StoryBot

Interactive Story Generation For Kids

Ram Senthamarai, Ghiwa Lamah, Nicolas Loffreda

Agenda

- Introduction and Motivation
- Data
- Models
- Experiments
- Manual Evaluation
- Conclusions & Next steps

Introduction and Motivation

Our Goal

Develop an NLP **sentence continuation model** that will enable an **eventual interactive story development experience** for children,
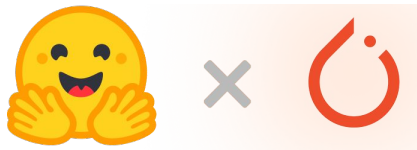
Data

- **Children Stories Text Corpus:** Dataset of over 200 children books from Kaggle
- **Post-Processed** using SpaCy Sentencizer Package
- **Three Fine-Tuning Datasets:**

Dataset	Variable	Label	Max Tokens
S1	**One** sentence as context	The following sentence	65
S2	**Two** sentences as context		110
S3	**Three** sentences as context		165

Model selection

Model	Checkpoint	Type	Parameters
OPT (baseline)	facebook/opt-350m	Decoder	~350M
OPT	facebook/opt-350m	Decoder	~350M
T5	google/t5-v1_1-base	Seq2Seq	~280M
BERT-2-BERT	bert-cased-base	Seq2Seq	~216M

Experiments

Parameters
- Epochs = 3
- Max tokens = 65, 110, 150 for S1, S2, S3

Outcomes
- **S3 fine-tuned versions chosen**
- Baseline: 25% blank output (test)
- T5, B2B: Short outputs

- **Need for manual evaluation**

Model	BLEU	ROUGE1	ROUGE2	ROUGEL	ROUGELSUM	BLEURT
Untuned OPT	0.005272	0.085343	0.007274	0.065647	0.065646	-0.997209
OPT S2	0.008313	0.137038	0.013192	0.101051	0.101071	-0.949812
OPT S3	**0.009127**	**0.140313**	**0.015348**	**0.015348**	**0.103603**	**-0.924052**
T5 S1	0.007994	0.125251	0.010452	0.101473	0.101447	-1.108518
T5 S2	0.011892	0.132282	0.017951	0.110494	0.110561	-1.107098
T5 S3	**0.013134**	**0.130802**	**0.020352**	**0.109915**	**0.109953**	**-1.08875**
Bert2Bert S1	0.010602	0.119237	0.014406	0.101568	0.101565	-1.109143
Bert2Bert S2	0.011956	0.126579	0.017824	0.108175	0.108119	-1.098777
Bert2Bert S3	**0.012474**	**0.130512**	**0.019387**	**0.111407**	**0.111362**	**-1.079424**

Manual Evaluation

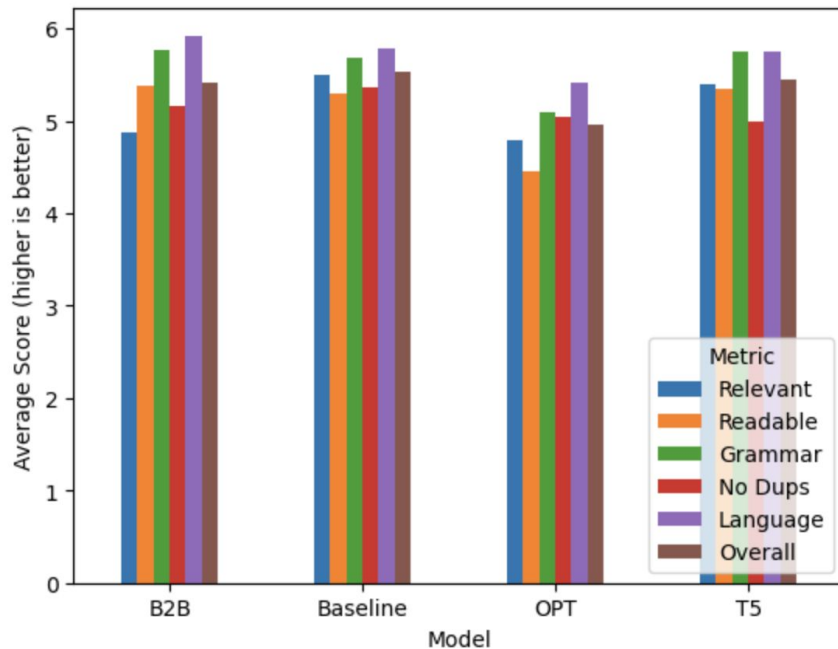## Generated Text from Thirty Prompts

Evaluated on 5 metrics

Relevancy	Readability	Grammar	Non-Redundancy	Kid-Friendly Language

Using 7-point Likert scale

1	2	3	4	5	6	7
Unusable	Very Bad	Bad	Neutral	Good	Very Good	Outstanding

Initial Results

#1
Close and above average scores

#2
Bert-2-Bert beat baseline in ⅘ metrics

#4
Tuned version of OPT performed worst

Model	Relevant	Readable	Grammar	No Dups	Language	Overall
B2B	4.8667	5.3778	5.7667	5.1667	5.9222	5.4200
Baseline	5.5000	5.2889	5.6889	5.3556	5.7778	5.5222
OPT	4.7889	4.4556	5.1000	5.0444	5.4111	4.9600
T5	5.4000	5.3444	5.7444	4.9889	5.7556	5.4467

#3
Baseline performed best overall

Analysis
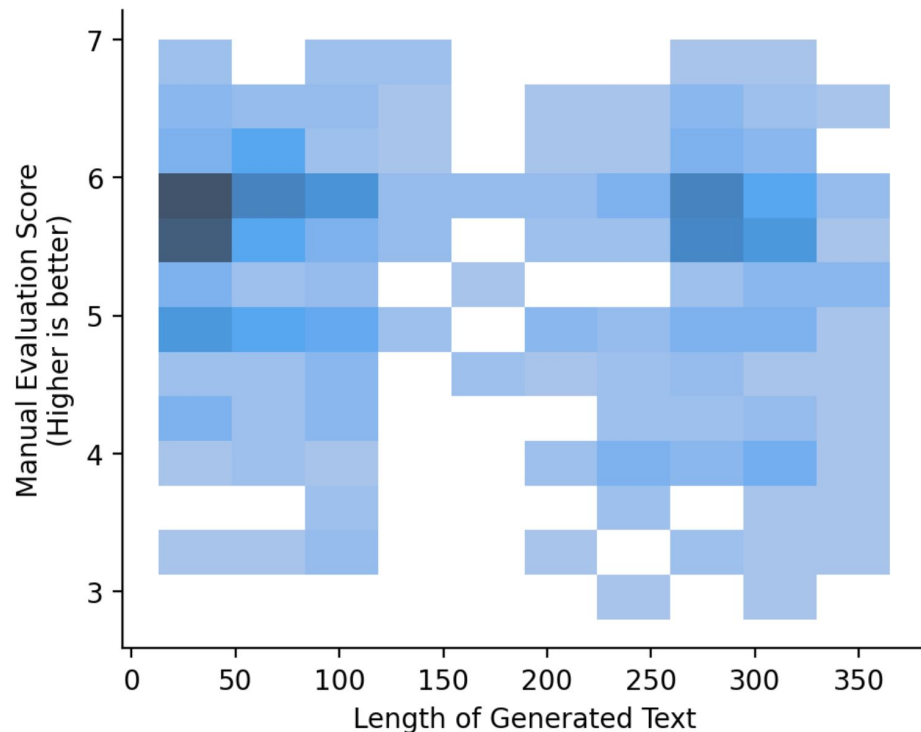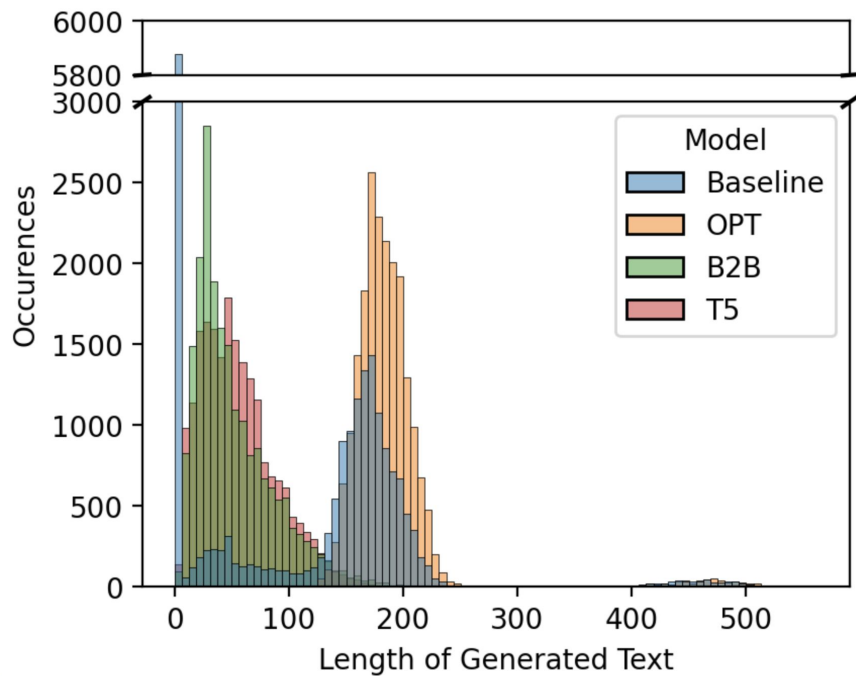
Inter-Annotator Agreement

Krippendorff's Alpha

Perfect disagreement ⟶ **−1 < α < 1** ⟵ **Perfect agreement**

Overall Score = 0.1766

⇓

Weak Correlation
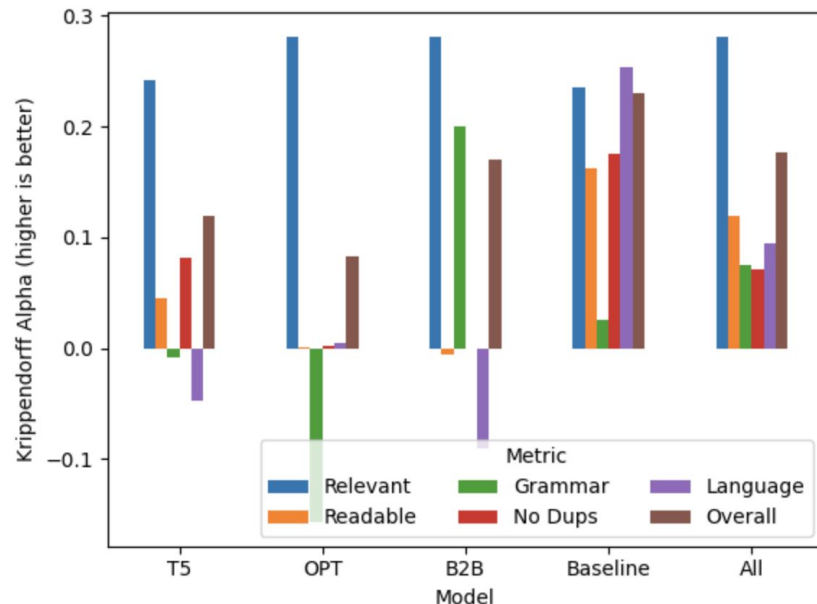
Model	Relevant	Readable	Grammar	No Dups	Language	Overall
T5	0.2418	0.0453	-0.0086	0.0821	-0.0479	0.1193
OPT	0.2811	0.0011	-0.1568	0.0016	0.0045	0.0827
B2B	0.2811	-0.0053	0.1999	-0.0011	-0.0903	0.1705
Baseline	0.2349	0.1619	0.0253	0.1751	0.2538	0.2303
All	0.2808	0.1195	0.0755	0.0711	0.0947	0.1766

Conclusions

1

Dataset selection and processing has a substantial impact on the success of fine-tuning.

2

Pre-trained models are susceptible to **degraded performance** during initial stages of fine-tuning.

3

Manual evaluation strategy with **sound criteria and clear cut guidelines** is key for evaluating text generation models.
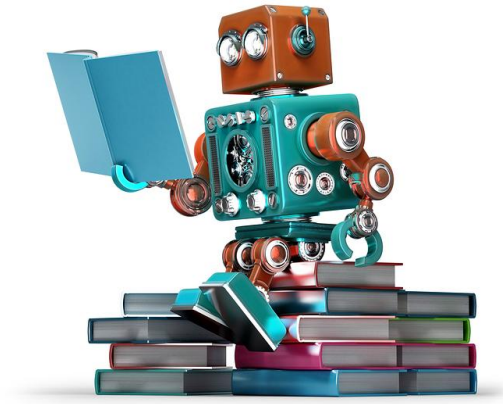
Next Steps

1

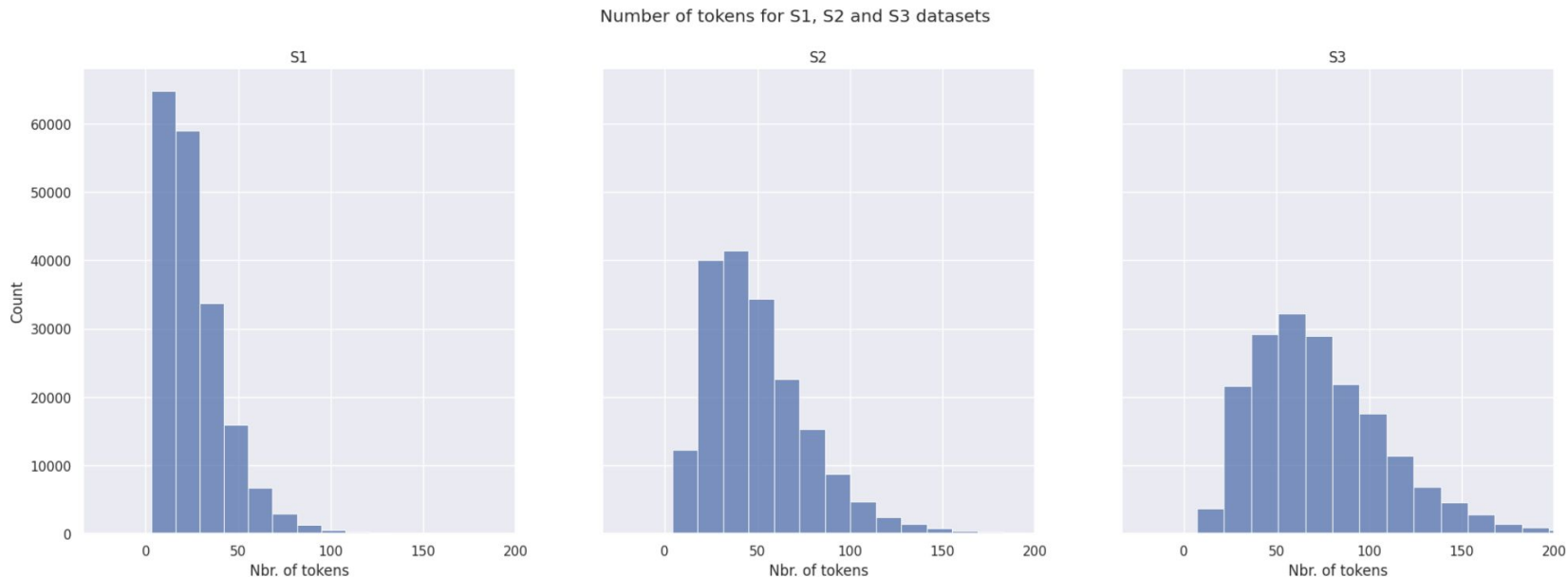Deep-dive corpus post-processing strategies
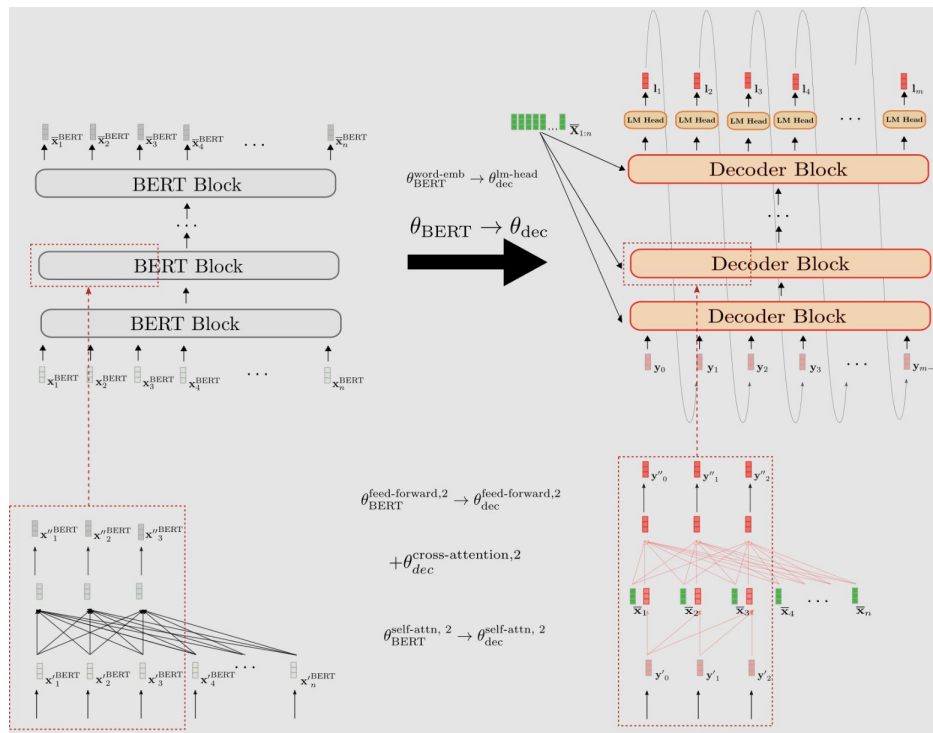
2

Further explore human evaluation strategies

Questions?

APPENDIX

Distribution of Number of Tokens

Number of tokens for S1, S2 and S3 datasets

B2B architecture

StoryBot In Action

Longest Outputs

Shorter Outputs

Image References

- https://robohub.org/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings/
- https://toistudent.timesofindia.indiatimes.com/news/bookmark/storytelling-robots-for-better-endings/44641.html
- https://www.presentation-guru.com/use-storytelling-to-present-with-power/
- https://www.pinterest.com/pin/131308145362081669/
- https://www.nicepng.com/maxp/u2q8r5q8r5i1q8w7/
- http://www.stickpng.com/img/cartoons/peter-pan/peter-pan-flying
- https://www.cleanpng.com/png-classic-sherlock-holmes-london-the-hound-of-the-ba-191602/download-png.html
- https://www.vhv.rs/viewpic/hbmRwTT_pinocchio-nose-png-pinocchio-with-long-nose-transparent/
- https://medium.com/pytorch/bettertransformer-out-of-the-box-performance-for-huggingface-transformers-3fbe27d50ab2
-

Berkeley
UNIVERSITY OF CALIFORNIA