] for?
	Buyer, Place	Who did [Buyer] buy things for in [Place]?
	Seller, Artifact	Who did buyer buy [Artifact] from [Seller] for?
	Seller, Place	Who did buyer buy things from [Seller] for in [Place]?
	Artifact, Place	Who did buyer buy [Artifact] for in [Place]?
	Buyer, Seller, Artifact	Who did [Buyer] buy [Artifact] from [Seller] for?
	Buyer, Seller, Place	Who did [Buyer] buy things from [Seller] for in [Place]?
	Buyer, Artifact, Place	Who did [Buyer] buy [Artifact] for in [Place]?
	Seller, Artifact, Place	Who did buyer buy [Artifact] from [Seller] for in [Place]?
	Buyer, Seller, Artifact, Place	Who did [Buyer] buy [Artifact] from [Seller] for in [Place]?
Artifact	-	What was bought?
	Buyer	What did [Buyer] buy?
	Seller	What did [Seller] sell?
	Beneficiary	What was bought for [Beneficiary]?
	Place	What was bought in [Place]?
	Buyer, Seller	What did [Buyer] buy from [Seller]?
	Buyer, Beneficiary	What did [Buyer] buy for [Beneficiary]?
	Buyer, Place	What did [Buyer] buy in [Place]?
	Seller, Beneficiary	What did buyer buy from [Seller] for [Beneficiary]?
	Seller, Place	What did [Seller] sell in [Place]?
	Beneficiary, Place	What was bought for [Beneficiary] in [Place]?
	Buyer, Seller, Beneficiary	What did [Buyer] buy from [Seller] for [Beneficiary]?
	Buyer, Seller, Place	What did [Buyer] buy from [Seller] in [Place]?
	Buyer, Beneficiary, Place	What did [Buyer] buy for [Beneficiary] in [Place]?
	Seller, Beneficiary, Place	What did buyer buy from [Seller] for [Beneficiary] in [Place]?
	Buyer, Seller, Beneficiary, Place	What did [Buyer] buy from [Seller] for [Beneficiary] in [Place]?
Place	-	Where did the sale take place?
	Buyer	Where did [Buyer] buy things?
	Seller	Where did [Seller] sell things?
	Beneficiary	Where did buyer buy things for [Beneficiary]?
	Artifact	Where did buyer buy [Artifact]?
	Buyer, Seller	Where did [Buyer] buy things from [Seller]?
	Buyer, Beneficiary	Where did [Buyer] buy things for [Beneficiary]?
	Buyer, Artifact	Where did [Buyer] buy [Artifact]?
	Seller, Beneficiary	Where did buyer buy things for [Beneficiary] from [Seller]?
	Seller, Artifact	Where did buyer buy [Artifact] from [Seller]?
	Beneficiary, Artifact	Where did buyer buy [Artifact] for [Beneficiary]?
	Buyer, Seller, Beneficiary	Where did [Buyer] buy things from [Seller] for [Beneficiary]?
	Buyer, Seller, Artifact	Where did [Buyer] buy [Artifact] from [Seller]?
	Buyer, Beneficiary, Artifact	Where did [Buyer] buy [Artifact] for [Beneficiary]?
	Seller, Beneficiary, Artifact	Where did buyer buy [Artifact] for [Beneficiary] from [Seller]?
	Buyer, Seller, Beneficiary, Artifact	Where did [Buyer] buy [Artifact] from [Seller] for [Beneficiary]?

Transaction. Transfer-Money	Giver	-	Who gave money to others?
		Recipient	Who gave money to [Recipient]?
		Beneficiary	Who gave money to others for [Beneficiary]?
		Place	Who gave money to others in [Place]?
		Recipient, Beneficiary	Who gave money to [Recipient] for [Beneficiary]?
		Recipient, Place	Who gave money to [Recipient] in [Place]?
		Beneficiary, Place	Who gave money to others for [Beneficiary] in [Place]?
		Recipient, Beneficiary, Place	Who gave money to [Recipient] for [Beneficiary] in [Place]?
	Recipient	-	Who was given money?
		Giver	Who did [Giver] give money to?
		Beneficiary	Who was given money for [Beneficiary]?
		Place	Who was given money in [Place]?
		Giver, Beneficiary	Who did [Giver] give money to for [Beneficiary]?
		Giver, Place	Who did [Giver] give money to in [Place]?
		Beneficiary, Place	Who was given money for [Beneficiary] in [Place]?
		Giver, Beneficiary, Place	Who did [Giver] give money to for [Beneficiary] in [Place]?
	Beneficiary	-	Who benefited from the transfer?
		Giver	Who did [Giver] give money for?
		Recipient	Who was [Recipient] given money for?
		Place	Who benefited from the transfer in [Place]?
		Giver, Recipient	Who did [Giver] give money to [Recipient] for?
		Giver, Place	Who did [Giver] give money for in [Place]?
		Recipient, Place	Who was [Recipient] given money for in [Place]?
		Giver, Recipient, Place	Who did [Giver] give money to [Recipient] for in [Place]?
	Place	-	Where was the amount transferred?
		Giver	Where did [Giver] give money to others?
		Recipient	Where was [Recipient] given money?
		Beneficiary	Where did giver give money for [Beneficiary]?
		Giver, Recipient	Where did [Giver] give money to [Recipient]?
		Giver, Beneficiary	Where did [Giver] give money to others for [Beneficiary]?
		Recipient, Beneficiary	Where was [Recipient] given money for [Beneficiary]?
		Giver, Recipient, Beneficiary	Where did [Giver] give money to [Recipient] for [Beneficiary]?
Business. Start-Org	Agent	-	Who started the organization?
		Org	Who started [Org]?
		Place	Who started the organization in [Place]?
		Org, Place	Who started [Org] in [Place]?
	Org	-	What organization was started?
		Agent	What organization was started by [Agent]?
		Place	What organization was started in [Place]?
		Agent, Place	What organization was started by [Agent] in [Place]?
	Place	-	Where was the organization started?
		Agent	Where was the organization started by [Agent]?
		Org	Where was [Org] started?
		Agent, Org	Where was [Org] started by [Agent]?
Business. Merge-Org	Org	-	What organization was merged?
Business. Declare-Bankruptcy	Org	-	What organization declared bankruptcy?
		Place	What organization declared bankruptcy in [Place]?
	Place	-	Where was the bankruptcy declared?
		Org	Where did [Org] declare the bankruptcy?
Business. End-Org	Org	-	What organization was ended?
		Place	What organization was ended in [Place]?
	Place	-	Where was the organization ended?
		Org	Where was [Org] ended?

Conflict. Attack	Attacker	-	Who was the attacking agent?
		Target	Who attacked [Target]?
		Instrument	Who used [Instrument] in the attack?
		Place	Who made the attack in [Place]?
		Target, Instrument	Who attacked [Target] using [Instrument]?
		Target, Place	Who attacked [Target] in [Place]?
		Instrument, Place	Who used [Instrument] in the attack in [Place]?
		Target, Instrument, Place	Who attacked [Target] using [Instrument] in [Place]?
	Target	-	Who was the target of the attack?
		Attacker	Who was attacked by [Attacker]?
		Instrument	Who was attacked with [Instrument]?
		Place	Who was the target of the attack in [Place]?
		Attacker, Instrument	Who was attacked by [Attacker] using [Instrument]?
		Attacker, Place	Who was attacked by [Attacker] in [Place]?
		Instrument, Place	Who was attacked with [Instrument] in [Place]?
		Attacker, Instrument, Place	Who was attacked by [Attacker] using [Instrument] in [Place]?
	Instrument	-	What instrument was used in the attack?
		Attacker	What instrument did [Attacker] use in the attack?
		Target	What instrument was used to attack [Target]?
		Place	What instrument was used in the attack in [Place]?
		Attacker, Target	What instrument did [Attacker] use to attack [Target]?
		Attacker, Place	What instrument did [Attacker] use in the attack in [Place]?
		Target, Place	What instrument was used to attack [Target] in [Place]?
		Attacker, Target, Place	What instrument did [Attacker] use to attack [Target] in [Place]?
	Place	-	Where did the attack take place?
		Attacker	Where did [Attacker] make an attack?
		Target	Where was [Target] attacked?
		Instrument	Where was [Instrument] used in the attack?
		Attacker, Target	Where did [Attacker] attack [Target]?
		Attacker, Instrument	Where did [Attacker] use [Instrument] to make an attack?
		Target, Instrument	Where was [Instrument] used to attack [Target]?
		Attacker, Target, Instrument	Where did [Attacker] attack [Target] using [Instrument]?
Conflict. Demonstrate	Entity	-	Who demonstrated?
		Place	Who demonstrated in [Place]?
	Place	-	Where did the demonstration take place?
		Entity	Where did [Entity] demonstrate?
Contact.Meet	Entity	-	Who met with others?
		Place	Who met others in [Place]?
	Place	-	Where did the meeting takes place?
		Entity	Where did [Entity] meet others?
Contact. Phone-Write	Entity	-	Who communicated with others?
Personnel. Start-Position	Person	-	Who is the employee?
		Entity	Who was hired by [Entity]?
		Place	Who was hired in [Place]?
		Entity, Place	Who was hired by [Entity] in [Place]?
	Entity	-	Who is the the employer?
		Person	Who hired [Person]?
		Place	Who hired employee in [Place]?
		Person, Place	Who hired [Person] in [Place]?
	Place	-	Where did the employment relationship begin?
		Person	Where was [Person] hired?
		Entity	Where did [Entity] hire employee?
		Person, Entity	Where did [Entity] hire [Person]?

Personnel. End-Position	Person	-	Who ended the position?
		Entity	Who was fired by [Entity]?
		Place	Who ended the position in [Place]?
		Entity, Place	Who was fired by [Entity] in [Place]?
	Entity	-	Who fired employee?
		Person	Who fired [Person]?
		Place	Who fired employee in [Place]?
		Person, Place	Who fired [Person] in [Place]?
	Place	-	Where did the employment relationship end?
		Person	Where did [Person] end the position?
		Entity	Where did [Entity] fire employee?
		Person, Entity	Where did [Entity] fire [Person]?
Personnel. Nominate	Person	-	Who was nominated?
		Agent	Who was nominated by [Agent]?
	Agent	-	Who is the nominating agent?
		Person	Who nominated [Person]?
Personnel. Elect	Person	-	Who was elected?
		Agent	Who was elected by [Agent]?
		Place	Who was elected in [Place]?
		Agent, Place	Who was elected by [Agent] in [Place]?
	Agent	-	Who was the voting agent?
		Person	Who elected [Person]?
		Place	Who elected person in [Place]?
		Person, Place	Who elected [Person] in [Place]?
	Place	-	Where did the election takes place?
		Person	Where was [Person] elected?
		Agent	Where did [Agent] elect person?
		Person, Agent	Where did [Agent] elect [Person]?
Justice. Arrest-Jail	Person	-	Who was arrested?
		Agent	Who was arrested by [Agent]?
		Place	Who was arrested in [Place]?
		Agent, Place	Who was arrested by [Agent] in [Place]?
	Agent	-	Who made the arrest?
		Person	Who arrested [Person]?
		Place	Who made the arrest in [Place]?
		Person, Place	Who arrested [Person] in [Place]?
	Place	-	Where did the arrest take place?
		Person	Where was [Person] arrested?
		Agent	Where did [Agent] arrest person?
		Person, Agent	Where did [Agent] arrest [Person]?
Justice. Release-Parole	Person	-	Who was released?
		Entity	Who was released by [Entity]?
		Place	Who was released in [Place]?
		Entity, Place	Who was released by [Entity] in [Place]?
	Entity	-	Who released the person?
		Person	Who released [Person]?
		Place	Who released the person in [Place]?
		Person, Place	Who released [Person] in [Place]?
	Place	-	Where did the release take place?
		Person	Where was [Person] released?
		Entity	Where did [Entity] release person?
		Person, Entity	Where did [Entity] release [Person]?
	Defendant	-	Who was on trial?
		Prosecutor	Who was on trial being prosecuted by [Prosecutor]?
		Adjudicator	Who was on trial being adjudicated by [Adjudicator]?
		Place	Who was on trial in [Place]?
		Prosecutor, Adjudicator	Who was tried by [Prosecutor] with being adjudicated by [Adjudicator]?
		Prosecutor, Place	Who was tried by [Prosecutor] in [Place]?
		Adjudicator, Place	Who was on trial being adjudicated by [Adjudicator] in [Place]?
		Prosecutor, Adjudicator, Place	Who was tried by [Prosecutor] with being adjudicated by [Adjudicator] in [Place]?

Justice. Trial-Hearing	Prosecutor	-	Who tried defendant?
		Defendant	Who tried [Defendant]?
		Adjudicator	Who tried the defendant being adjudicated by [Adjudicator]?
		Place	Who tried defendant in [Place]?
		Defendant, Adjudicator	Who tried [Defendant] being adjudicated by [Adjudicator]?
		Defendant, Place	Who tried [Defendant] in [Place]?
		Adjudicator, Place	Who tried the defendant being adjudicated by [Adjudicator] in [Place]?
		Defendant, Adjudicator, Place	Who tried [Defendant] being adjudicated by [Adjudicator] in [Place]?
	Adjudicator	-	Who adjudicated the trial?
		Defendant	Who adjudicated the trial [Defendant] was on?
		Prosecutor	Who adjudicated the trial being prosecuted by [Prosecutor]?
		Place	Who adjudicated the trial in [Place]?
		Defendant, Prosecutor	Who adjudicated the trial [Defendant] was on being prosecuted by [Prosecutor]?
		Defendant, Place	Who adjudicated the trial [Defendant] was on in [Place]?
		Prosecutor, Place	Who adjudicated the trial being prosecuted by [Prosecutor] in [Place]?
		Defendant, Prosecutor, Place	Who adjudicated the trial [Defendant] was on being prosecuted by [Prosecutor] in [Place]?
	Place	-	Where did the trial take place?
		Defendant	Where was [Defendant] tried?
		Prosecutor	Where did [Prosecutor] try the defendant?
		Adjudicator	Where did [Adjudicator] adjudicate the trial?
		Defendant, Prosecutor	Where did [Prosecutor] try [Defendant]?
		Defendant, Adjudicator	Where did [Adjudicator] adjudicate the trial [Defendant] was on?
		Prosecutor, Adjudicator	Where did [Prosecutor] try the defendant with being adjudicated by [Adjudicator]?
		Defendant, Prosecutor, Adjudicator	Where did [Prosecutor] try [Defendant] with being adjudicated by [Adjudicator]?
	Defendant	-	Who was indicated for crime?
		Prosecutor	Who was indicated for crime by [Prosecutor]?
		Adjudicator	Who was indicated for crime being adjudicated by [Adjudicator]?
		Place	Who was indicated for crime in [Place]?
		Prosecutor, Adjudicator	Who was indicated for crime by [Prosecutor] being adjudicated by [Adjudicator]?
		Prosecutor, Place	Who was indicated for crime by [Prosecutor] in [Place]?
		Adjudicator, Place	Who was indicated for crime being adjudicated by [Adjudicator] in [Place]?
		Prosecutor, Adjudicator, Place	Who was indicated for crime by [Prosecutor] being adjudicated by [Adjudicator] in [Place]?
	Prosecutor	-	Who executed the indictment?
		Defendant	Who indicated [Defendant] for crime?
		Adjudicator	Who executed the indictment being adjudicated by [Adjudicator]?
		Place	Who executed the indictment in [Place]?
		Defendant, Adjudicator	Who indicated [Defendant] for crime being adjudicated by [Adjudicator]?
		Defendant, Place	Who indicated [Defendant] for crime in [Place]?
		Adjudicator, Place	Who executed the indictment being adjudicated by [Adjudicator] in [Place]?
		Defendant, Adjudicator, Place	Who indicated [Defendant] for crime being adjudicated by [Adjudicator] in [Place]?

Justice. Charge-Indict	Adjudicator	-	Who adjudicated the indictment?
		Defendant	Who adjudicated the indictment [Defendant] was charged in?
		Prosecutor	Who adjudicated the indictment executed by [Prosecutor]?
		Place	Who adjudicated the indictment in [Place]?
		Defendant, Prosecutor	Who adjudicated the indictment [Defendant] was charged in by [Prosecutor]?
		Defendant, Place	Who adjudicated the indictment [Defendant] was charged in in [Place]?
		Prosecutor, Place	Who adjudicated the indictment executed by [Prosecutor] in [Place]?
		Defendant, Prosecutor, Place	Who adjudicated the indictment [Defendant] was charged in by [Prosecutor] in [Place]?
	Place	-	Where did the indictment take place?
		Defendant	Where was [Defendant] indicated?
		Prosecutor	Where did [Prosecutor] execute the indictment?
		Adjudicator	Where did [Adjudicator] adjudicate the indictment?
		Defendant, Prosecutor	Where did [Prosecutor] indicate [Defendant] for crime?
		Defendant, Adjudicator	Where was [Defendant] indicated for crime being adjudicated by [Adjudicator]?
		Prosecutor, Adjudicator	Where did [Prosecutor] execute the indictment being adjudicated by [Adjudicator]?
		Defendant, Prosecutor, Adjudicator	Where did [Prosecutor] indicate [Defendant] for crime being adjudicated by [Adjudicator]?
Justice.Sue	Plaintiff	-	Who sued defendant?
		Defendant	Who sued [Defendant]?
		Adjudicator	Who sued defendant being adjudicated by [Adjudicator]?
		Place	Who sued defendant in [Place]?
		Defendant, Adjudicator	Who sued [Defendant] being adjudicated by [Adjudicator]?
		Defendant, Place	Who sued [Defendant] in [Place]?
		Adjudicator, Place	Who sued defendant being adjudicated by [Adjudicator] in [Place]?
		Defendant, Adjudicator, Place	Who sued [Defendant] being adjudicated by [Adjudicator] in [Place]?
	Defendant	-	Who was sued?
		Plaintiff	Who was sued by [Plaintiff]?
		Adjudicator	Who was sued for crime being adjudicated by [Adjudicator]?
		Place	Who was sued in [Place]?
		Plaintiff, Adjudicator	Who was sued by [Plaintiff] for crime being adjudicated by [Adjudicator]?
		Plaintiff, Place	Who was sued by [Plaintiff] in [Place]?
		Adjudicator, Place	Who was sued for crime being adjudicated by [Adjudicator] in [Place]?
		Plaintiff, Adjudicator, Place	Who was sued by [Plaintiff] for crime being adjudicated by [Adjudicator] in [Place]?
	Adjudicator	-	Who adjudicated the suing?
		Plaintiff	Who adjudicated the suing made by [Plaintiff]?
		Defendant	Who adjudicated the suing against [Defendant]?
		Place	Who adjudicated the suing in [Place]?
		Plaintiff, Defendant	Who adjudicated the suing against [Defendant] made by [Plaintiff]?
		Plaintiff, Place	Who adjudicated the suing made by [Plaintiff] in [Place]?
		Defendant, Place	Who adjudicated the suing against [Defendant] in [Place]?
		Plaintiff, Defendant, Place	Who adjudicated the suing against [Defendant] made by [Plaintiff] in [Place]?
		-	Where did the suit take place?
		Plaintiff	Where did [Plaintiff] sue defendant?
		Defendant	Where was [Defendant] sued?
		Adjudicator	Where did [Adjudicator] adjudicate the suing?

	Place	Plaintiff, Defendant Plaintiff, Adjudicator Defendant, Adjudicator Plaintiff, Defendant, Adjudicator	Where did [Plaintiff] sue [Defendant]? Where did [Plaintiff] sue defendant being adjudicated by [Adjudicator]? Where was [Defendant] sued being adjudicated by [Adjudicator]? Where did [Plaintiff] sue [Defendant] being adjudicated by [Adjudicator]?
Justice. Convict	Defendant	- Adjudicator Place Adjudicator, Place	Who was convicted for crime? Who was convicted by [Adjudicator] for crime? Who was convicted for crime in [Place]? Who was convicted by [Adjudicator] for crime in [Place]?
	Adjudicator	- Defendant Place Defendant, Place	Who convicted defendant for crime? Who convicted [Defendant] for crime? Who convicted defendant for crime in [Place]? Who convicted [Defendant] for crime in [Place]?
	Place	- Defendant Adjudicator Defendant, Adjudicator	Where did the conviction take place? Where was [Defendant] convicted for crime? Where did [Adjudicator] convict the defendant for crime? Where did [Adjudicator] convict [Defendant] for crime?
Justice. Sentence	Defendant	- Adjudicator Place Adjudicator, Place	Who was sentenced for crime? Who was sentenced by [Adjudicator] for crime? Who was sentenced for crime in [Place]? Who was sentenced by [Adjudicator] for crime in [Place]?
	Adjudicator	- Defendant Place Defendant, Place	Who sentenced the defendant for crime? Who sentenced [Defendant] for crime? Who sentenced the defendant for crime in [Place]? Who sentenced [Defendant] for crime in [Place]?
	Place	- Defendant Adjudicator Defendant, Adjudicator	Where did the sentencing take place? Where was [Defendant] sentenced for crime? Where did [Adjudicator] sentence the defendant for crime? Where did [Adjudicator] sentence [Defendant] for crime?
Justice.Fine	Entity	- Adjudicator Place Adjudicator, Place	Who was fined for crime? Who was fined by [Adjudicator] for crime? Who was fined for crime in [Place]? Who was fined by [Adjudicator] for crime in [Place]?
	Adjudicator	- Entity Place Entity, Place	Who fined the entity for crime? Who fined [Entity] for crime? Who fined the entity for crime in [Place]? Who fined [Entity] for crime in [Place]?
	Place	- Entity Adjudicator Entity, Adjudicator	Where did the fining take place? Where was [Entity] fined for crime? Where did [Adjudicator] fine the entity for crime? Where did [Adjudicator] fine [Entity] for crime?
Justice. Execute	Person	- Agent Place Agent, Place	Who was executed for crime? Who was executed by [Agent] for crime? Who was executed for crime in [Place]? Who was executed by [Agent] for crime in [Place]?
	Agent	- Person Place Person, Place	Who executed person for crime? Who executed [Person] for crime? Who executed person for crime in [Place]? Who executed [Person] for crime in [Place]?
	Place	- Person Agent Person, Agent	Where did the execution take place? Where was [Person] executed for crime? Where did [Agent] execute person for crime? Where did [Agent] execute [Person] for crime?
	Agent	- Destination Origin Destination, Origin	Who extradited person? Who extradited person to [Destination]? Who extradited person from [Origin]? Who extradited person from [Origin] to [Destination]?

Justice. Extradite	Destination	- Agent Origin Agent, Origin	Where was the person extradited to? Where did [Agent] extradite person to? Where was the person extradited to from [Origin]? Where did [Agent] extradite person to from [Origin]?
	Origin	- Agent Destination Agent, Destination	Where was the person extradited from? Where did [Agent] extradite person from? Where was the person extradited from to [Destination]? Where did [Agent] extradite person from to [Destination]?
Justice. Acquit	Defendant	- Adjudicator	Who was acquited of crime? Who was acquited of crime by [Adjudicator]?
	Adjudicator	- Defendant	Who acquited the defendant of crime? Who acquited [Defendant] of crime?
Justice. Pardon	Defendant	- Adjudicator Place Adjudicator, Place	Who was pardoned for crime? Who was pardoned by [Adjudicator] for crime? Who was pardoned for crime in [Place]? Who was pardoned by [Adjudicator] for crime in [Place]?
	Adjudicator	- Defendant Place Defendant, Place	Who pardoned defendant for crime? Who pardoned [Defendant] for crime? Who pardoned defendant for crime in [Place]? Who pardoned [Defendant] for crime in [Place]?
	Place	- Defendant Adjudicator Defendant, Adjudicator	Where did the pardon take place? Where was [Defendant] pardoned for crime? Where did [Adjudicator] pardon the defendant for crime? Where did [Adjudicator] pardon [Defendant] for crime?
Justice. Appeal	Defendant	- Adjudicator Place Adjudicator, Place	Who made the appeal? Who made the appeal to [Adjudicator]? Who made the appeal in [Place]? Who made the appeal to [Adjudicator] in [Place]?
	Adjudicator	- Defendant Place Defendant, Place	Who adjudicated the appeal? Who adjudicated the appeal made by [Defendant]? Who adjudicated the appeal in [Place]? Who adjudicated the appeal made by [Defendant] in [Place]?
	Place	- Defendant Adjudicator Defendant, Adjudicator	Where did the appeal take place? Where did [Defendant] make the appeal? Where did [Adjudicator] adjudicate the appeal? Where did [Defendant] make the appeal to [Adjudicator]?

Table 12: Complete Templates for argument roles in ACE ontology.

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
Section 4.5 and Limitation Section

☒ A2. Did you discuss any potential risks of your work?
the work is for foundational research and experiments are conducted on public dataset

☑ A3. Do the abstract and introduction summarize the paper's main claims?
Abstract and Section 1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

section 3.1, section 3.2, section 4.1, section 4.2

☑ B1. Did you cite the creators of artifacts you used?
section 3.1, section 3.2, section 4.1, section 4.2

☒ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
ACE 2005 Corpus was released by Linguistic Data Consortium with license: LDC User Agreement for Non-Members The code of our work will be released on Github with license TBD.

☒ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
ACE 2005 Corpus is for research purposes only Our work is for research purposes only

☒ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
The task itself is designed for identifying person/organization involved in events. The dataset contains violent events such as attack, die, and injure.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
section 3.1 and appendix

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
appendix

C ☑ Did you run computational experiments?

section 4

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
section 3.3 and appendix

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
section 3 and appendix

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
section 4 and appendix

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
appendix

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?
Left blank.

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
No response.

☐ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
No response.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
No response.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No response.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
No response.