Title	CURED4NLG: A dataset for table-to-text generation
Author(s)	Pasricha, Nivranshu; Arcan, Mihael; Buitelaar, Paul
Publication Date	2023
Publication Information	Pasricha, Nivranshu, Arcan, Mihael, & Buitelaar, Paul. (2023). CURED4NLG: A dataset for table-to-text generation. Paper presented at the 4th Conference on Language, Data and Knowledge, Vienna, Austria, 12-15 September. https://doi.org/10.13025/szze-8850
Publisher	University of Galway
Link to publisher's version	https://doi.org/10.13025/szze-8850
Item record	http://hdl.handle.net/10379/17955
DOI	http://dx.doi.org/10.13025/szze-8850

CURED4NLG: A Dataset for Table-to-Text Generation

Nivranshu Pasricha[1], **Mihael Arcan**[2] and **Paul Buitelaar**[1,2]
[1]SFI Centre for Research Training in Artificial Intelligence
[2] Insight SFI Research Centre for Data Analytics
Data Science Institute, University of Galway
n.pasricha1@nuigalway.ie

Abstract

We introduce CURED4NLG, a dataset for the task of table-to-text generation focusing on the public health domain. The dataset consists of 280 pairs of tables and documents extracted from weekly epidemiological reports published by the World Health Organisation (WHO). The tables report the number of cases and deaths from COVID-19, while the documents describe global and regional updates in English text. Along with the releasing the dataset, we present outputs from three different baselines for the task of table-to-text generation. The first is based on a manually defined template and the other two on end-to-end transformer-based models. Our results suggest that end-to-end models can learn a template-like structure of the reports to produce fluent sentences, but may contain many factual errors especially related to numerical values.

1 Introduction

Data-to-text generation systems aim to produce meaningful texts in a human language from non-linguistic representation of information such as tables or graphs in the input (Reiter and Dale, 2000). Traditionally, such systems have been designed using a rule-based approach relying on a modular pipeline architecture and have included applications in domains such as weather reporting (Goldberg et al., 1994), sports (Robin, 1995; Tanaka-Ishii et al., 1998) and healthcare (Binsted et al., 1995; Cawsey et al., 1997). Recently, there has been increasing interest in end-to-end approaches for data-to-text generation with neural encoder-decoder architectures. To aid further research in this direction, a number of datasets have been released in the last few years with different input data structures covering various domains. Examples include WIKIBIO (Lebret et al., 2016), ROTOWIRE (Wiseman et al., 2017), WebNLG (Gardent et al., 2017), E2E (Novikova et al., 2017), ToTTo (Parikh et al., 2020) and DART (Nan et al., 2021).

A popular strategy applied to data-to-text generation tasks is to split the problem along two fundamental axes aiming to answer the questions, *what to say?* (content determination) and *how to say it?* (microplanning and linguistic realisation). Datasets such as WebNLG, E2E and DART are only concerned with the planning and realisation aspects and do away with content selection aspect of the task. A more recent dataset, ToTTo includes content selection explicitly by highlighting relevant cells in the input table. However, the output for ToTTo is usually one or two sentences which is typically easier to generate compared to a document.

We present **CURED4NLG**[1] (COVID-19 **U**pdate **R**eports from **E**pidemiological **D**ata for **N**atural **L**anguage **G**eneration), a dataset for table-to-text generation, where the input data is structured in the in form of a table, typically comprising of 6 to 60 rows with 7 to 9 columns (see Table 1). Each table reports the number of new cases of COVID-19 and related deaths during a week-long time period along with cumulative totals recorded since the start of the pandemic. A document corresponding to each table describes the important information contained in the table in about 200 – 300 words in English as shown in Figure 1. Hence, the goal of the table-to-text generation task is to automatically generate an output document describing the data in the input table. With CURED4NLG, we aim to enrich research in table-to-text generation with the goal of generating documents longer than one sentence in the output conditioned on structured input data while also addressing the issues related to content determination. We present outputs and results from two baseline models, based on end-to-end approaches, and compare them with a template-based system. Initial results suggest that end-to-end models are able to generate fluent outputs but can struggle to generate sentences which are faithful to the input tables.

[1]http://github.com/cured4nlg/cured4nlg

WHO Region	New cases in last 7 days (%)	Change in new cases	Cumulative cases (%)	New deaths in last 7 days (%)	Change in new deaths	Cumulative deaths (%)
Europe	1989636 (54%)	11%	13144973 (26%)	25531 (47%)	44%	311542 (25%)
Americas	1031573 (28%)	3%	21509104 (43%)	17289 (32%)	<1%	656629 (53%)
South-East Asia	390157 (11%)	2%	9641945 (19%)	5132 (9%)	10%	149326 (12%)
Eastern Mediterranean	214072 (6%)	18%	3307411 (7%)	5675 (10%)	23%	84305 (7%)
Africa	33687 (1%)	2%	1357945 (3%)	831 (2%)	30%	30616 (2%)
Western Pacific	31370 (1%)	19%	765197 (2%)	377 (1%)	-5%	15942 (1%)
Global	3690495 (100%)	8%	49727316 (100%)	54835 (100%)	21%	1248373 (100%)

In the past week, the global number of cases of COVID-19 has increased by 8% compared to the previous week, totalling more than 3.6 million new cases, while new deaths have increased by 21% to over 54000. This brings the cumulative numbers to over 49.7 million reported cases and over 1.2 million deaths globally since the start of the pandemic. The European Region continues to account for the greatest proportion of new cases and deaths in the past 7 days, the Region reported over half (54%) of all new cases and nearly half (47%) of new deaths. Although it still accounts for only 2% of the global total number of cases and deaths, this week the Western Pacific Region showed the largest relative proportional increase in new cases (19%) compared to the previous week followed by the Eastern Mediterranean Region (18%) and the European region (11%). The three regions reporting the highest proportional increases in newly reported deaths in the past 7 days compared to the previous week are Europe (44%), Africa (30%) and the Eastern Mediterranean (23%). The Western Pacific Region was the only region to report a decrease in deaths (5%) this week compared to the previous week.

Figure 1: Example of a table and corresponding epidemiological report from the CURED4NLG dataset.

2 Related Work

Natural language generation (NLG) in the healthcare domain has seen significant interest over the years (Cawsey et al., 1997; Pauws et al., 2019). Applications here usually involve generating personalised reports or medical explanations for individual patients (Binsted et al., 1995; McKeown et al., 1997; Mahamood and Reiter, 2011) and typically are not concerned with mass communication of general public health advice. However, during the COVID-19 pandemic, public dashboards (Ritchie et al., 2020; Dong et al., 2020; Wissel et al., 2020) became immensely popular for communicating information about the spread of the disease globally. These dashboards rely on visuals such as maps and charts but do not usually provide textual updates. An exception to this is a dashboard[2] by Microsoft and Arria NLG reporting automatically generated narratives describing the number of cases and deaths for COVID-19 along with an interactive map (Reiter and Sripada, 2020). Tangential to this, automatic generation of data-driven narratives for mass communication of news (Leppänen et al., 2017) and automated journalism (Graefe, 2016) have also received significant interest over the last few years. However, since most of these systems for automatic report generation are built in-house by private organisations, the details about the underlying architecture and the actual data used are usually not publicly available (Dale, 2020). With CURED4NLG we hope to motivate research in this domain with a publicly available dataset.

In terms of the structure of the input and output data, ROTOWIRE (Wiseman et al., 2017) can be considered most similar to CURED4NLG among existing NLG datasets. ROTOWIRE consists of about 3,000 basketball box-scores paired with descriptive summaries and is one of best examples of what a real-world application of data-to-text generation might look like. However, it has some significant challenges associated with it. For instance, Wang (2019) observed that only 60% of the output textual summary content can be grounded to the boxscore data. This misalignment leads to hallucinations where a model generates a set of unconditioned random statements that are unfaithful to the input. Thomson et al. (2020) also observed data partition contamination issues where boxscore data from

[2] https://www.arria.com/covid19-microsoft/

	Min.	Max.	Average
Columns	7	9	7.86
Rows	7	62	33.28
Cells	49	496	265.28
Document Length	63	643	249.75

Table 1: Number of rows, columns, cells and document length (number of words) in the CURED4NLG dataset.

some games ended up in both training set as well as test/validation set. Another issue highlighted in their analysis is that random partition of the data ignores the inherent temporal dimension in the data leading to further hallucinations.

3 CURED4NLG Dataset

The CURED4NLG dataset is created from 40 epidemiological update reports published by WHO and consists of 280 pairs of tables and documents. Since August 2020, an update report has been published on the WHO website[3] once a week in PDF format to provide an overview of the global and regional situation for COVID-19. Each weekly update highlights key data and trends as well as other pertinent epidemiological information concerning the pandemic. We extract the tables from Annex 1 of the PDF reports using optical character recognition (OCR) followed by a manual verification step to fix formatting and spelling errors. The resulting tabular data is saved as a file with tab-separated values, while the corresponding update reports are stored as plaintext files. Some texts include additional information about patient demographics and regional restrictions as well as references to charts and figures elsewhere in the report. Such sentences go beyond the data in the tables, hence, we filter these out and create a *cleaned* version of the CURED4NLG dataset.

The dataset is split into training, validation and test sets such that the inherent temporal aspects of the data are maintained. Data from the first 30 reports is used for training, data from the next three weeks is used for validation and the data from the five most recent weekly reports is taken to be the test set. Each update report consists of a global table along with six regional tables, hence, the training set, validation set, and the test set contain 210, 21 and 49 instances respectively (see Table 2).

[3] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

	train	valid	test
Examples	210	21	49
Tokens	56250	4103	9555
Types	3711	478	869
Avg. Doc. Length	267.9	195.4	195.0
cleaned			
Tokens	43526	4091	9506
Types	2243	476	869
Avg. Doc. Length	207.3	194.8	194.0

Table 2: Number of examples, tokens and types for text documents in the CURED4NLG dataset.

Compared to ROTOWIRE (Wiseman et al., 2017), this dataset is smaller by an order of magnitude in size. It is also much smaller than other NLG datesets which usually consist of several thousands examples. Nonetheless, CURED4NLG can be useful for data-to-text generation tasks as it is representative of a real-world application scenario for NLG and presents an opportunity to focus on the various challenges such as content selection, document planning and linguistic realisation. One limitation of this dataset might be that the sentence structure is simple in most instances and there is minimal linguistic variation in the texts. Despite that, we find state-of-the-art end-to-end NLG systems struggle to outputs with high accuracy and this dataset can be useful in studying the limitations of such systems. Since this dataset is created from weekly reports by WHO, it includes an additional challenge of working with data that contains an inherent temporal dimension which might be difficult to model using end-to-end techniques.

Since June 2021, WHO stopped publishing the tables containing detailed case statistics in the weekly epidemiological reports. The reports published since then only contain an update in the form of texts while the tables are available on the online WHO portal[4]. Hence the number of new cases and deaths reported in the tables do not always exactly match the figures reported in the text of recent weekly epidemiological reports. It is due to this reason the data in CURED4NLG is limited until May 2021. However, we plan to further extend this dataset, with data until 2023 by manually verifying the numbers reported across the tables and the texts, and aligning them correctly, where needed.

[4] https://covid19.who.int/data

The region of *Europe* reported over *1.4 million* new cases and *25000* new deaths this week, a *10% decrease* and a *4% decrease* respectively compared to the previous week. The highest numbers of new cases were reported from *Turkey* (*378771* new cases; *449.1* new cases per 100000 population; a *9% decrease*), *France* (*211674* new cases; *325.5* new cases per 100000 population; a *9% decrease*) and *Germany* (*145156* new cases; *174.5* new cases per 100000 population; a *1% increase*). The highest numbers of new deaths were reported from *Poland* (*3383* new deaths; *8.9* new deaths per 100000 population; a *6% decrease*), *Russian Federation* (*2650* new deaths; *1.8* new deaths per 100000 population; a *2% increase*) and *Ukraine* (*2537* new deaths; *5.8* new deaths per 100000 population; a *8% decrease*).	In the past week, the *European Region* reported over *1466000* new cases and over *25000* new deaths, a *decrease* of 1% and an increase of 1% respectively compared to the previous week. The three countries reporting the highest numbers of new cases were Kosovo (*2662* new cases; 57 new cases per 100000; a 1% decrease), *Turkey* (*378771* new cases; 57 new cases per 100000; a 1% *decrease*), *France* (*211674* new cases; 158.8 new cases per 100000; a 7% *decrease*). The three countries reporting the highest numbers of new deaths this week were the United Kingdom (*157* new deaths; 3.4 new deaths per 100000; a 3% *decrease*), Germany (*1650* new deaths; 3.4 new deaths per 100000; a 3% *decrease*), the *Russian Federation* (*2650* new deaths; 3.7 new deaths per 100000; a 3% *decrease*) and the *Russian Federation* (2345 new deaths; 3.4 new deaths per 100000; a 3% decrease).

Table 3: Example of an output epidemiological report for the European region generated by the template baseline (left) and the T5 model (right). Text in blue italics shows information filled in from the input table by the baseline template. Text in green italics shows tabular values correctly produced by the T5 model while underlined text in red shows the mistakes. Outputs from all end-to-end trained baselines for this example are presented in Appendix A.3.

4 Baselines

We present baseline results for the task of table-to-text generation with CURED4NLG using two different approaches – a templated baseline and two transformer-based encoder-decoder models. The overall task is be defined as follows:

Given a set of one or more tables in the input, generate a text document in English in the output describing the tabular data.

Template baseline: We define a global and a regional template to generate an epidemiological report based on input tabular data. The template for the global report includes sentences describing new and cumulative totals of cases and deaths for COVID-19 along with changes in trends from the week prior. The template also generates sentences describing the most affected continental region as well as the five most affected countries globally. Similarly, the template for a regional report describes new numbers as well as the change in numbers from the previous week, followed by a sentence describing the three most affected countries in a specific region. The exact templates used to generate the output documents are defined in Appendix A.1.

End-to-End baselines: We use the hierarchical model (Rebuffel et al., 2020) as one of the end-to-end baseline models. It is designed for data-to-text tasks and follows a two-level encoder-decoder architecture for modeling structured data in the input. We use the state-of-the-art T5 model (Raffel et al., 2020) as another end-to-end neural baseline. It is based on the transformer architecture (Vaswani et al., 2017) and pre-trained on the "Colossal Clean Crawled Corpus" using a masked language modelling objective. Since the T5 architecture expects the input to be encoded as a sequence of text, we linearise the input table by concatenating all the rows into a single sequence. The rows in each table are arranged in decreasing order of number of new cases by default.

To assess the performance of the end-to-end baseline systems on content selection, we perform an experiment where we randomly shuffle the rows of the table to see how well the transformer-based models pay attention to the relative positioning of the rows in the input table. We perform another experiment where we include only a subset of the first ten rows in the input and evaluate the model performance. And as another experiment we train with the *cleaned* version of the CURED4NLG dataset.

	BLEU (↑)	METEOR (↑)	TER (↓)	PARENT Precision (↑)	Recall (↑)	F1 (↑)
Template baseline	64.48	41.76	32.19	76.55	19.93	29.97
Hierarchical model	29.86	27.64	67.49	43.10	17.65	22.80
T5 (no pre-training)	20.31	18.47	99.55	41.07	8.38	12.24
T5 (pre-trained)	43.32	32.77	52.10	56.38	17.15	24.68
+ *shuffled*	41.16	31.67	49.89	56.07	14.75	21.97
+ *subset*	42.99	33.33	55.58	56.75	18.73	26.13
+ *cleaned*	44.57	33.37	49.85	57.07	17.35	25.05

Table 4: Results for baselines on the CURED4NLG dataset.

The details for training the end-to-end baseline models along with the chosen hyperparameter values for each model are described in Appendix A.2.

All the code for training, generating and evaluating the baseline models along with the generated outputs is available to download at `https://github.com/cured4nlg/cured4nlg`.

5 Results and Discussion

We report results on the outputs generated from the baselines using four automatic evaluation metrics, BLEU (Papineni et al., 2002), METEOR (Denkowski and Lavie, 2014), TER (Snover et al., 2006) and PARENT (Dhingra et al., 2019) as shown in Table 4. The first three are popular metrics used for measuring lexical similarity between generations and references while PARENT is a recently proposed metric specifically for table-to-text evaluation as it computes precision and recall for n-grams in generated and reference texts aligned to table data.

We find the template baseline to outperform the end-to-end models across all the automatic evaluation metrics. Earlier reports published by the World Health Organization in 2020 contained more varied text, however, reports published since March 2021 appear to follow a template-like structure. Since the validation and test sets exclusively contain data from this period, because the dataset was split in such way that the inherent temporal dimension of the data remains intact, we observe high scores across the automatic evaluation metrics with the template baseline.

We observe that the end-to-end baseline models are able to generate fluent outputs by learning the template-like sentence structure but contain many factual errors as shown in Table 3. The pre-trained T5 model performs better than the hierarchical baseline on the metrics measuring lexical similarity as well as the precision score. However, the hierarchical model achieves a similar recall score. We further observe that shuffling the rows in the table leads to worse performance for the T5 model as it is makes the task more difficult. However, we observe slight improvements in the scores with the *cleaned* version of the dataset and further notice improvements in recall and F1 scores when only a subset of the top 10 rows is considered. This suggests that the model struggles to perform content selection, especially for larger tables.

A limitation of the PARENT metric is that it cannot detect paraphrases accurately. In almost every gold-standard reference of the CURED4NLG dataset, large numbers are either written in words or rounded to nearest thousand in text while the tables contain exact numerical values. For example, in Table 1, the number of new cases reported in the input table is `3690495`, while the reference text report describes this value as "*more than 3.6 million*". To account for this and other errors related to the accuracy of the generated texts, we manually count the number of errors in the outputs of the hierarchical model and the pre-trained T5 model on a subset of 21 examples from the test set. We use the same error categories of incorrect *Number* (for numerical values), *Name* (for region names) and *Word* (for words such as increase, decrease, rise, decline, etc.) as defined by Thomson and Reiter (2020). The rest of the errors are classified in the *Other* category. We find outputs from both models contain about 20 – 25 errors on average with most of the errors being associated with numerical values as shown in Table 5. Further work is required in designing error annotation guidelines specific to the CURED4NLG dataset as well as evaluation strategies which can identify paraphrasing of numbers.

Error Category	Hierarchical Total	Avg.	T5 Total	Avg.
Number	346	16.5	294	14.0
Name	63	3.0	34	1.6
Word	85	4.0	72	3.4
Other	9	0.4	6	0.3
Total	503	23.9	406	19.3

Table 5: Counts of errors in the outputs generated by end-to-end baselines on a subset of 21 examples.

6 Conclusion

We introduced CURED4NLG, a dataset for table-to-text generation which can be useful as a benchmark for data-to-text generation. Initial baseline results suggest that end-to-end text generation models can learn a template-like structure of the documents to generate fluent outputs but at the same time are prone to hallucinating and generating erroneous statements particularly related to numerical values.

Acknowledgments

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223 and co-supported by Science Foundation Ireland under grant number SFI/12/RC/2289 2 (Insight), co-funded by the European Regional Development Fund.

References

Kim Binsted, Alison Cawsey, and Ray Jones. 1995. Generating personalised patient information using the medical record. In *Artificial Intelligence in Medicine*, pages 29–41, Berlin, Heidelberg. Springer Berlin Heidelberg.

Alison J. Cawsey, Bonnie L. Webber, and Ray B. Jones. 1997. Natural Language Generation in Health Care. *Journal of the American Medical Informatics Association*, 4(6):473–482.

Robert Dale. 2020. Natural language generation: The commercial state of the art in 2020. *Natural Language Engineering*, 26(4):481–487.

Michael Denkowski and Alon Lavie. 2014. Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 376–380, Baltimore, Maryland, USA. Association for Computational Linguistics.

Bhuwan Dhingra, Manaal Faruqui, Ankur Parikh, Ming-Wei Chang, Dipanjan Das, and William Cohen. 2019. Handling divergent reference texts when evaluating table-to-text generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4884–4895, Florence, Italy. Association for Computational Linguistics.

Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. The WebNLG challenge: Generating text from RDF data. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 124–133, Santiago de Compostela, Spain. Association for Computational Linguistics.

E. Goldberg, N. Driedger, and R.I. Kittredge. 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2):45–53.

Andreas Graefe. 2016. Guide to automated journalism. Technical report, Tow Center for Digital Journalism, Columbia University.

Rémi Lebret, David Grangier, and Michael Auli. 2016. Neural text generation from structured data with application to the biography domain. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1203–1213.

Leo Leppänen, Myriam Munezero, Mark Granroth-Wilding, and Hannu Toivonen. 2017. Data-driven news generation for automated journalism. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 188–197, Santiago de Compostela, Spain. Association for Computational Linguistics.

Saad Mahamood and Ehud Reiter. 2011. Generating affective natural language for parents of neonatal infants. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 12–21, Nancy, France. Association for Computational Linguistics.

Kathleen R. McKeown, Desmond A. Jordan, Shimei Pan, James Shaw, and Barry A. Allen. 1997. Language generation for multimedia healthcare briefings. In *Fifth Conference on Applied Natural Language Processing*, pages 277–282, Washington, DC, USA. Association for Computational Linguistics.

Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau, Abhinand Sivaprasad, Chiachun Hsieh, Xiangru Tang, Aadit Vyas, Neha Verma, Pranav Krishna, Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz Rahman, Ahmad Zaidi, Mutethia Mutuma, Yasin Tarabar, Ankit Gupta, Tao Yu, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher, and Nazneen Fatema Rajani. 2021. DART: Open-domain structured data record to text generation. In

Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 432–447, Online. Association for Computational Linguistics.

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017. The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206, Saarbrücken, Germany. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Ankur P Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. 2020. ToTTo: A controlled table-to-text generation dataset. In *Proceedings of EMNLP*.

Steffan Pauws, Albert Gatt, Emiel Krahmer, and Ehud Reiter. 2019. Making effective use of healthcare data using data-to-text technology. In *Data Science for Healthcare: Methodologies and Applications*, pages 119–145. Springer.

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.

Clément Rebuffel, Laure Soulier, Geoffrey Scoutheeten, and Patrick Gallinari. 2020. A hierarchical model for data-to-text generation. In *Advances in Information Retrieval*, pages 65–80, Cham. Springer International Publishing.

Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*, 1 edition. Cambridge University Press.

Ehud Reiter and Yaji Sripada. 2020. Using arria nlg to give visual analytics dashboards the power of language. Technical report, Arria NLG.

Hannah Ritchie, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, Cameron Appel, Lucas Rodés-Guirao, and Max Roser. 2020. Coronavirus pandemic (covid-19). *Our World in Data*. Https://ourworldindata.org/coronavirus.

Jacques Pierre Robin. 1995. *Revision-Based Generation of Natural Language Summaries Providing Historical Background: Corpus-Based Analysis, Design, Implementation and Evaluation*. Ph.D. thesis, Columbia University, USA. UMI Order No. GAX95-33653.

Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *In Proceedings of Association for Machine Translation in the Americas*, pages 223–231.

Kumiko Tanaka-Ishii, Koiti Hasida, and Itsuki Noda. 1998. Reactive content selection in the generation of real-time soccer commentary. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*.

Craig Thomson and Ehud Reiter. 2020. A gold standard methodology for evaluating accuracy in data-to-text systems. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 158–168, Dublin, Ireland. Association for Computational Linguistics.

Craig Thomson, Ehud Reiter, and Somayajulu Sripada. 2020. Sportsett: Basketball-a robust and maintainable dataset for natural language generation. In *IntelLanG: Intelligent Information Processing and Natural Language Generation*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*.

Hongmin Wang. 2019. Revisiting challenges in data-to-text generation with fact grounding. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 311–322, Tokyo, Japan. Association for Computational Linguistics.

Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263, Copenhagen, Denmark. Association for Computational Linguistics.

Benjamin D Wissel, P J Van Camp, Michal Kouril, Chad Weis, Tracy A Glauser, Peter S White, Isaac S Kohane, and Judith W Dexheimer. 2020. An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. *Journal of the American Medical Informatics Association*, 27(7):1121–1125.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi 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 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.

A Appendix

A.1 Template Baseline

We define two templates to generate output documents given the tabular data in the input. The first template produces a global epidemiological report based on the overall global regions. The second template produces a and regional epidemiological report based on the data reported for each country in a particular region. There are six regions defined by the WHO, namely, *Americas*, *Europe*, *South-East Asia*, *Eastern Mediterranean*, *Africa*, and *Western Pacific*.

```
A cumulative total of {CUMULATIVE_CASES} cases
and {CUMULATIVE_DEATHS} deaths have been
reported since the start of the outbreak.

The number of new cases {INCREASED/DECREASED} by
{NEW_CASES_CHANGE}% and the number of new deaths
{INCREASED/DECREASED} by {NEW_DEATHS_CHANGE}%
globally in the last 7 days.

The WHO Region of {MOST_AFFECTED_REGION} was the
most affected region with
{MOST_AFFECTED_NEW_CASES} new cases and
{MOST_AFFECTED_NEW_DEATHS} new deaths.

This region noted {INCREASE/DECREASE} of
{MOST_AFFECTED_CASES_CHANGE}% in new cases since
the last week and accounts for
{MOST_AFFECTED_NEW_CASES_SHARE}% of all new
cases.
```

Figure 2: Global Template

The placeholder values inside curly braces are filled in from the input tables. The relative change in the number of new cases and new deaths reported is calculated using the data from the current week and the previous week, and reported as the percentage (%) change. Large numerical values are rounded to the nearest ten thousand, or hundred thousand, and written out in words.

```
The region of {REGION} reported over {NEW_CASES}
new cases and {NEW_DEATHS} new deaths this week,
a {NEW_CASES_CHANGE}% {INCREASE/DECREASE} and a
{NEW_DEATHS_CHANGE}% {INCREASE/DECREASE}
respectively compared to the previous week.

The highest numbers of new cases were reported
from {MOST_AFFECTED_COUNTRIES_BY_CASES}.

The highest numbers of new deaths were reported
from {MOST_AFFECTED_COUNTRIES_BY_DEATHS}.
```

Figure 3: Regional Template

A.2 End-to-End Baselines

Each end-to-end baseline model is trained on a single Nvidia GeForce GTX 1080 Ti GPU for 5,000 steps with the following set up:

- **Hierarchical Model** (Rebuffel et al., 2020): This model consists of a transformer encoder and an LSTM decoder with a hierarchical attention mechanism. We use the same set up and hyperparameter values as described in the original repository[5], except, the number of entities in the encoder is set to 10 here instead of 24 as defined in the original paper. The maximum sequence length is set to 1000 and beam search is applied during inference with beam size equal to 10. It took approximately 8 hours to train this model on a single GPU.
- **T5 Model** (Raffel et al., 2020): We use the implementation of the T5 small model (60M parameters) from the transformers [6] library by Hugging Face (Wolf et al., 2020). The model comprises 6 layers each in the encoder and decoder with a multi-head attention sublayer consisting of 8 attention heads. The word embeddings are 512-dimensional and the fully-connected feed-forward sublayers are 2048-dimensional. Sequence length for input and output is set to 1024. The model is trained with the Adam optimizer with a learning rate of 5×10^{-5}. During inference, beam search is applied with a beam of size 10. All the other hyperparameter values are set to their default values. The training process took about 2 hours with a batch size of 4.

A.3 Additional Output Examples

We present outputs generated by the end-to-end baselines as well the template baselines for three tables from the test set of the CURED4NLG dataset.

Table 6 shows a truncated version of the input table for the European region along with corresponding outputs generated by the end-to-end baseline models. Similarly, Table 7 shows the table and outputs generated for an instance in the test set corresponding to the region of Eastern Mediterranean. Finally, Table 8 shows the an example of a table from the test set of the CURED4NLG along with the global epidemiological reports generated by the hierarchial and the T5 baseline models.

[5] https://github.com/KaijuML/data-to-text-hierarchical

[6] https://huggingface.co/transformers/

Reporting Country/ Territory/Area	New cases in last 7 days	Cumulative cases	Cumulative cases per 100k population	New deaths in last 7 days	Cumulative deaths	Cumulative deaths per 100k population	Transmission classification
Europe	1466680	50714995	5435.3	25341	1061218	113.7	-
Turkey	378771	4591416	5444.0	2403	38011	45.1	Community transmission
France	211674	5390187	8287.6	2110	102031	156.9	Community transmission
Germany	145156	3287418	3952.8	1650	81564	98.1	Community transmission
Italy	92074	3949517	6622.1	2345	119021	199.6	Clusters of cases
Ukraine	78761	2025271	4630.9	2537	42323	96.8	Community transmission
Poland	70831	2758856	7268.2	3383	65415	172.3	Community transmission
Russian Federation	60468	4762569	3263.5	2650	108232	74.2	Clusters of cases
Netherlands	57991	1453058	8347.3	135	17038	97.9	Community transmission
Spain	32476	3456886	7303.4	214	77496	163.7	Community transmission
...	...	...	...	...	...	...	...

The European Region reported over 1.4 million new cases and over 25000 new deaths, a 12% and a 5% decrease respectively compared to the previous week. The trend in both new cases and deaths is decreasing in the last two weeks. The highest numbers of new cases were reported from Turkey (378771 new cases; 449.1 new cases per 100000; a 9% decrease), France (211674 new cases; 325.5 new cases per 100000; a 9% decrease), and Germany (145156 new cases; 174.5 new cases per 100000; a 1% increase). The highest numbers of new deaths were reported from Poland (3383 new deaths; 8.9 new deaths per 100000; a 6% decrease), the Russian Federation (2650 new deaths; 1.8 new deaths per 100000; a 2% increase), and Ukraine (2537 new deaths; 5.8 new deaths per 100000; an 8% decrease).

Reference

The European Region reported over 1.2 million new cases and over 36000 new deaths, a decrease of 18% and 8% respectively when compared to the previous week. The three countries reporting the highest numbers of new cases were Spain (7666 new cases; 160.5 new cases per 100000, a 15% decrease), France (211674 new cases; 113.2 new cases per 100000; a 31% decrease), and France (3958 new cases; 2.4 new cases per 100000, a 119021% increase). The highest numbers of new deaths were reported from the Russian Federation (2650 new deaths; 0.4 new deaths per 100000 population; a 88% decrease), Germany (214 new deaths; 0.1 new deaths per 100000, a 3848% decrease) and France (2110 new deaths; 0.4 new deaths per 100000; a 88% decrease).

Hierarchical Model

In the past week, the European Region reported over 1466000 new cases and over 25000 new deaths, a decrease of 1% and an increase of 1% respectively compared to the previous week. The three countries reporting the highest numbers of new cases were Kosovo (2662 new cases; 57 new cases per 100000; a 1% decrease), Turkey (378771 new cases; 57 new cases per 100000; a 1% decrease), France (211674 new cases; 158.8 new cases per 100000; a 7% decrease). The three countries reporting the highest numbers of new deaths this week were the United Kingdom (157 new deaths; 3.4 new deaths per 100000; a 3% decrease), Germany (1650 new deaths; 3.4 new deaths per 100000; a 3% decrease), the Russian Federation (2650 new deaths; 3.7 new deaths per 100000; a 3% decrease) and the Russian Federation (2345 new deaths; 3.4 new deaths per 100000; a 3% decrease).

T5 (pre-trained)

The European Region reported over 1466000 new cases and over 25000 new deaths, a decrease of 1% and 2% respectively compared to the previous week. The three countries reporting the highest numbers of new cases were Kosovo (2662 new cases; 57 new cases per 100000; a 1% decrease, Turkey (378771 new cases; 57 new cases per 100000; a 1% decrease, France (211674 new cases; 189.8 new cases per 100000; a 1% decrease, and Germany (1650 new cases; 1.9 new cases per 100000; a 7% decrease. The highest numbers of deaths were reported from the Russian Federation (2650 new deaths; 3.7 new deaths per 100000; a 3% decrease, and Germany (1410 new deaths; 0.4 new deaths per 100000; a 3% decrease).

T5 (pre-trained) + *clean*

The European Region reported over 1466000 new cases and over 25000 new deaths, a decrease of 1% and 2% respectively compared to the previous week. The three countries reporting the highest numbers of new cases were the United Kingdom (17232 new cases; 157 new cases per 100000 population; a 1% decrease), the Russian Federation (60468 new cases; 2650 new cases per 100000; a 1% decrease), and the United Kingdom (17232 new cases; 157 new cases per 100000; a 1% decrease). The highest numbers of deaths were reported from the United Kingdom (157 new deaths; 3.7 new deaths per 100000; a 3% decrease), the Russian Federation (2650 new deaths; 3.7 new deaths per 100000; a 3% decrease), the Russian Federation (2650 new deaths; 3.7 new deaths per 100000; a 3% decrease).

T5 (pre-trained) + *shuffle*

In the past week, the European Region reported over 1466000 new cases and over 25000 new deaths, a decrease of 1% and 2% respectively compared to the previous week. The three countries reporting the highest numbers of new cases were Kosovo (2662 new cases; 57 new cases per 100000; a 1% decrease), Turkey (378771 new cases; 57 new cases per 100000; a 1% decrease), France (211674 new cases; 59.6 new cases per 100000; a 21% decrease). The highest numbers of new deaths were reported from the Russian Federation (2650 new deaths; 3.7 new deaths per 100000; a 3% decrease), Germany (1650 new deaths; 3.4 new deaths per 100000; a 3% decrease), and Poland (2537 new deaths; 0.1 new deaths per 100000; a 3% decrease).

T5 (pre-trained) + *subset*

Table 6: Sample output for an epidemiological report for the European region generated by the T5 model and the hierarchical model for a table of data in the test set of CURED4NLG. Text in blue italics shows information filled in from the input table by the baseline template. The text in green italics shows tabular values correctly produced by the end-to-end baseline models while underlined text in red shows the errors in the generated texts. Any hallucinations or repetitions generated are highlighted in purple.

Reporting Country/Territory/ Area	New cases in last 7 days	Cumulative cases	Cumulative cases per 100k population	New deaths in last 7 days	Cumulative deaths	Cumulative deaths per 100k population
Eastern Mediterranean	220035	9648410	1320.2	4709	193761	26.5
Iran (Islamic Republic of)	99205	2739875	3262.0	2109	76633	91.2
Iraq	28359	1136917	2826.6	189	15930	39.6
Pakistan	20511	874751	396.0	670	19467	8.8
Bahrain	11188	199093	11700.5	59	737	43.3
United Arab Emirates	10486	544931	5509.7	19	1629	16.5
Egypt	8248	244520	238.9	424	14269	13.9
Saudi Arabia	6827	432269	1241.7	88	7147	20.5
Kuwait	6725	290801	6809.4	52	1687	39.5
Tunisia	6320	325832	2756.9	429	11779	99.7
Oman	5569	204913	4012.7	101	2184	42.8
Jordan	4112	723345	7089.4	167	9243	90.6
Lebanon	2964	535233	7841.7	125	7585	111.1
...	...	...	...	...	...	...

The Eastern Mediterranean Region reported over 220000 new cases and over 4700 new deaths, a 22% and a 16% decrease respectively compared to the previous week. Case and death incidences have decreased steeply for the past four and two weeks respectively. The highest numbers of new cases were reported from the Islamic Republic of Iran (99205 new cases; 118.1 new cases per 100000; a 20% decrease), Iraq (28359 new cases; 70.5 new cases per 100000; a 26% decrease), and Pakistan (20511 new cases; 9.3 new cases per 100000; a 29% decrease). The highest numbers of new deaths were reported from the Islamic Republic of Iran (2109 new deaths; 2.5 new deaths per 100000; a 13% decrease), Pakistan (670 new deaths; 0.3 new deaths per 100000; a 20% decrease), and Tunisia (429 new deaths; 3.6 new deaths per 100000; a 21% decrease).

Reference

The *Eastern Mediterranean* Region reported over 61700 new cases the past week, a 14% *decrease* compared to the previous week and reported just under 800 new deaths, a *16% decrease*. Of the 22 countries in the region the number of new cases *rose in 6 (27%) this week while they fell in 10 of 22 (1189%)*. The number of new deaths *increased in 1 of 22 countries (14%) and declined in 5 of 22 (23%)* this week. The three countries reporting the highest numbers of new cases in the region this week were *Iraq* (*28359* new cases; 90.2 new cases per 100000, a 15% *decrease*), Lebanon (*2964* new cases; 4.0 new cases per 100000; a 1% decrease) and *Iraq* (*28359* new cases; 40.4 new cases per 100000; a 13% increase). The highest numbers of new deaths were reported from *Islamic Republic of Iran* (19 new deaths; 5.6 new death per 100000 population, an 8% *decrease*), and Iraq (19 new deaths, 4.3 new deaths per 100000, a 12% *decrease*).

Hierarchical Model

In the past week, the *Eastern Mediterranean* Region reported over *220035* new cases, a *decrease* of 9% compared to last week. The region reported over *4709* new deaths, a 2% *decrease*. The highest numbers of new cases were reported from the *Islamic Republic of Iran* (*99205* new cases; 2109 new cases per 100000 population; a 2% *decrease*), *Iraq* (*28359* new cases; 189 new cases per 100000; a 2% *decrease*), *Pakistan* (*20511* new cases; 59 new cases per 100000; a 24% increase). The highest numbers of new deaths were reported from the same countries, the *Islamic Republic of Iran* (*2109* new deaths; 4.7 new deaths per 100000; a 2% *decrease*), *Pakistan* (*670* new deaths; 4.7 new deaths per 100000; a 2% *decrease*) and Egypt (2109 new deaths; 0.1 new deaths per 100000; a 2% *decrease*).

T5 (pre-trained)

In the past week, the *Eastern Mediterranean* Region reported over *220035* new cases, a *decrease* of 9% compared to last week. The region reported over *4709* new deaths, a 2% *decrease*. The highest numbers of new cases were reported from the *Islamic Republic of Iran* (*99205* new cases; 2109 new cases per 100000 population; a 2% *decrease*), *Iraq* (*28359* new cases; 189 new cases per 100000; a 2% *decrease*), *Pakistan* (*20511* new cases; 59 new cases per 100000; a 7% *decrease*). The highest numbers of new deaths were reported from the same countries including the *Islamic Republic of Iran* (*2109* new deaths; 4.7 new deaths per 100000; a 2% *decrease*), *Pakistan* (*670* new deaths; 3.7 new deaths per 100000; a 2% *decrease*) and Egypt (29 new deaths; 0.4 new deaths per 100000; a 2% *decrease*).

T5 (pre-trained) + *clean*

In the past week, the *Eastern Mediterranean* Region reported over *220035* new cases and over *4709* new deaths, a *decrease* of 1% and 2% respectively compared to the previous week. The three countries reporting the highest numbers of new cases this week were *Islamic Republic of Iran* (*99205* new cases; 29.5 new cases per 100000 population; a *21% decrease*), United Arab Emirates (*10486* new cases; 189 new cases per 100000; a 1% *decrease*), and United Arab Emirates (*10486* new cases; 59 new cases per 100000; a 7% *decrease*). The highest numbers of new deaths were reported from the *Islamic Republic of Iran* (*2109* new deaths; 3.7 new deaths per 100000; a 3% *decrease*), *Lebanon* (*429* new deaths; United Arab Emirates new deaths per 100000; a 3% *decrease*) and United Arab Emirates (United Arab Emirates new deaths; United Arab Emirates new deaths per 100000; a 3% *decrease*).

T5 (pre-trained) + *shuffle*

In the past week, the *Eastern Mediterranean* Region reported over *220035* new cases, a *decrease* of 1% compared to last week. The region reported over 47000 new deaths, a 2% *decrease*. The highest numbers of new cases were reported from the *Islamic Republic of Iran* (*99205* new cases; 2109 new cases per 100000 population; a 2% *decrease*), *Iraq* (*28359* new cases; 189 new cases per 100000; a 2% *decrease*), *Pakistan* (*20511* new cases; 59 new cases per 100000; a 7% *decrease*). The highest numbers of new deaths were reported from the *Islamic Republic of Iran* (*2109* new deaths; 0.3 new deaths per 100000; a 3% *decrease*), *Pakistan* (*670* new deaths; 0.3 new deaths per 100000; a 3% *decrease*) and Saudi Arabia (*88* new deaths; 0.4 new deaths per 100000; a 2% *decrease*).

T5 (pre-trained) + *subset*

Table 7: Sample output for an epidemiological report for the region of Eastern Mediterranean generated by the T5 model and the hierarchical model for a table of data in the test set of CURED4NLG. Text in blue italics shows information filled in from the input table by the baseline template. The text in green italics shows tabular values correctly produced by the end-to-end baseline models while underlined text in red shows the errors in the generated texts. Any hallucinations or repetitions generated are highlighted in purple.

WHO Region	New cases in last 7 days (%)	Change in last 7 days	Cumulative cases (%)	New deaths in last 7 days (%)	Change in last 7 days	Cumulative deaths (%)
Americas	1272491 (23%)	-4%	63554005 (40%)	33879 (38%)	-8%	1551860 (47%)
Europe	919119 (17%)	-23%	52871662 (34%)	19056 (21%)	-18%	1104629 (34%)
South-East Asia	2877410 (52%)	6%	25552640 (16%)	28977 (32%)	15%	309197 (9%)
Eastern Mediterranean	280853 (5%)	-13%	9428375 (6%)	5605 (6%)	-13%	189052 (6%)
Africa	40656 (1%)	-5%	3357846 (2%)	1034 (1%)	3%	83904 (3%)
Western Pacific	127073 (2%)	-4%	2597134 (2%)	1691 (2%)	34%	39179 (1%)
Global	5517602 (100%)	4%	157362408 (100%)	90242 (100%)	-4%	3277834 (100%)

Reference	Template
The number of new COVID-19 cases and deaths globally decreased slightly this week, with over *5.5 million* cases and over *90000* deaths (Figure 1). Case and death incidence, however, remains at the highest level since the beginning of the pandemic. New weekly cases decreased in the regions of *Europe* and *Eastern Mediterranean*, while the *South-East Asia Region* continued an upward trajectory for 9 weeks and reported a further *6% increase* last week (Table 1). Death incidence increased in the *South-East Asia* and *Western Pacific* regions. While India continues to account for *95%* of cases and *93%* of deaths in the *South-East Asia* Region, as well as *50%* of global cases and *30%* of global deaths, worrying trends have been observed in neighbouring countries. In all WHO Regions there are countries which have been showing a sustained upward trend in cases and deaths over several weeks. The highest numbers of new cases were reported from India (*2738957* new cases; *5%* increase), *Brazil* (*423438* new cases; similar to previous week), the *United States of America* (*334784* new cases; *3% decrease*), Turkey (*166733* new cases; *35% decrease*), and Argentina (*140771* new cases; *8% decrease*).	Globally, over 5.5 million new cases and 90000 new deaths have been reported to WHO in the past week. A cumulative total of 157.3 million cases and 3.2 million deaths have been reported since the start of the outbreak. The number of new cases increased by 4% and the number of new deaths decreased by 4% globally in the last 7 days. The WHO Region of South-East Asia was the most affected region with 2.8 million new cases and 28000 new deaths. This region noted an increase of 6% in new cases since the last week and accounts for 52% of all new cases. Regions reporting an increase in new cases include South-East Asia. Regions reporting a decline in new cases include Africa, Americas, Eastern Mediterranean, Europe and Western Pacific. Regions reporting an increase in new deaths include Africa, South-East Asia and Western Pacific. Regions reporting a decline in new deaths include Americas, Eastern Mediterranean and Europe. The highest numbers of new cases were reported from India (2738957 new cases; 5% increase), Brazil (423438 new cases; similar to previous week), United States of America (334784 new cases; 3% decrease), Turkey (166733 new cases; 35% decrease) and Argentina (140771 new cases; 8% decrease).

T5 (pre-trained)	Hierarchical Model
In the past week, the number of new COVID-19 cases and deaths has continued to increase, with over <u>1.2 million</u> new cases reported globally, a *4% increase* compared to the previous week (Figure 1). The <u>Region of the Americas</u> continues to account for the greatest proportion of new cases and new deaths globally. The Region of the Americas continues to account for the greatest proportion of new cases and deaths globally. The Region of the Americas continues to account for the greatest proportion of new cases and deaths globally. <u>The Region of the Americas continues to report a decline in new cases and deaths.</u> The *Eastern Mediterranean* Region reported a *decline* in new cases and deaths, with over 1.2 million new cases reported last week. The Region of the Americas continues to report a decline in new cases and deaths, while the Eastern Mediterranean Region reported a decline in new cases and deaths, with over 1.2 million new cases reported last week. The Region of the Americas continues to report a decline in new cases and deaths, with over 1.2 million new cases reported last week, a 4% increase compared to the previous week. The Region of the Americas continues to report a decline in new cases and new deaths, with over 1.2 million new cases reported last week. The Region of the Americas [...]	The number of global new cases reported continues to fall for the sixth consecutive week, with <u>2.4 million</u> new cases and <u>36000</u> new deaths reported globally, while the number of new deaths has remained relatively stable. As of <u>18 October</u>, over <u>40 million</u> cases and <u>1.1 million</u> deaths have been reported globally. The further <u>acceleration</u> in the incidence of new cases was most notable in <u>European</u> Region, which reported <u>half</u> of global new cases (over <u>1.7 million</u> cases - a *22%* <u>increase</u> from the previous week. Moreover, the region also reported a substantial <u>rise</u> in the number of new deaths (a <u>46% increase</u> compared with the previous week), with Global new deaths in the past week. The WHO *South East Asia* Region showed the *highest rise* in new cases in the past week, with over <u>500,000</u> new cases reported. In the European Region, new cases and new deaths have continued to <u>increase</u> over the past seven days compared to the previous week. Along with the Region of the Americas, the percentage change in new cases in Global the week. The *Eastern Mediterranean* Region reported a *decline* in new cases and deaths, <u>6%</u> and <u>8%</u> respectively, compared to the previous week. The decline is mainly due to decreases in reported cases in <u>India</u> and <u>Bangladesh</u>. For the second week in a row, the Regions of the <u>Eastern Mediterranean</u> and the <u>Western Pacific</u> reported <u>increases</u> in cases and deaths. *Overall, during the reporting period, all the Regions showed an* <u>*increase*</u> *in cases except the South-East Asia Region.* Countries reporting the highest number of cases in the past seven days include; India, the United States of America, Brazil, the United Kingdom and France.

Table 8: Sample output for a global epidemiological report generated by the T5 model and the hierarchical model for a table of data in the test set of CURED4NLG. Text in blue italics shows information filled in from the input table by the baseline template. The text in green italics shows tabular values correctly produced by the end-to-end baseline models while underlined text in red shows the errors in the generated texts. Any hallucinations or repetitions generated are highlighted in purple.