NewsStories: Illustrating articles with visual summaries supplementary

Reuben Tan[1] Bryan A. Plummer[1] Kate Saenko[1,2] JP Lewis[3]
Avneesh Sud[3] Thomas Leung[3]
[1]Boston University, [2]MIT-IBM Watson AI Lab, [3]Google Research
{rxtan, bplum, saenko}@bu.edu, {jplewis, asud, leungt}@google.com

In this supplementary material, we provide the following:

1. Additional statistics of the images, articles and videos included in our NewsStories dataset.
2. Ablation experiments over the length of the input text sequence using the best-performing MIL-SIM model.
3. An algorithm for generating suitable image sets to describe any text narrative or article.
4. Additional qualitative retrieval results on the test splits of the GoodNews dataset.

To begin, we compute and present histograms over the number of articles, images and videos that are present in the computed story clusters in Section A. Next, we conduct an ablation study in Section B over the length of the input text sequences in the MIL-SIM approach to determine the importance of using additional context for learning the correspondences between text narratives and groups of complementary images. In Section C, we describe an algorithm that an author may use to leverage the finetuned models to select a visually illustrative set of images for a given text narrative. Finally, we provide additional qualitative retrieval results of the MIL-SIM model on the test splits of the NewsStories and GoodNews datasets in Section D.

A Additional statistics of the NewsStories dataset

Article and images statistics. Figure 1 shows the histograms of the number of images and articles that are contained in each news story cluster. Note that we only show the first n bins that contain 95% of all images and articles for conciseness. As mentioned in the main paper, the number of articles per story cluster varies greatly across different clusters. Story clusters with unusually high numbers of articles are generally very noisy and tend to revolve around the theme of entertainment such as reality television updates or music videos. Additionally, a high percentage of story clusters contain between 1 and 20 images, where some of the images sourced from different media channels on the same story may be near-duplicates of each other. However, we note that such noise is prevalent in uncurated real-world data and being able to leverage these publicly available data to address the proposed research problem effectively remains an open question.

News videos statistics. We provide a histogram of the number of videos contained in each story cluster of our unfiltered NewsStories dataset in Figure 2. As corroborated by the data in Table 1, we observe that a large percentage of news story clusters do not contain any videos at all. Additionally, the number of videos in a news story cluster varies considerably, from a minimum of zero to a maximum of 4005 videos. However, with over 450K stories containing at least one corresponding video, our NewsStories dataset still provides a rich environment for learning to reason about multimodal correspondences between text, images, videos and audio.

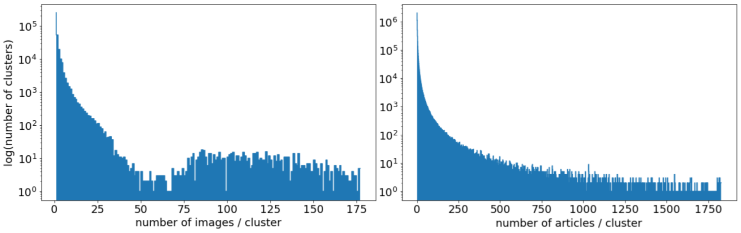


Fig. 1. Histogram of the number of images / articles that are contained in each news story cluster. For improved visualization, we show first n bins that capture 95% of total image / articles

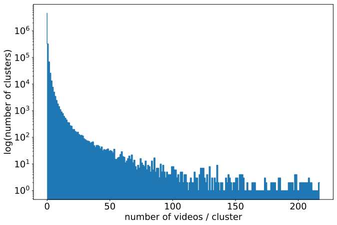


Fig. 2. Histogram of number of videos per story cluster.

B Input text length ablation

To evaluate the importance of the number of input word tokens in retrieving the most relevant image set, we conduct an ablation study over the length of the input text sequences for the best-performing MIL-SIM approach and report the results in Table 2. In our experiments, we use both article-level and sentence-level alignment objectives. In the former objective, we begin from the start of an

Table 1. Video story cluster statistics

	Unfiltered	Filtered
Min. # videos in a cluster	0	0
Max. # videos in a cluster	4005	2255
Std. dev. # videos per cluster	4.22	6.79
Mode # videos per cluster	0	0
# clusters with $\geq$ 1 video	451,228	81,957

Table 2. Ablation of our best-performing MIL-SIM approach over the length of the input text sequences for the MIL-SIM approach on the validation split of NewsStories

Input Text	R@1			R@5			R@10			Median Rank		
Length	3	4	5	3	4	5	3	4	5	3	4	5
32	34.28	36.46	38.08	56.66	60.19	62.64	65.11	68.52	70.43	3	3	3
64	**41.10**	45.31	48.04	65.78	70.02	73.37	74.36	78.42	81.10	2	2	2
128	39.96	45.48	48.58	67.01	72.15	75.74	76.47	80.65	84.02	2	2	2
256	41.06	**45.91**	**49.23**	**67.16**	**73.09**	**76.35**	**77.22**	**81.60**	**84.70**	2	2	2
512	39.44	44.88	48.73	66.39	72.29	75.74	76.50	81.44	84.69	2	2	2

article and limit the number of words to the desired lengths. Since the pretrained CLIP model accepts a maximum of 77 word tokens, we modify the original model by zero-padding the pretrained positional embeddings for the additional word positions and finetune the entire set of positional embeddings during training.

In the latter objective, we split each article into sentences and limit the number of sentences by the selected number of words. In Table 3, we observe that using more input words generally helps to improve retrieval accuracy.

Interestingly, the best retrieval performance is obtained when 256 input word tokens are used. One possible reason is that there are more redundant sentences, which do not really contribute to the overall semantics of the story. The performance drop when 512 word tokens are used also suggests that a solution to this problem has to be able to better filter out non-salient sentences.

C Algorithm for selecting multiple images for text narratives

More often than not, a journalist seeking to obtain suitable images to visually illustrate their article may not be able to find suitable images that are already grouped into sets. With this in mind, we present a general algorithm below that allows an author to search for individual candidate images before grouping them and using our finetuned models to select the best set.

```
Given a text narrative, extract a set of named entities $E = \{e_1, \cdots, e_n\}$
images ← set()
```

Table 3. Ablation of MIL-SIM over the length of input text sequences on zero-shot evaluations of article-to-image-set retrieval approaches on the GoodNews [1] dataset. Each test split has 3, 4, or 5 images in each article.

Input Text	R@1			R@5			R@10			Median Rank		
Length	3	4	5	3	4	5	3	4	5	3	4	5
32	26.36	26.18	25.46	49.53	47.64	46.96	59.06	57.10	56.47	6	7	7
64	27.73	27.82	27.13	50.87	50.19	49.85	61.29	59.16	59.02	5	6	7
128	28.23	29.89	28.86	51.76	52.90	51.72	62.24	62.28	60.96	5	5	6
256	**29.42**	**30.59**	**30.23**	**52.07**	49.82	51.44	**60.51**	**61.73**	**62.58**	**4**	**4**	**5**
512	28.26	28.14	27.06	50.97	50.54	49.27	59.26	59.94	58.52	5	5	6

```
for k in range(n) do
    look up a suitable image in an image database using $e_k$
    images.append($I_k$) where $I_k$ is the looked-up image
end for
Generate all possible combinations C of images of size X where X is the desired number of images
$max\_similarity \leftarrow$ inf
best set $\leftarrow None$
for k in range(C) do
    compute similarity between text representation and image set representations $score$
    if $score > max\_similarity$ then
        $max\_similarity = score$
        best set $\leftarrow C_k$
    end if
end for
return best set
```

D Retrieval results using the image search algorithm

We present visual examples of the top-ranked image sets for randomly selected articles from the GoodNews dataset using the proposed image set search algorithm and the finetuned MIL-SIM model. For each detected named entity in an article, we search for suitable images using Google Search and select the top 5 results as candidates images. To provide a basis of comparison, we put ourselves in the shoes of a journalist and create image sets for the selected articles using a naive approach. Specifically, we randomly select 5 named entities present in each article and select the top-ranked image returned by the Google Search API for each named entity. While this is a simplistic alternative approach, it can be easily adopted by any journalist without considering one's aesthetic preferences. Compared to these randomly selected image sets, the image sets selected by the

MIL-SIM model appear to be more visually descriptive. It is possible that the image sets may be more diverse if a user increases the number of candidate images for each query named entity. However, there is a trade-off between increasing the diversity of the image sets and inference time since the time required to compute the combinations of images increases significantly with more candidate images.

Finally, we conduct a qualitative analysis of the correspondences between individual images in a set and specific sentences in an article. Recall that the MIL-SIM approach relies on the assumption that each image should be related to at least one sentence in an article, even if their relationship is loose and illustrative at best rather than literal. We compute the similarity scores between each sentence in an article and the image set selected by the MIL-SIM model. Each image and its best matching sentence are outlined and highlighted in the same color.

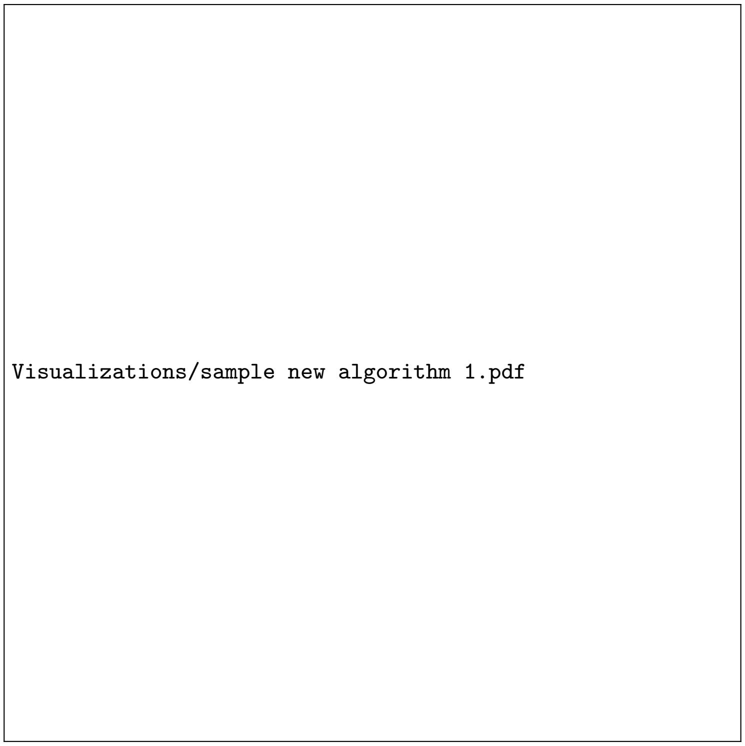
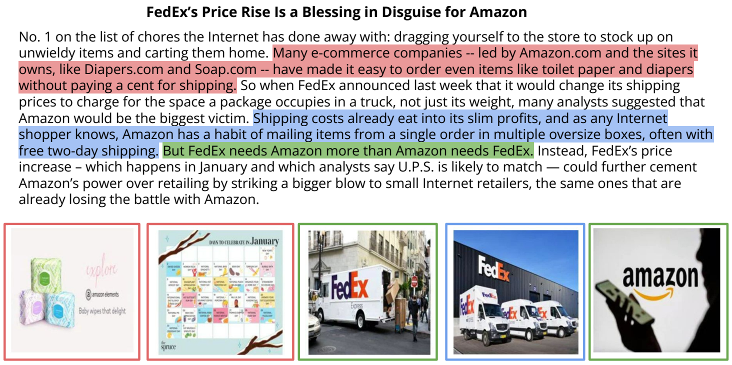


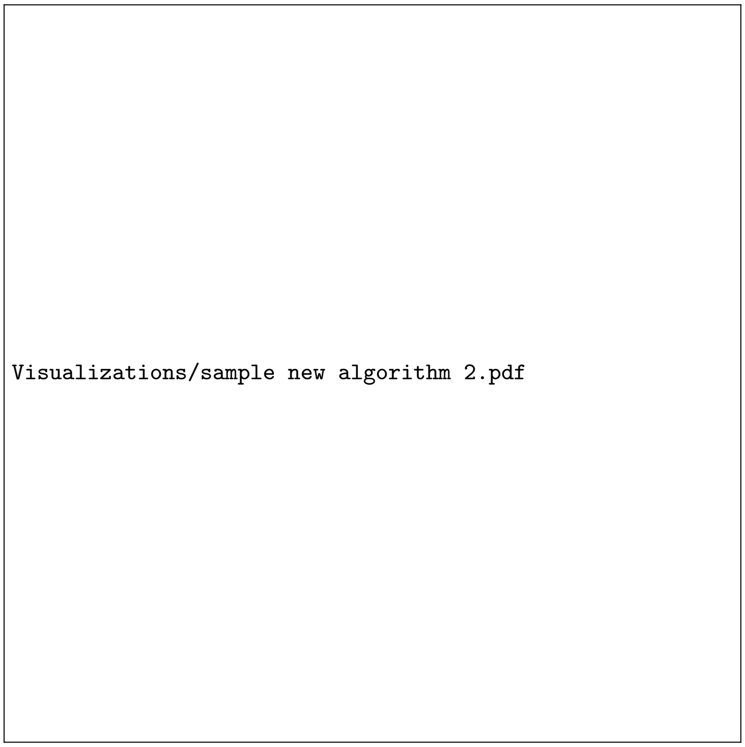
Fig. 3. Example of retrieved image sets given a query article using our algorithm.



FedEx's Price Rise Is a Blessing in Disguise for Amazon

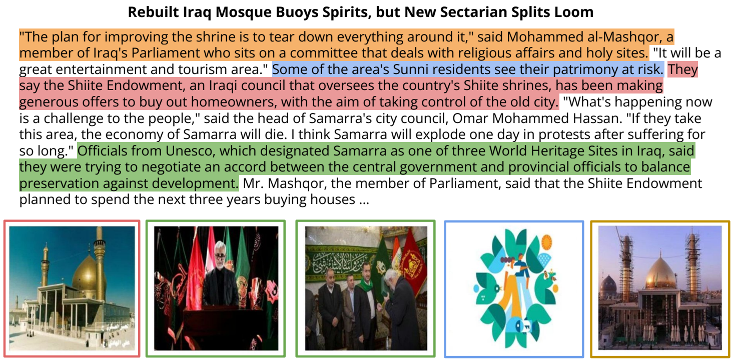
No. 1 on the list of chores the Internet has done away with: dragging yourself to the store to stock up on unwieldy items and carting them home. Many e-commerce companies -- led by Amazon.com and the sites it owns, like Diapers.com and Soap.com -- have made it easy to order even items like toilet paper and diapers without paying a cent for shipping. So when FedEx announced last week that it would change its shipping prices to charge for the space a package occupies in a truck, not just its weight, many analysts suggested that Amazon would be the biggest victim. Shipping costs already eat into its slim profits, and as any Internet shopper knows, Amazon has a habit of mailing items from a single order in multiple oversize boxes, often with free two-day shipping. But FedEx needs Amazon more than Amazon needs FedEx. Instead, FedEx's price increase – which happens in January and which analysts say U.P.S. is likely to match — could further cement Amazon's power over retailing by striking a bigger blow to small Internet retailers, the same ones that are already losing the battle with Amazon.

Fig. 4. Example of the most relevant sentences for each image in the top-ranked image set, as determined by the MIL-SIM model.



Visualizations/sample new algorithm 2.pdf

Fig. 5. Example of retrieved image sets given a query article using our algorithm.



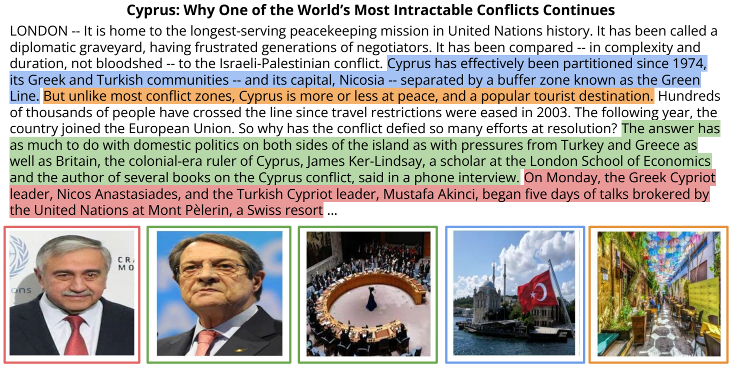
Rebuilt Iraq Mosque Buoys Spirits, but New Sectarian Splits Loom

"The plan for improving the shrine is to tear down everything around it," said Mohammed al-Mashqor, a member of Iraq's Parliament who sits on a committee that deals with religious affairs and holy sites. "It will be a great entertainment and tourism area." Some of the area's Sunni residents see their patrimony at risk. They say the Shiite Endowment, an Iraqi council that oversees the country's Shiite shrines, has been making generous offers to buy out homeowners, with the aim of taking control of the old city. "What's happening now is a challenge to the people," said the head of Samarra's city council, Omar Mohammed Hassan. "If they take this area, the economy of Samarra will die. I think Samarra will explode one day in protests after suffering for so long." Officials from Unesco, which designated Samarra as one of three World Heritage Sites in Iraq, said they were trying to negotiate an accord between the central government and provincial officials to balance preservation against development. Mr. Mashqor, the member of Parliament, said that the Shiite Endowment planned to spend the next three years buying houses ...

Fig. 6. Example of the most relevant sentences for each image in the top-ranked image set, as determined by the MIL-SIM model.

Visualizations/sample new algorithm 3.pdf

Fig. 7. Example of retrieved image sets given a query article using our algorithm.



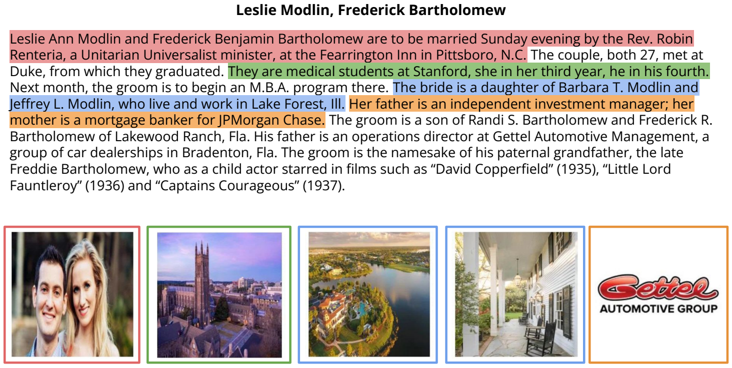
Cyprus: Why One of the World's Most Intractable Conflicts Continues

LONDON -- It is home to the longest-serving peacekeeping mission in United Nations history. It has been called a diplomatic graveyard, having frustrated generations of negotiators. It has been compared -- in complexity and duration, not bloodshed -- to the Israeli-Palestinian conflict. Cyprus has effectively been partitioned since 1974, its Greek and Turkish communities -- and its capital, Nicosia -- separated by a buffer zone known as the Green Line. But unlike most conflict zones, Cyprus is more or less at peace, and a popular tourist destination. Hundreds of thousands of people have crossed the line since travel restrictions were eased in 2003. The following year, the country joined the European Union. So why has the conflict defied so many efforts at resolution? The answer has as much to do with domestic politics on both sides of the island as with pressures from Turkey and Greece as well as Britain, the colonial-era ruler of Cyprus, James Ker-Lindsay, a scholar at the London School of Economics and the author of several books on the Cyprus conflict, said in a phone interview. On Monday, the Greek Cypriot leader, Nicos Anastasiades, and the Turkish Cypriot leader, Mustafa Akinci, began five days of talks brokered by the United Nations at Mont Pèlerin, a Swiss resort ...

Fig. 8. Example of the most relevant sentences for each image in the top-ranked image set, as determined by the MIL-SIM model.

Visualizations/sample new algorithm 4.pdf

Fig. 9. Example of retrieved image sets given a query article using our algorithm.



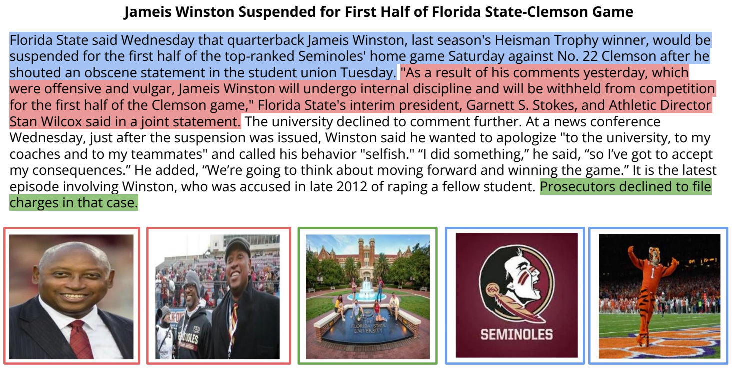
Leslie Modlin, Frederick Bartholomew

Leslie Ann Modlin and Frederick Benjamin Bartholomew are to be married Sunday evening by the Rev. Robin Renteria, a Unitarian Universalist minister, at the Fearrington Inn in Pittsboro, N.C. The couple, both 27, met at Duke, from which they graduated. They are medical students at Stanford, she in her third year, he in his fourth. Next month, the groom is to begin an M.B.A. program there. The bride is a daughter of Barbara T. Modlin and Jeffrey L. Modlin, who live and work in Lake Forest, Ill. Her father is an independent investment manager; her mother is a mortgage banker for JPMorgan Chase. The groom is a son of Randi S. Bartholomew and Frederick R. Bartholomew of Lakewood Ranch, Fla. His father is an operations director at Gettel Automotive Management, a group of car dealerships in Bradenton, Fla. The groom is the namesake of his paternal grandfather, the late Freddie Bartholomew, who as a child actor starred in films such as "David Copperfield" (1935), "Little Lord Fauntleroy" (1936) and "Captains Courageous" (1937).

Fig. 10. Example of the most relevant sentences for each image in the top-ranked image set, as determined by the MIL-SIM model.

Visualizations/sample new algorithm 5.pdf

Fig. 11. Example of retrieved image sets given a query article using our algorithm.



Jameis Winston Suspended for First Half of Florida State-Clemson Game

Florida State said Wednesday that quarterback Jameis Winston, last season's Heisman Trophy winner, would be suspended for the first half of the top-ranked Seminoles' home game Saturday against No. 22 Clemson after he shouted an obscene statement in the student union Tuesday. "As a result of his comments yesterday, which were offensive and vulgar, Jameis Winston will undergo internal discipline and will be withheld from competition for the first half of the Clemson game," Florida State's interim president, Garnett S. Stokes, and Athletic Director Stan Wilcox said in a joint statement. The university declined to comment further. At a news conference Wednesday, just after the suspension was issued, Winston said he wanted to apologize "to the university, to my coaches and to my teammates" and called his behavior "selfish." "I did something," he said, "so I've got to accept my consequences." He added, "We're going to think about moving forward and winning the game." It is the latest episode involving Winston, who was accused in late 2012 of raping a fellow student. Prosecutors declined to file charges in that case.

Fig. 12. Example of the most relevant sentences for each image in the top-ranked image set, as determined by the MIL-SIM model.

Image set retrievals for articles

D.1 Additional qualitative retrieval results

In this section, we provide additional qualitative retrieval results obtained by our best-performing MIL-SIM approach on our test splits of the NewsStories and GoodNews datasets. For each query article, we show the ground-truth set of corresponding images as well as the top five retrieved image sets. In the visualizations, we also provide the corresponding article titles that correspond to the retrieved image sets. The ground-truth and incorrect image sets are also outlined in green and red boxes, respectively. The retrieved image sets for a query article are ordered from top to bottom.

References

1. Biten, A.F., Gomez, L., Rusinol, M., Karatzas, D.: Good news, everyone! context driven entity-aware captioning for news images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12466–12475 (2019)

`Visualizations/updated correct newsstories retrievals 1.pdf`

Fig. 13. Correct retrievals on the test split of the NewsStories dataset.

`Visualizations/updated correct newsstories retrievals 2.pdf`

Fig. 14. Correct retrievals on the test split of the NewsStories dataset.

Visualizations/updated correct newsstories retrievals 3.pdf

Fig. 15. Correct retrievals on the test split of the NewsStories dataset.

Visualizations/updated correct newsstories retrievals 4.pdf

Fig. 16. Correct retrievals on the test split of the NewsStories dataset.

Visualizations/updated correct newsstories retrievals 5.pdf

Fig. 17. Correct retrievals on the test split of the NewsStories dataset.

Visualizations/updated incorrect newsstories retrievals 1.pdf

Fig. 18. Incorrect retrievals on the test split of the NewsStories dataset.

Visualizations/updated incorrect newsstories retrievals 2.pdf

Fig. 19. Incorrect retrievals on the test split of the NewsStories dataset.

Visualizations/updated incorrect newsstories retrievals 3.pdf

Fig. 20. Incorrect retrievals on the test split of the NewsStories dataset.

```
Visualizations/updated incorrect newsstories retrievals 4.pdf
```

Fig. 21. Incorrect retrievals on the test split of the NewsStories dataset.

`Visualizations/updated incorrect newsstories retrievals 5.pdf`

Fig. 22. Incorrect retrievals on the test split of the NewsStories dataset.

Visualizations/updated correct goodnews retrievals 1.pdf

Fig. 23. Correct retrievals on the GoodNews dataset.

Visualizations/updated correct goodnews retrievals 2.pdf

Fig. 24. Correct retrievals on the GoodNews dataset.

Visualizations/updated correct goodnews retrievals 3.pdf

Fig. 25. Correct retrievals on the GoodNews dataset.

Visualizations/updated correct goodnews retrievals 4.pdf

Fig. 26. Correct retrievals on the GoodNews dataset.

Visualizations/updated correct goodnews retrievals 5.pdf

Fig. 27. Correct retrievals on the GoodNews dataset.

`Visualizations/updated incorrect goodnews retrievals 1.pdf`

Fig. 28. Incorrect retrievals on the GoodNews dataset.

`Visualizations/updated incorrect goodnews retrievals 2.pdf`

Fig. 29. Incorrect retrievals on the GoodNews dataset.

Visualizations/updated incorrect goodnews retrievals 3.pdf

Fig. 30. Incorrect retrievals on the GoodNews dataset.

Visualizations/updated incorrect goodnews retrievals 4.pdf

Fig. 31. Incorrect retrievals on the GoodNews dataset.

Visualizations/updated incorrect goodnews retrievals 5.pdf

Fig. 32. Incorrect retrievals on the GoodNews dataset.