BART-light: One Decoder Layer Is Enough

Anonymous ACL submission

Abstract

BART (Lewis et al., 2020), an encoder-decoder transformer language model (LM), has reached state-of-the-art results on several tasks in natural language generation and understanding. Similar to other pretrained encoder-decoder LMs, it uses the same number of hidden layers in the encoder and the decoder. In this paper, we show that one can easily remove all but one or two decoder layers for text generation tasks and even remove the whole decoder for classification tasks, with little to no compromises on performance. Our study presents that a shallow decoder is sufficient for most tasks when a deep encoder is used.

1 Introduction

Encoder-decoder language models (LMs) with the transformer architecture (Vaswani et al., 2017), such as BART (Lewis et al., 2020), became state of the art for many sequence-to-sequence (seq2seq) generation and sequence classification tasks. At the same time, pretrained LMs have been growing in size (Sanh et al., 2019), leading to several issues. First, adoption of these models is often infeasible due to hardware requirements or high latency. Second, rising computational cost poses questions on the environmental fallout (Strubell et al., 2019; Schwartz et al., 2019). Finally, we still lack full understanding of how the models work and whether their size is required and justifiable.

The model size issue has been explored with knowledge distillation (KD) (Bucilå et al., 2006), which was adapted to compress large neural-network (NN) models (Hinton et al., 2015) and later applied to transformers (Sanh et al., 2019; Sun et al., 2020; Jiao et al., 2020). Nonetheless, there has been little specific focus on compressing the decoder, which is a crucial question in text generation tasks. Another research branch focused on the analysis and pruning of transformer's attention (Voita et al., 2019; Michel et al., 2019), however, these have been confined only to the encoder.

In this work, we focus on compressing the decoder with an additional analysis of decoder attention heads. Our contributions are as follows:

- We present a simple compression technique for pretrained encoder-decoder LMs via removing most decoder layers or even the whole decoder (for classification tasks).
- We demonstrate the technique on BART, showing that many variants of our pruned BART-light are not significantly worse (at the 95% level) than the original BART-*base* on multiple downstream tasks (summarization, dialogue response generation, classification). The approach also brings performance benefits: The pruned BART-light with one decoder layer is up to 1.36× faster and 1.51× smaller than original BART.
- The visualization of the attention head weights demonstrates that the BART-light uses its attention layers more efficiently than the BART-*base*.

We will publish our source code on GitHub repository with the final version of this paper.

2 Background and Related Work

Pretrained LMs have reached state-of-the-art results on a whole range of natural language tasks. Earlier encoder-only LMs such as BERT (Devlin et al., 2018) and decoder-only LMs such as GPT (Radford et al., 2018) have been unified by encoder-decoder transformers pioneered by BART (Lewis et al., 2020), which we use as a base for our work. BART is pretrained as a denoising autoencoder with the task of recovering a sentence corrupted by functions such as word shuffling or removal.

Due to increasing model sizes, a large body of research has focused on the area of KD (Bucilå et al., 2006) to compress large NN models (Hinton et al., 2015). Recently, Sanh et al. (2019), Jiao et al. (2020), Sun et al. (2020) or Kim and

Model	Layers	Params	XSum R1	XSum R2	XSum RL	CNN/DailyMail R1	CNN/DailyMail R2	CNN/DailyMail RL
BERTSUMABS (Liu and Lapata, 2019)	6-6	110M	38.76	16.33	31.15	41.72	19.39	38.76
BERTSUMEXTABS (Liu and Lapata, 2019)	6-6	110M	38.81	16.50	31.27	42.13	19.60	39.18
BART-Large (Lewis et al., 2020)	12-12	406M	45.14	22.27	37.25	44.16	21.28	40.90
DistilBART (Shleifer and Rush, 2020)	12-1	222M	✗	17.98	33.31	✗	✗	✗
BART-base (ours)	6-6	139M	**39.96**	**18.31**	**32.78**	**40.23**	18.80	37.56
BART-light-top+bottom	6-2	102M	39.04	17.52	32.07	39.71	18.61	37.22
+ init			39.90*	18.19*	32.71*	40.05*	**18.92***	**37.59***
+ KD			37.97	16.63	31.16	39.35	18.24	36.83
+ KD + init			39.48	18.07	32.59	39.94	18.85	37.49
BART-light-top	6-1	92M	37.76	16.99	31.44	38.85	18.00	36.51
+ init			38.78	17.74	32.22	39.56	18.64*	37.35
+ KD			36.03	15.60	29.98	38.30	17.48	35.95
+ KD + init			37.16	16.68	31.14	39.00	18.24	36.72
BART-light-bottom	6-1	92M	36.78	15.97	30.44	38.82	17.78	36.48
+ init			38.14	16.98	31.54	39.15	18.05	36.82
+ KD			32.76	12.52	26.95	36.47	15.60	34.28
+ KD + init			32.72	12.57	26.95	37.52	16.57	35.31

Table 1: Results for summarization on XSum and CNN/DailyMail. We used paired bootstrap resampling (Koehn, 2004) to test significance of performance drop for BART-light-top+bottom (+init) and BART-light-top (+init) compared to BART-base (at 95% confidence level). "*" denotes that no significant performance drop was found.

Hassan (2020) successfully distilled the encoder-only BERT/RoBERTa LMs, compressing models either during pretraining or fine-tuning. KD in a decoder-only LM has been demonstrated with the DistilGPT-2 LM.[1] In encoder-decoder setups, Kim and Rush (2016) first showed positive results for decoder KD in neural machine translation (NMT). Recently, Kasai et al. (2021) showed that a single-layer decoder may be enough in NMT. Closest to our study, Shleifer and Rush (2020) showed KD with lightweight decoders in text summarization.

Orthogonal to KD, Voita et al. (2019) investigated the role and importance of individual heads in multi-head attention in a NMT transformer model. They focused on the encoder and found some heads had a specific linguistically interpretable function while others did not, and successfully pruned 38 out of 48 heads with a marginal drop in performance.

3 Methods

In this section, we introduce BART-light and BART-encoder-only, two core ideas of this paper. Unlike in other studies, e.g. (Lewis et al., 2020; Shleifer and Rush, 2020), both model configurations and all experiments use BART-*base* which has 6 encoder and decoder layers, with 12 attention heads each and the model inner dimension of 768 (ca 139 M parameters in total). This decision was driven by (1) available computational resources and (2) a desire to derive a more compact architecture than one based on the BART-*large* which is ca. 3× bigger.

3.1 BART-light: Reduced Decoder, Knowledge Distillation

For the BART-light, we retain only one or both of the top and the bottom decoder layers from the base model, since the bottom layer is trained to process the encoder's last hidden state and the top one to model the output probability space. Other than that, the model specification corresponds to BART-*base*.

For generation tasks, we fine-tune the pruned model in the basic setting. In addition, we use a KD approach inspired by Sanh et al. (2019) and Shleifer and Rush (2020). Our student (pruned) model learns to imitate a teacher (full) model via minimizing Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) over student's and teacher's predictions to help the student to model the probability space for the next predicted token.

KL divergence over a sentence is calculated as:

$$D_{\mathrm{KL}} = \sum_{i=1}^{L} \sum_{w \in \mathcal{V}} \mathbf{a}^{(i)} \cdot \mathbf{y}_t^{(i)}(w) \log \left(\frac{\mathbf{y}_t^{(i)}(w)}{\mathbf{y}_s^{(i)}(w)} \right),$$

where $\mathbf{y}_s$ and $\mathbf{y}_t$ are output probabilities of student and teacher models assigned to generated tokens, and $\mathbf{a}^{(i)} \in \{0, 1\}$ denotes an attention mask.

Similarly to Sanh et al. (2019) and Shleifer and Rush (2020), we also experiment with initializing the student from a fine-tuned teacher. First, we fine-tune the full model and then use corresponding layers for the distilled model's weight initialization.

[1] https://huggingface.co/distilgpt2

Model	Layers	Params	ConvAI2 F1	R1	R2	RL	BLEU	Eval (ms)	Speed-up
BART-base	6-6	139M	28.27	29.33	18.44	28.78	23.29	550	1.00×
BART-light-top+bottom	6-2	102M	28.35*	29.18*	18.45*	28.61*	**23.42***	443	1.24×
+ init			28.22	29.14	18.33	28.56	22.98		
+ KD			27.01	28.93	18.16	28.40	20.89		
+ KD + init			27.14	29.03	18.09	28.45	20.78		
BART-light-top	6-1	92M	27.71	29.03	18.19	28.48	22.44	**405**	**1.36**×
+ init			**28.64***	**29.44***	**18.50***	**28.83***	23.16*		
+ KD			27.36	29.24	18.38	28.69	21.15		
+ KD + init			27.48	29.29	18.34	28.74	20.86		
BART-light-bottom	6-1	92M	25.62	27.33	16.44	26.78	19.12	**405**	**1.36**×
+ init			27.20	28.39	17.29	27.81	21.03		
+ KD			26.20	28.30	17.25	27.70	19.64		
+ KD + init			26.28	28.21	17.10	27.62	19.30		

Table 2: Results on ConvAI2. We tested BART-light-top+bottom and BART-light-top (+init) for significant performance drops and mark with "*" cases where no significant drop was found (see Table 1 for details).

3.2 BART-encoder-only Model

Lewis et al. (2020) feed both the encoder and decoder with the same input for sequence classification tasks. In our BART-encoder-only model, we remove the whole decoder and move BART's classification head directly onto the encoder.

The input sentence with an appended EOS token is fed into the encoder. The EOS token representation from the final encoder layer is used by the classification head. This approach is thus similar to the BERT classification scenario (Devlin et al., 2018), but we can leverage the same pretrained model for both classification and generative tasks.

For both binary and multi-label classification, we retain the KL divergence as our KD objective.

4 Experimental Setup

For our experiments, we rely on a pretrained BART-*base* checkpoint from HuggingFace (Wolf et al., 2020). We train each model with 5 random seeds and present the mean of each metric on either dev or test set. We use standard training and decoding parameters (see Appendix for details).

4.1 Data and Tasks

We choose 5 different datasets to cover a diverse range of text generation and classification tasks that BART is known to handle well:

XSum (Narayan et al., 2018) – a news summarization task designed for abstractive models.

CNN/DailyMail (Hermann et al., 2015) – a news extractive summarization dataset with summaries being very similar to source documents.

ConvAI2 (Dinan et al., 2019) – a dialogue response generation task conditioned on personas and previous turns. We conduct experiments on a human-bot evaluation dataset.

MNLI (Williams et al., 2018) – a corpus for a multi-class textual entailment recognition.

SST-2 (Socher et al., 2013) – a movie summary dataset for a binary sentiment classification. The GLUE version is used (Wang et al., 2018).

4.2 Evaluation metrics

To evaluate all generation tasks, we use ROUGE (Lin, 2004). For ConvAI2, we also measure token F1 (Dinan et al., 2019), BLEU score (Post, 2018), and inference time. A bootstrap hypothesis testing (200 random runs) is used to assess if compressed seq2seq models are statistically significantly (at the 95% level) worse than our BART-*base*. Accuracy is used for classification tasks. In addition, we analyze the attention heads (Voita et al., 2019) on the ConvAI2 in order to evaluate if "confident" heads are retained in the model after pruning.

5 Results and Discussion

BART-light achieves comparable results to BART-*base* over all examined tasks. For generation tasks, quantitative results are confirmed by cursory manual checks of generated outputs (see examples in the Appendix). When only one decoder layer is retained, the model with the top layer retained consistently outperforms the one with the bottom layer retained. For classification tasks, the choice of which layers are retained is unimportant. Moreover, BART-encoder-only also performs well. Findings on individual tasks are elaborated below.

Summarization Results are shown in Table 1. BART-light with two decoder layers does not per-

Model	Layers	MNLI m/mm	SST-2
BERT (Devlin et al., 2018)	12	84.6/83.4	93.5
DistilB. (Sanh et al., 2019)	6	✗/✗	91.7
MobileB. (Sun et al., 2020)	24*	84.3/83.4	92.6
TinyB. (Jiao et al., 2020)	6	84.6/83.2	93.1
BART-base (ours)	6-6	**85.4/86.1**	93.2
BART-light-top+bottom	6-2	84.6/85.3	92.8
+ init		84.6/84.8	92.7
+ KD		85.1/85.1	**94.1**
+ KD + init		85.0/85.0	92.8
BART-light-top	6-1	84.7/85.4	93.1
+ init		85.0/84.8	92.5
+ KD		84.9/85.1	93.8
+ KD + init		85.0/85.0	92.8
BART-light-bottom	6-1	84.7/84.7	92.7
+ init		85.2/85.0	92.4
+ KD		84.8/84.8	94.0
+ KD + init		84.9/85.1	92.8
BART-encoder-only	6	84.5/84.6	91.9
+ init		84.5/84.3	92.4
+ KD		84.4/84.6	93.4
+ KD + init		84.8/84.9	92.4

Table 3: Accuracy on MNLI (m/mm sections) and SST-2, with the following baselines: BART-base, DistilBERT, MobileBERT (*has only 25.1 M parameters. due to narrower layers) and TinyBERT.

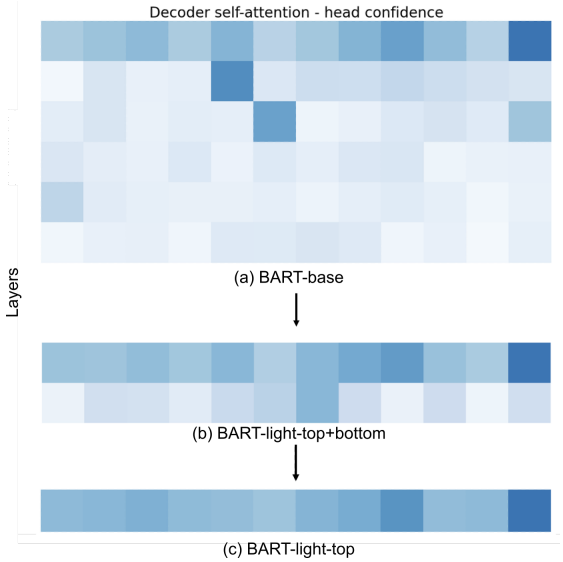

Figure 1: Mean maximum attention weight on ConvAI2 (darker color means higher weight; see Voita et al. (2019) for a more detailed definition), for the following models: (a) the original BART-base model, (b) BART-light-top+bottom, (c) BART-light-top.

form significantly worse than BART-*base* on either extractive or abstractive summarization tasks. When only one layer is retained, our pruned model still reaches 96-99% of BART-*base*'s performance. As expected, the drop in the performance tends to be higher for abstractive summarization than for the extractive one.

Student initialization leads to the best performance for BART-light; this approach looks most promising also for its simple implementation and lower training cost. On the other hand, KD tends to perform poorly on summarization. These results are opposed to Shleifer and Rush (2020), suggesting more hyper-parameter tuning might be needed.

Dialogue Response Generation BART-light with only the top or top+bottom decoder layers does not perform statistically worse than the full model on any observed metrics as shown in Table 2. Generally, utilizing KD helps to improve token recall, but has a negative effect on precision.

On this task, we also compare the models' generation latency. Our findings show that BART-light-top is not only slightly better than BART-*base*, but also provides 1.36× speed-up and requires less memory as it has 1.51× less parameters, making it more suitable for interactive tasks.

Classification The results are shown in Table 3. All versions of BART-light and BART-encoder-only models slightly outperform BART-*base* on SST-2. On this task, using KD leads to an increase of 1.2 % in accuracy on average. Multi-class textual entailment classification is a more difficult task than binary sentiment analysis task, which is a plausible explanation why BART-light tends to be slightly worse here. Nevertheless, BART-encoder-only and BART-light models, regardless of the choice of decoder layers, retain around 99% accuracy of BART-*base*. The classification results suggest that one can simply pretrain an encoder-decoder-based model and then use only the encoder for problems where the decoder is not required.

Attention head analysis Our attention head analysis demonstrates that pruning the decoder layers does not lead to losing confident attention heads. As depicted in Figure 1, when we retain only the top and bottom layers, new confident heads are learned, standing in for the ones from the removed layers. The confidence of the weights in the top layer is mostly preserved. When further only the top layer is kept, all heads end up quite confident.

In future work, we plan to experiment with pruning out certain attention heads from the remaining layer to investigate further model size reductions.

References

Cristian Bucilå, Rich Caruana, and Alexandru Niculescu-Mizil. 2006. Model compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, page 535–541, New York, NY, USA. Association for Computing Machinery.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, et al. 2019. The second conversational intelligence challenge (convai2). *arXiv preprint arXiv:1902.00098*.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, and Qun Liu. 2020. Tinybert: Distilling bert for natural language understanding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 4163–4174.

Jungo Kasai, Nikolaos Pappas, Hao Peng, James Cross, and Noah Smith. 2021. Deep encoder, shallow decoder: Reevaluating non-autoregressive machine translation. In *International Conference on Learning Representations*.

Yoon Kim and Alexander M. Rush. 2016. Sequence-level knowledge distillation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1317–1327, Austin, Texas. Association for Computational Linguistics.

Young Jin Kim and Hany Hassan. 2020. Fastformers: Highly efficient transformer models for natural language understanding. In *Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing*, pages 149–158. Association for Computational Linguistics.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395, Barcelona, Spain. Association for Computational Linguistics.

Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

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*, pages 7871–7880. Association for Computational Linguistics.

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

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. 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)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

Paul Michel, Omer Levy, and Graham Neubig. 2019. Are sixteen heads really better than one? In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting bleu scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191. Association for Computational Linguistics.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. 2019. Green ai. *arXiv preprint arXiv:1907.10597*.

Sam Shleifer and Alexander M Rush. 2020. Pre-trained summarization distillation. *arXiv preprint arXiv:2010.13002*.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, Washington, USA. Association for Computational Linguistics.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.

Zhiqing Sun, Hongkun Yu, Xiaodan Song, Renjie Liu, Yiming Yang, and Denny Zhou. 2020. Mobilebert: a compact task-agnostic bert for resource-limited devices. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2158–2170. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. 2019. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5797–5808, Florence, Italy. Association for Computational Linguistics.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.

Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122, New Orleans, Louisiana. Association for Computational Linguistics.

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*, pages 38–45, Online. Association for Computational Linguistics.

Appendix

Table 4 shows statistics of datasets used. MNLI and SST-2 do not have publicly available test sets, therefore, we evaluate on the development sets and choose a portion of the training data at random for validation. For ConvAI, individual examples were generated from whole dialogues drawn from the ConvAI2 dataset version available from HF's datasets. The source code for example extraction is available in the ZIP file.

Table 5 shows used training parameters. When knowledge distillation used, we caclculate the mean of label-smoothing cross entropy loss and Kullback–Leibler divergence.

Table 6 then presents parameters used for generation.

The remaining tables show example XSum summaries and CNN/DailyMail highlights. Each table contains the source document (article) and target summary (highlights) together with generated predictions of BART-base models, BART-Light-Top+Bottom and BART-Light-Top.

Dataset	Training	(-for validation)	Development	Test
XSum	204,045		11,332	11,334
CNN/DailyMail	287,713		13,368	11,490
ConvAI2	32,838		4,064	4,098
MNLI	392,702	(-8,702)	9,815 / 9,832	-
SST-2	67,349	(2,069)	872	-

Table 4: Dataset size statistics. We show the number of training examples, number of training examples used for validation (in case of MNLI and SST-2), size of the development set (used for validation in XSum, CNN/DailyMail and ConvAI; used for evaluation on MNLI, SST-2) and the size of the test set (if available). The development set numbers for MNLI show the "m" and "mm" sections, respectively.

	XSum	**CNN**	**ConvAI2**	**MNLI**	**SST**
Batch size	256	64	32	128	128
Input length	512	1024	512	128	64
Target length	50	64	128	✗	✗
Max epochs	8	4	20	5	10
Learning rate	0.00002	0.00001	0.00002	0.00003	0.00002
Label smoothing	0.1	0.1	0.1	0.1	0.1
(β_1, β_2)	(0.9, 0.999)	(0.9, 0.999)	(0.9, 0.999)	(0.9, 0.98)	(0.9, 0.98)
ε	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$

Table 5: Training hyperparameters

	XSum	**CNN**	**ConvAI2**
Min length	10	56	5
Max length	60	142	160
Length penalty	1.0	0.8	0.8
Beam size	5	5	5
No repeat n-grams	3	3	✗

Table 6: Generation hyperparameters

Document	The country's consumer watchdog has taken Apple to court for false advertising because the tablet computer does not work on Australia's 4G network. Apple's lawyers said they were willing to publish a clarification. However the company does not accept that it misled customers. The Australian Competition and Consumer Commission (ACCC) said on Tuesday: "Apple's recent promotion of the new 'iPad with wi-fi + 4G' is misleading because it represents to Australian consumers that the product can, with a sim card, connect to a 4G mobile data network in Australia, when this is not the case." The watchdog then lodged a complaint at the Federal Court in Melbourne. At a preliminary hearing, Apple lawyer Paul Anastassiou said Apple had never claimed the device would work fully on the current 4G network operated by Telstra. Apple says the new iPad works on what is globally accepted to be a 4G network. The matter will go to a full trial on 2 May. The Apple iPad's third version went on sale earlier this month, with Australia the first country where it was available. Shoppers lined up by the hundreds at Apple stores on opening day and the company said it had been its strongest iPad launch to date. The ACCC said it was seeking an injunction on sales as well as a financial penalty against Apple, corrective advertising and refunds to consumers. On its website, Apple does state that 4G LTE is only supported on selected networks in the US and Canada.
Target summary	US technology firm Apple has offered to refund Australian customers who felt misled about the 4G capabilities of the new iPad.
BART-Base	Apple has been accused of misleading customers in Australia over its new iPad tablet computer's use of 4G mobile data.
BART-Top+Bottom	Apple has been accused of misleading Australian consumers over its promotion of the iPad with wi-fi and 4G mobile data.
BART-Top	Hundreds of people lined up in Australia to sell its new iPad with wi-fi and 4G LTE (4G LTE) tablet in Australia.

Table 7: XSUM example

Document	Alex Williams, 22, from Christchurch, had failed to appear in court for sentencing in November, but was arrested in the town on Sunday. He appeared at Bournemouth Crown Court earlier and was sentenced to 20 months in prison for violent disorder. Williams was also sentenced to a further two months for failing to appear in court. Detectives had appealed to the public to help locate him after he failed to show up. The gang's offences included throwing a liquid, thought to be ammonia, at two victims who were each left blind in one eye. Four men were jailed for the attacks. In November Joe Warne, 21, was sentenced to 20 years; Reece Watkins, 22, was sentenced to 18 years; Dominic Barker, 20, was given 16 years and Piers Fox-Havilland was jailed for 12 years. The men, who lived in the London area, were also sentenced to 12 years for conspiracy to rob and eight for possession of imitation firearms - to run concurrently. Dorset Police said the gang was charged following a lengthy investigation into attacks at locations including Boscombe and Christchurch.
Target summary	A man who was part of a five-strong gang that committed violent acts in Dorset has been jailed.
BART-base	A man has been jailed for a string of violent attacks on two women in Dorset which left them blind in one eye.
BART-Top+Bottom	A man who went missing after a series of acid attacks in Dorset has been jailed.
BART-Top	A man who went missing after a series of acid attacks in Dorset has been jailed

Table 8: XSUM example

Document	Cracks in its structure were spotted in March and the area was sealed off. The IFA has confirmed to the BBC that it received a preliminary engineer's report into the damage and is studying the findings. The venue is scheduled to host the Euro 2016 qualifier against Romania on 13 June. Kevin McShane of the Institute of Civil Engineers told BBC Radio Ulster's Good Morning Ulster programme that it was a surprise "nothing more serious" had happened because the foundations had moved. "We can see at the base of the columns that the wall and the beams have moved," he said. "That's caused stress throughout the structure, given the shape of the structure it's very top heavy, it's leaning on one side towards the pitch. "So if there's some problems with the foundation, the natural reaction of the whole structure is to try to topple over onto the pitch." On Wednesday, the IFA's chief executive Patrick Nelson said staging the international at Windsor Park remained its "only focus". He said an alternative venue for the qualifier, such as the Kingspan Stadium, had not yet been considered. It is believed that at least two independent reports are being carried out on the damaged stand, with cracks first spotted on 31 March, two days after Northern Ireland's Euro 2016 qualifier against Finland. Linfield are unable to use their home venue and the Irish Cup final on 2 May, which was scheduled to be played at Windsor Park, has been switched to the Oval.
Target summary	The Irish Football Association (IFA) has been advised to demolish the damaged west stand at Windsor Park, the BBC understands.
BART-base	The Irish Football Association (IFA) is investigating the damage caused to the stand at Windsor Park.
BART-Top+Bottom	The Irish Football Association (IFA) is investigating reports that a damaged stand at Windsor Park is to be toppled over.
BART-Top	The Irish Football Association (IFA) has confirmed that it is understood to have been damaged by the Irish FA's building at Windsor Park.

Table 9: XSUM example

Document	Mr Capriles has been at the forefront of efforts calling for a recall referendum on President Nicolas Maduro. There was no immediate comment from the government. The ban comes after a week of several large opposition-led protests in the capital Caracas. They accuse the government of stifling dissent. On Thursday a man was shot dead during the demonstrations. Over the last few days, authorities have accused Mr Capriles of inciting violence and bloodshed by leading protests against the unpopular president. The protesters were demonstrating against a decision by the Venezuelan Supreme Court to assume control of the opposition-led congress. Although the court's decision was quickly overturned, the street protests continued. Mr Capriles, who is a former presidential candidate and the governor of the state of Miranda wrote on social media: "URGENT: I inform the country and international public opinion that I am being notified at this very moment of a BAN for 15 years." He accused the government of running a smear campaign against him. Venezuela: What's behind the turmoil? The most expensive Nutella in the world Leopoldo Lopez loses appeal In 2015 another prominent opposition leader, Leopoldo Lopez was sentenced to nearly 14 years in prison on charges of inciting violence during anti-government protests in 2014. Mr Lopez was himself barred from office in 2008 when he was the popular mayor of a Caracas district. The Venezuelan president, Nicolas Maduro, narrowly defeated Mr Capriles in the 2013 elections, a result that sparked controversy and debate as the opposition claimed electoral fraud. Mr Maduro's government have said that a US-backed business elite is responsible for Venezuela's economic downturn and that it was trying to organise a coup to impose right wing rule.
Target summary	A leading Venezuelan opposition leader, Henrique Capriles, says he has been formally banned for 15 years from political activity.
BART-base	Venezuelan opposition leader Henrique Capriles has been barred from office for 15 years for inciting violence.
BART-Top+Bottom	Venezuelan opposition leader Henrique Capriles has been banned from office for 15 years.
BART-Top	Venezuela's opposition leader Henrique Capriles has been banned from office for 15 years for inciting violence.

Table 10: XSUM example

Document	The 21-year-old, who scored 10 goals in 50 appearances for the U's last season, has signed a three-year contract with the Championship side. O'Dowda made his Republic of Ireland debut against Belarus in May, and scored in Oxford's Johnstone's Paint Trophy final defeat by Barnsley. "He's a real emerging talent," said Bristol City head coach Lee Johnson. "Callum isn't exactly under the radar because there was a lot of interest in him, so we're delighted he's joined us - we believe he can continue the progression he has already shown." O'Dowda, who had two years left on his contract with hometown club Oxford, had missed the start of their pre-season tour to Spain because of "illness". The fee could rise to £1.6m, with Oxford owed a percentage of any future sale. Find all the latest football transfers on our dedicated page.
Target summary	A Bristol City have signed Oxford United winger Callum O'Dowda for a fee of at least £1m.
BART-base	Bristol City have signed striker Callum O'Dowda from Oxford United for an undisclosed fee.
BART-Top+Bottom	Bristol City have signed Oxford United striker Callum O'Dowda for an undisclosed fee.
BART-Top	Bristol City have signed Republic of Ireland striker Callum O'Dowda from Oxford United for an undisclosed fee.

Table 11: XSUM example

Document	The FTSE 100 index closed 2.27% higher, while the FTSE 250 closed up 1.68%. The pound fell 1.2% against the dollar to $1.3264 after Mr Carney said a deteriorating outlook meant action from the Bank was likely during the summer. Shares in the US rose, with both the Dow Jones and S&P 500 index up by 0.6% in morning trading. The past few days have seen huge swings on the world's financial markets as traders and investors struggled to assess the impact of the UK's vote to leave the European Union. After two days of heavy falls following the vote, shares spent two days recovering, and on Wednesday, the FTSE 100 ended above the level it had closed at last Thursday. However, the FTSE 250 index - which contains more UK-focused companies - still remains well below its pre-referendum level. "The blue chip [FTSE 100] index is stuffed full with big international firms and it is these companies propelling the rise. The heavy weighting towards these groups is very misleading," said Joe Rundle, head of trading at ETX Capital. "The likes of Fresnillo, Randgold, AstraZeneca, Royal Dutch Shell, British American Tobacco - these are hardly dependent on, or reflective of, the UK economy. They're listed in London but their earnings come from abroad. "There is a key distinction. UK-focused firms are doing much, much worse. Easyjet, Lloyds, Barclays, RBS, Barratt, Taylor Wimpey - they've all recorded 20% losses since the Brexit vote." Shares in RBS were one of the biggest fallers on the FTSE 100 on Thursday, dropping 4.8%. The pound is well below the $1.50 rate it touched last Thursday before the outcome of the referendum became clear. As well as falling against the dollar after Mr Carney's speech, it also tumbled 1.1% against the euro at â,¬1.1962. Analysts continued to warn that the recent market recovery might only be temporary. "The odds may be against them, but investors are hoping that the worst is over for currencies and equities," Kathy Lien, managing director of foreign exchange strategy at BK Asset Management, said in a note. "But considering there's been no additional clarity on the terms of Brexit or the outlook for the UK economy and global economy since Britain's decision to leave the European Union on Friday, we don't see fundamental support for the recent moves." Analysts at Swiss bank UBS have cut their end of year forecast for the FTSE 100 to 5,500 from 6,500. "We see a significant amount of uncertainty around the UK over the next few months. We do not know who the prime minister will be, or when, or if, Article 50 is invoked and there are even possibilities of a general election given the current fluidity of UK politics," UBS said. The price of gold was little changed at $1,318.10 an ounce on Thursday. The precious metal is viewed as a safe investment in times of uncertainty and gold price hit a near two-year high of $1,358.20 on Friday in the wake of the referendum.
Target summary	UK shares have risen and the pound has fallen after Bank of England governor Mark Carney hinted at fresh stimulus measures after the Brexit vote.
BART-base	(Close): London's leading shares closed higher on Thursday, as Bank of England governor Mark Carney said action was needed to stabilise the UK economy.
BART-Top+Bottom	UK shares have risen after Bank of England governor Mark Carney gave a speech to Prime Minister Mark Carney.
BART-Top	UK shares have risen after Bank of England governor Mark Carney's speech on Friday.

Table 12: XSUM example

Article	(CNN)Richard Glatzer, who directed a powerful film about a professor battling Alzheimer's as he faced his own harrowing health struggles, has died. Glatzer died in Los Angeles on Tuesday after having ALS for four years, his publicist said. He was 63. Glatzer co-directed "Still Alice" with his husband, Wash Westmoreland. The 2014 film earned a number of major awards for its lead actress, Julianne Moore. Directing the movie was a challenge that Glatzer embraced, even as he faced a growing number of health obstacles after his ALS diagnosis in 2011. "On set, he inspired the cast and crew with his perseverance, (co-directing) the film by typing with one finger into a text-to-speech app on his iPad," his publicist's statement said. In a Twitter post Wednesday, Westmoreland said he was devastated. "Richard was my soul mate, my collaborator, my life," he said. "A true artist and a brilliant man." Opinion: Why 'Still Alice' is about you . When she accepted her Academy Award for best actress last month for her role in the film, Moore noted Glatzer's absence. "Finally, to our filmmakers, Wash Westmoreland and Richard Glatzer, who had hoped to be here tonight, but they can't because of Richard's health," she said. "When Richard was diagnosed with ALS, Wash asked him what he wanted to do. Did he want to travel? Did he want to see the world? He said he wanted to make movies. And that's what he did." People we've lost in 2015 . CNN's Topher Gauk-Roger contributed to this report.
Target highlights	Richard Glatzer, who co-directed "Still Alice," has died . He had been battling ALS for four years .
BART-base	Richard Glatzer died in Los Angeles on Tuesday after having ALS for four years. Glatzer co-directed "Still Alice" with his husband, Wash Westmoreland. The 2014 film earned a number of major awards for its lead actress, Julianne Moore.
BART-Top+Bottom	Richard Glatzer died in Los Angeles on Tuesday after having ALS for four years. Glatzer co-directed "Still Alice" with his husband, Wash Westmoreland. The 2014 film earned a number of major awards for its lead actress Julianne Moore.
BART-Top	Richard Glatzer died in Los Angeles on Tuesday after having ALS for four years. Glatzer co-directed "Still Alice" was my soul mate, my collaborator, my life," his publicist says. Westmoreland says he was devastated by his health.

Table 13: CNN/DailyMail example

Article	Educator: Los Angeles teacher Thelmo Garcia, 37, is accused of placing a plastic over the head of a 10-year-old female student in class last month . A Los Angeles elementary school teacher was arrested this week after police say he placed a plastic bag over the head of a 10-year-old student. Thelmo Garcia, 37, a fourth-grade teacher at Rosewood Elementary School, pleaded not guilty to one count of child abuse Friday, two days after he was taken into custody in connection to the February 5 incident. According to the Los Angeles County District Attorney's Office, the educator placed a bag over the head of a girl in his class 'for several seconds' during school hours. The child reportedly did not pass out after the stunt, according to CBS Los Angeles. Investigators would not say whether Garcia was disciplining the girl, or just joking with her, as he's been known to do with other students. Garcia has been with the Los Angeles Unified School District since 1999 and taught fourth grade at Rosewood since 2010. The 37-year-old man was ordered held on $100,000 bail. If convicted, he could face up to six years in prison. The school has suspended Garcia with pay pending an investigation. Scroll down for video . Place of learning: The alleged incident took place at Rosewood Elementry School, where Garcia has taught fourth grade since 2010 . The 37-year-old pleaded not guilty and was ordered held on $100,000 bail . 'We take these allegations very seriously,' Ellen Morgan, of the Los Angeles Unified School District, said in a statement. 'The district is fully cooperating with the LAPD investigation.' Some parents and students at the LA school were stunned to learn of the allegations against Garcia, who was described as a caring and playful teacher. According to CBS, Garcia had no history of infractions over the course of his teaching career.
Target highlights	Thelmo Garcia, 37, pleaded not guilty Friday to one count of child abuse . Prosecutors say on February 5, Garcia placed a bag over the head of a 10-year-old girl for several seconds at Rosewood Elementary School .
BART-base	Thelmo Garcia, 37, pleaded not guilty to one count of child abuse Friday, two days after he was taken into custody in connection to the February 5 incident. Garcia has been with the Los Angeles Unified School District since 1999 and taught fourth grade at Rosewood since 2010.
BART-Top+Bottom	Thelmo Garcia, 37, a fourth-grade teacher at Rosewood Elementary School, pleaded not guilty to one count of child abuse Friday, two days after he was taken into custody in connection to the February 5 incident. Garcia has been with the Los Angeles Unified School District since 1999 and taught
BART-Top	Thelmo Garcia, 37, is a fourth-grade teacher at Rosewood Elementary School. The 37-year-old pleaded not guilty and was ordered held on $100,000 bail. Garcia has been with the Los Angeles Unified School District since 1999 and taught fourth grade since 2010.

Table 14: CNN/DailyMail example

Article	(CNN)The misuse of outer protective garments may have led to the exposure of a potentially deadly strain of bacteria at the Tulane National Primate Research Center near New Orleans, the Centers for Disease Control and Prevention said Friday. An employee at the center has tested positive for the bacterium, which is kept at the facility. The employee is not sick, and Jason McDonald, a CDC spokesman, said the bacteria probably aren't a threat to the general population. Inspectors from the CDC and the U.S. Animal and Plant Health Inspection Service said the misuse of outer garments "could have led to the bacteria clinging to inner garments and getting carried out of the select agent lab where research was being conducted with the bacteria on mice," a news release said. "Additionally, CDC and APHIS inspectors determined that Tulane primate center staff frequently entered the select agent lab without appropriate protective clothing, which would increase the risk of bringing the bacteria out of the lab or becoming infected themselves." The bacterium, Burkholderia pseudomallei, was being tested on mice in a biosafety level 3 lab at the Covington, Louisiana campus. It can cause can cause melioidosis, also known as Whitmore's disease. All research with the agent at the facility was suspended on February 11 and will remain suspended until it can be shown that there are no more risks and that proper procedures are being followed, the CDC said. The CDC says the primate facility can resume that research when Tulane officials show inspectors that: . • Entity-wide procedures exist to ensure animals accidentally exposed in the future are managed appropriately; . • All personal protective equipment procedures are thoroughly reviewed and revised appropriately to lessen the risk of future breaches; . • All Tulane primate center personnel are trained on any new or revised protective clothing procedures; and . • Improved entry and exit procedures to the outside enclosures housing non-human primates are in place to stop any further transmission among the animals." The CDC and U.S. Department of Agriculture say they have completed their investigation, which began in November when two monkeys were diagnosed with Whitmore's disease. Six others had antibodies indicating exposure to the bacterium. According to the CDC, "the bacteria causing melioidosis are found in contaminated water and soil. It is spread to humans and animals through direct contact with the contaminated source." It is not transmitted between humans or animals, "and the risk of acquiring melioidosis is low," the CDC said. Melioidosis "is predominately a disease of tropical climates, especially in Southeast Asia and northern Australia where it is widespread," according to the CDC website.
Target highlights	The error "could have led to the bacteria clinging to inner garments and getting carried out" of a lab, inspectors say . Also, "staff frequently entered the select agent lab without appropriate protective clothing" The CDC says the bacteria probably aren't a threat to the general population .
BART-base	An employee at the center has tested positive for the bacterium, which is kept at the facility. The CDC says the bacteria probably aren't a threat to the general population. It can cause can cause melioidosis, also known as Whitmore's disease.
BART-Top+Bottom	An employee at the Tulane National Primate Research Center tested positive for the bacterium. The CDC says the bacteria probably aren't a threat to the general population. It can cause melioidosis, also known as Whitmore's disease, is not transmitted between humans or animals.
BART-Top	An employee at the Tulane National Primate Research Center near New Orleans. The bacteria was being tested on mice in a biosafety level 3 lab at the Covington, Louisiana campus. It can cause melioidosis, also known as Whitmore's disease.

Table 15: CNN/DailyMail example

Article	A woman in Oregon has pleaded guilty to having an illegal relationship with her biological father, who she never knew growing up. Chalena Moody, 25, was sentenced to 10 days in jail on the incest charge, but got credit for time already served and won't have to serve any additional days behind bars. She also received 18 months probation. The authorities say Moody and her father, Eric Lee Gates, 49, were living as a couple in Springfield, before moving to Klamath Falls and had two children under two years of age. Chalena Moody, 25, has pleaded guilty to having an illegal relationship with her biological father Eric Lee Gates, 49. They have been living as couple in Springfield, Oregon, and have two young kids . A third died in utero, according to county prosecutors. The two children have numerous medical issues and are now in the custody of the state, reports The Register-Guard. 'There's issues with the children, and they've been put into foster care,' said Sgt. Dave Lewis with Springfield Police. 'It's not a good situation all around.' As part of her probation, Moody cannot have any contact with Gates. He is serving 60 days in the Lane County Jail for violating his probation and moving in with Moody following his guilty plea and subsequent sentencing to the same probationary terms. Gates was also ordered to pay restitution to the State of Oregon for the medical expenses of the two children. Gates is serving 60 days in the Lane County Jail for violating his probation and moving in with Moody following his guilty plea to the incest charge . Authorities say Gates was not Moody's custodial parent during her childhood, as he served several stints in prison while she was growing up. Chalena Moody was sentenced to 10 days in jail on the incest charge and also received 18 months probation . Court records show at least one of those prison terms was for a felony conviction of driving with a suspended license. However, Gates and Moody reconnected in recent years and started their relationship. Springfield police Detective Dave Lewis said Gates and Moody openly admitted to their consensual relationship, to being father-daughter and also conceiving children together. DNA tests were performed on the family, and it was determined that Gates and Moody were related and had children from an incestuous relationship, Lewis said. Oregon State law defines incest as when a person marries or engages in sexual intercourse or deviate sexual intercourse with another whom the person knows to be related to them, either legitimately or illegitimately, as an ancestor, descendant, brother or sister of either whole or half-blood. Incest is a Class C felony, punishable by up to five years in prison and $125,000 in fines.
Target highlights	Chalena Moody, 25, has pleaded guilty to having an illegal relationship with her biological father Eric Lee Gates, 49 . Authorities say they had been living as a couple in Springfield, Oregon, and have two children under two years of age . The two children have numerous medical issues and are now in the custody of the state . Gates has been ordered to pay restitution to the state for the medical expenses of the children . He was not Moody's custodial parent during her childhood, as he served several stints in prison .
BART-base	Chalena Moody, 25, was sentenced to 10 days in jail on the incest charge, but got credit for time already served and won't have to serve any additional days behind bars. She also received 18 months probation. The authorities say Moody and her father, Eric Lee Gates, 49, were
BART-Top+Bottom	Chalena Moody, 25, was sentenced to 10 days in jail on the incest charge, but got credit for time already served and won't have to serve any additional days behind bars. The authorities say Moody and her father, Eric Lee Gates, 49, were living as a couple in Springfield, Oregon
BART-Top	Chalena Moody, 25, was sentenced to 10 days in jail on the incest charge, but got credit for time already served and won't have to serve any additional days behind bars. Moody and her father, Eric Lee Gates, 49, were living as a couple in Springfield, Oregon,

Table 16: CNN/DailyMail example

Article	The FA are set to decide Jonny Evans's fate today after the Manchester United defender denied allegations of spitting at Newcastle striker Papiss Cisse. A disciplinary hearing was hastily set up last night after Evans confirmed his intention to fight the charge which will carry a six-game ban if he is found guilty. The FA wanted to make sure the matter was dealt with swiftly before the Northern Ireland international played in United's FA Cup quarter-final clash with Arsenal at Old Trafford on Monday night. Jonny Evans has decided to contest the FA charge of spitting on Newcastle forward Papiss Cisse . Evans appears to launch spit in the direction of the Newcastle United striker, who then retaliates . Evans and Cisse clash near the half-way line at St James' Park following the unpleasant exchange . Arsenal (h), Monday; Tottenham (h), March 15; Liverpool (a), March 22; Aston Villa (h), April 4; Manchester City (h) April 12; Chelsea (a) April 18. They tweeted confirmation of Evans's decision on Friday, just over an hour after the 6pm deadline he had been given to respond to the charge. It read: 'Manchester United's Jonny Evans denies FA charge in relation to incident involving Papiss Cisse. Independent Regulatory Commission hearing tonight, decision expected tomorrow.' It is understood that United will submit video evidence to support Evans's claim that he did not spit directly at Cisse. Both players were charged over the unsavoury incident in United's 1-0 win at St James' Park on Wednesday night. Cisse issued a public apology and accepted his guilt, triggering an immediate seven-match ban. The Senegal striker's suspension is one game longer because he had already served a three-match ban earlier this season. Evans was expected to contest the charge after issuing a statement on Thursday denying any knowledge of a spitting incident, which followed a tackle between the two players in the first half. He said: 'Having woken up this morning I am shocked to have seen the media coverage from last night's match. I would like to make it clear that I did not spit at Papiss Cisse. 'I was totally unaware of any spitting incident and had assumed that the issue at the time was with the challenge and his attempted retaliation to the tackle from the floor. Cisse has been banned for seven games after accepting the charge from the FA on Thursday . Tempers threaten to boil over as team-mates from Newcastle and United gather on the pitch . Cisse was banned for an extra game after being sent off for elbowing Seamus Coleman in December . 'During the game Papiss Cisse and I spoke about the incident and it is clear by my reaction in the television footage that I was totally surprised by any suggestion of spitting. 'It is not in my character or in my nature to spit at anybody nor is it something I have ever done or would ever do. It is certainly not something that I did last night.' United boss Louis van Gaal also defended his player, insisting that he believed Evans's version of events. Spitting is understandably reviled among the majority of Premier League players, but Stoke striker Jonathan Walters mounted a surprisingly robust objection on Friday by threatening to leave any opponent who targeted him in such a manner 'eating through a straw'. The FA are unlikely to take any action against Walters but will not welcome any talk of physical violence on the pitch. Walters told talkSPORT: 'As a player you take anything off the ball; pinches and kicks that are meant to wind you up. Little things like that go on in any match. 'But spitting is pretty low. If it happened to me and someone spat at my face or towards me then I think he'd be eating his supper through a straw that night. I wouldn't be that happy. It's the lowest of the low.'
Target highlights	Jonny Evans and Papiss Cisse involved in spitting row on Wednesday . Cisse admitted the FA charge and was given a seven-match ban . Evans denied that he spat on the Senegal international . Disciplinary hearing for the defender happening on Friday night . The FA set to announce their decision on Saturday .
BART-base	Manchester United defender Jonny Evans has decided to contest the FA charge of spitting on Newcastle striker Papiss Cisse. A disciplinary hearing was hastily set up last night after Evans confirmed his intention to fight the charge which will carry a six-game ban if he is found guilty. The FA wanted to make
BART-Top+Bottom	Jonny Evans has decided to contest FA charge of spitting on Papiss Cisse. The Manchester United defender appeared to launch spit in the direction of the Newcastle striker. Cisse has been banned for seven games after accepting the charge from the FA on Thursday. Evans was expected to contest the charge
BART-Top	Jonny Evans has decided to contest the FA charge of spitting on Papiss Cisse. Evans will contest the charge which will carry a six-game ban if he is found guilty. The Northern Ireland international played in United's FA Cup quarter-final clash with Arsenal.

Table 17: CNN/DailyMail example

Article	Roy Hodgson must convince Saido Berahino that his international future is with England after the forward received an official approach from Burundi. Berahino, who has scored 16 times for West Brom in the Barclays Premier League this season, was called into the England squad for the first time in November. The striker was born in Burundi and a delegation of officials from the tiny country in south-east Africa have been courting him this week. Saido Berahino (left, in action for West Brom against Aston Villa) has impressed England boss Roy Hodgson . Berahino was called in to the senior England squad last November but is also wanted by his native Burundi . Burundi, 126th in the FIFA rankings, play Mauritius in a friendly on March 25 and want to name the West Brom forward in their squad. England's head coach names his squad on Thursday for the Euro 2016 clash with Lithuania on March 27 at Wembley and the friendly in Turin against Italy four days later. Berahino has yet to represent the England team and he would still be permitted to switch nationalities under FIFA rules. Hodgson's striking options have been swelled by the return from injury of Daniel Sturridge, along with regular forwards Danny Welbeck and England captain Wayne Rooney. Hodgson has already made it clear that he intends to select Tottenham forward Harry Kane when he names his squad at the national stadium next week. Berahino celebrates scoring for England's U21s but can switch allegiances having not played for the seniors . Berahino trains with the senior England squad in November (left) and in action for the U21s . Kane, who is the Spurs squad travelling to Manchester United on Sunday, has scored 16 times in the Barclays Premier League and was named player of the month for the second time on Friday. His emergence means Hodgson has a dilemma over who to select as his fifth striker. Berahino, an England Under-21 international, was named in the full squad for the first time in November for the clashes with Slovenia in a Euro 2016 qualifier and the friendly with Scotland at Celtic Park. He was an unused substitute in both games, but Hodgson predicted a bright future for the forward after he spent a week training with the national team.
Target highlights	Saido Berahino's native Burundi send delegation of officials to court him . Roy Hodgson also has Daniel Sturridge, Danny Welbeck, Wayne Rooney and Harry Kane to pick and must decide whether to also call up Berahino . Berahino is yet to play for senior side so can still switch allegiances .
BART-base	Saido Berahino has scored 16 times for West Brom this season. The striker was called into the England squad for the first time in November. Burundi, 126th in the FIFA rankings, play Mauritius in a friendly on March 25 and want to name the West Brom forward in their
BART-Top+Bottom	Saido Berahino received official approach from Burundi. Berahino has scored 16 times for West Brom in the Barclays Premier League this season. Burundi want to name West Brom forward in their squad. Hodgson has already made it clear he intends to select Harry Kane
BART-Top	Saido Berahino has scored 16 times for West Brom in the Premier League. The striker was born in Burundi and a delegation of officials from the tiny country in south-east Africa. Burundi, 126th in the FIFA rankings, will be permitted to switch nationalities under FIFA

Table 18: CNN/DailyMail example