

Sibil AI: Children Story Generator in Sinhala using Transformers

RMVD Bandara¹, HAA Sanja¹ and B Hettige¹

¹General Sir John Kotelawala Defence University, Sri Lanka

#budditha@kdu.ac.lk

Abstract: Ever since the birth of humankind, stories have been used as a means of sharing information and educating people. Stories are more than just a form of entertainment; they impart lessons that often help children to develop the skills they need to thrive in life. Research and development teams have unquestionably mastered the practicalities of producing human-like creative text tales, which has been a significant barrier in natural language processing in recent years. A system based on artificial intelligence that generates children's stories can serve as a resource for parents and children to connect with. The ability to generate natural language stories that people can understand, remember, and enjoy is difficult to achieve with current technology. A new model based on transformers is introduced in this paper. This new approach for generating stories for children based on the GPT-2 model with the help of a web application. The GPT-2 is a model based on neural network that is designed to imitate the human behaviour of producing creative and coherent text. It can generate stories in different genres and starting captions. The web application takes advantage of the GPT-2 model's ability to generate fluent texts, including proper punctuations, complex syntaxes, and grammar rules. The solution allows users to generate creative stories from different genres with starting captions. Especially, using the proof-of-concept to support the narration given in Sinhala language, one of the native languages in Sri Lanka.

Keywords: Artificial Intelligence, GPT-2, Story Generator

1. Introduction

Communication done through a behaviour called storytelling is a way of communicating through many different forms of art, like poetry or music. The study found that storytelling using the native language of the children brought an emotional upbeat utilizing analogies to communicate stories using metaphors. It has been proven and seen through every single person's life that the roots of a person how well they turn out to be in their adult lives go a long way back to their childhood. These root causes can be interactions of a range of biological, psychological, and social variables which can act as the accelerating and maintenance factors. Storytelling can be an effective means to provoke positive emotions in people. This is especially obvious in children who are above the age of five. By seeing outcomes, they will also see how they make them happen. This also helps children to understand that they are capable of achieving something when they set a goal

and work hard enough towards it. The dataset we use for training and testing have a significant impact on how well the generation of the narrative model performs. DL (Deep Learning) architects and designers consequently, using these self-learning mechanisms, transformer models and DL techniques employed several models. This paper aims to generate a transformers-based model that can produce stories with an amount of descriptive data that they have not been able to achieve in any previous literature.

This paper is formed with six sections which the research contributed in-depth.

Section I: Introduction sets the tone for the research through a brief overview of the problem, current solution, and the proposed solution with its major promising novel features.

Section II: Literature review is a collection of studies on the research findings that have been carried over creation of stories using a machine. Because the focus of this study is on GPT-2 transformer model and NLP problems in narrative production, the methods as well as the strategies that contributes to overall coherency and comprehension in the producing context, were critically examined.

Section III: Methodology sets forth the configuration of the technology utilized in the study with the innovative techniques used to build the narrative generator.

Section IV: Design and Implementation illustrate the proposed system's architecture and the action taken to implement the intended unique methods to developing the web application.

Section V: Results describes the proposed solution outputs after being implemented. The plan of evaluation and the actions were detailed in here.

Section VI: Conclusion concludes the study with the comments on the study endeavour. It specifies how the system's objectives were met, as well as its limitations. Furthermore, it makes recommendations for further research in this area.

2. Literature Review

Case-based reasoning model, data-driven approach model, suspenseful story generator, location-based text generators, and the go transformer models were built upon in the field of text generation in the early days.

Natural Language Processing evolved from intelligent question answering, dialogues, machine translation, and other systems into template-driven Natural Language Generation (NLG), that fits into pre-existing templates. Then it progressed to advance NLG, which divides facts from data to comprehend the most significant and fascinating one, which is referred to as deep learning. Simultaneously, the related technology has achieved several advancements including Recurrent Neural Network (RNN), Long

Short-Term Memory (LSTM), and the Sequence to Sequence(S2S) model through the emergence of the GPT-2 and Transformer-based BERT language model.

As this case study continues, (Hoppe and Toussaint, 2020) has shown that language models learn these tasks of question answering, machine translation, reading comprehension, and summarization; without explicit supervision when trained on WebText, a new dataset made up of millions of web pages. Typically, these tasks are approached with supervised learning on task-specific learning on the corresponding datasets. The GPT-2 model used here with a 1.5B parameter transformer delivers cutting-edge performance on 7 out of 8 evaluated language modelling datasets, although it still unfits WebText. These findings point to a viable direction for developing language processing algorithms that learn from examples that occur in real life.

Automated storytelling aims to construct a cohesive tale using intermediary documents were briefly researched (Rigsby and Barbara, 2017). By offering quantitative techniques of narrative quality evaluation that have been proved to have good agreement with human judgment, this study advanced the state-of-art. There were two developed automatic narrative evaluation techniques as dispersion, a metric for narrative flow, and coherence, a metric for the effectiveness of the middle-of-the-story articles conveys information about the connectivity between the start and the end. Chapters from Anne of Green Gables, Black Beauty, Peter Pan, and Treasure Island were divided into a single dataset and later combined with the Atlantic Storm dataset.

The encoder and decoder-equipped conventional networks or complicated Recurrent Neural Networks (RNN) serve as the foundation for the Sequence transformer models. To create poetry in Chinese, the sequence2sequence model combine the attention mechanism is applied in (Lin, Gao, and Chang, 2019). It is challenging to utilize in this context since the Chinese tale generation follows certain norms. In this case, the “Demi-Gods and Semi-Devils” dataset and a sizable dataset of Chinese short text summarization will be combined by FastText as input data.

As a novel solution, two Thai sentence creation machine was studied (Krukaset, Krukaset and Khancome, 2018). These computers use fixed patterns from the sentences that were manually evaluated. All fixed patterns are implemented within the machines using directed graphs. Tables are used to construct and display every path in a graph.

The Folktale Generator System (FGS) which served as the story grammar, is based on Vladimir Propp’s theory (Zhang, Tran and Fangbemi, 2016) of narrative function. FGS enables users to intuitively create a variety of story plots by naturally sketching curves. Each useful Propp’s theory segment is altered into ten versions, one of which is the original given a score attribute from one to 10 and then stored as text scripts in a database. By utilizing the numerous story possibilities, it provides, FGS may give old folktales a fresh perspective and strengthen the preservation of their cultural values. By utilizing the numerous story possibilities, it provided, FGS may give old folktales a fresh perspective and strengthen the preservation of their cultural value.

A revolutionary method for a control-and-edit transformer methodology that supports deleting policy and adding policy by using controlled imitation learning of editing distance from dynamic programming was studied by (Chen *et al.*, 2021). In the study, a weighting-reward with corpus stat possession which measures continuous rewarding for the controlled goal was described in the context.

An important area of NLP is the creation of sentences from provide starting words or the completion of incomplete phrases. It shows, in one way, whether a computer is capable of human thought and creativity. The study done by (Qu *et al.*, 2020) discuss programming the machine to perform a specific task before using it in NLP to aid with application in situations like summary creation, machine translation, and automatic question answering. For text generation and prediction, the OpenAI GPT-2 and BERT models were utilized extensively.

An interactive personalized story generation was discussed in (Yu and Riedl, 2012). A system demonstrated on a simple interactive story generation based on choose-your-own-adventure stories to evaluate the algorithms. In this study, user handle the narration they have seen up to and then, continue the narration by providing their ratings.

With the help of Smart Game Format (SGF), which provides a written description of move sequences, (Ciolino, Kalin and Noever, 2020) have trained the Generative Pretrained Transformer (GPT-2) to play Go game like a pro. Results show that language modelling is capable of capturing both the strategic formations and the sequencing pattern of championship Go games, and that random play takes undesirable beginning locations, like as the board edges and second rows, hence there are some obvious differences between random and Go transformer play that may be attributed to training. The resulting model generates a nearly complete and coherent game sequence without any human knowledge, heuristic rules, or strategic guidance.

Automated tale creation employing a computational model of focalization and several plan libraries for the story’s characters were studied in (Author, 2011). By telling the same tale from numerous angles, multiple internal focalizations can provide complex events with rich and varied explanations. However, because of the repetition, it can also weary the audience or reader.

3. Methodology

A. The Population, Qualitative and Quantitative Research
The population of users for the proposed state-of-art text generation model solution included:

- All 5 to 6 years old.
- Preschool children.
- Childcare authorities.
- Children’s storybook authors.
- Children’s storybook publishers.
- Children’s home wardens.
- Especially abled children.
- Their guardians.

The sample group was selected in multi-stage randomly. To gain a better understanding of the case study, a survey questionnaire was carried out. Out of the 171 responses collected, 91 responses marked a parent of at least one kid.

Number of Kids	Percentage of parents who responded (%)	Age of the Youngest Child	Percentage of the families who responded (%)
1	24.2	Less than 1 year	13.2
2	38.5	1-3 years	27.5
3	25.3	3-6 years	11
4	2.2	6+ years	48.4
5	9.9		

Table 1: Survey Statistics

As the result shows, the majority of families had two children. The findings for the questionnaire question on who remains with the youngest child in the household were then as follows:

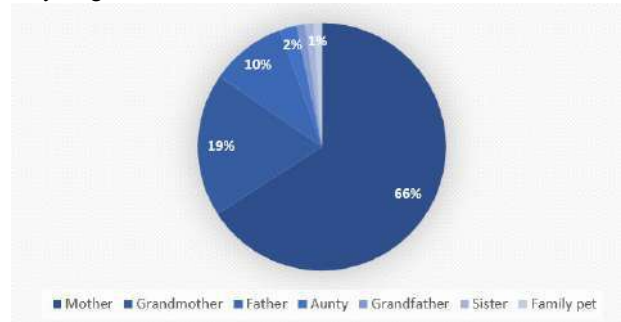


Figure I: Survey Statistics on "Who Stays with The Youngest"

This results in the mother of the child spending more time with the child, however working moms only have a noticeably short window of time to spend with their children. So, that causes a major issue when it comes to kids getting into tech autism, being violent, or less creative of perceptions and imaginations. interactions of the kid develop with the cause of time when the children spend more time with their family.

To build a case on the facts, the participants have answered regarding the behaviour of the youngest as follows,

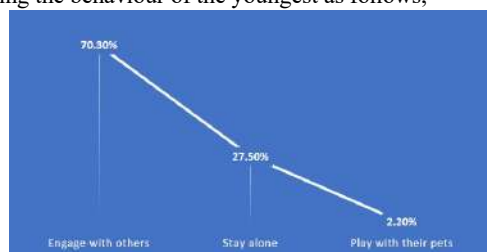


Figure II: Survey Statistics on "Children Behaviour"

As a result, the study found that children are more engaged with one another and had no problem reading books alone or listening to someone read aloud to them. But the kids who stay alone require some encouragement towards reading books.

Then the questionnaire also included responses on digital device usage of the kids in their families, and the results are shown as follows,

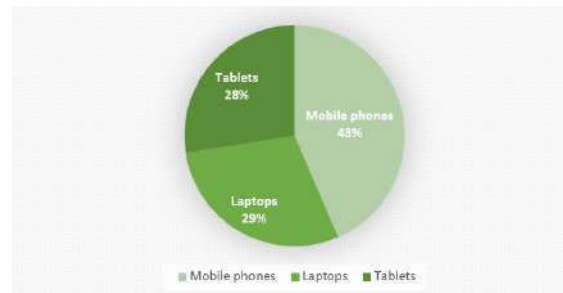


Figure III: Survey Statistics on "Digital Device Usage of Children"

This concluded with the fact that mobile phone usage was seen as an addiction on the benefit of the doubt.

The researcher also came across a campaign held by UNICEF at LAOS. This campaign was carried out by UNICEF because most schools have been closed temporarily due to the coronavirus pandemic which has hit hard on kids who are in the age group of 5 to 13 years of age.

It was able to niche down the captured observations where there was one specific kid who was affected with tech autism. The research was able to gain observation on him as:

- Child only reacted to a set of specific keywords.
- Child has fewer interactions with strangers, but better interactions with parents and siblings.
- Children have a good sense of knowledge of tech devices, and they are fast learners and identify patterns.
- Child understands English and Sinhala language and is interested in reading stories about princesses, and books with large pages and illustrations.
- Child shows aggressiveness when there is a situation that carries out a big noise but is affectionate to animals.

Following identifications were abstracts from the UNICEF at LAOS read-out loud campaign.

- Children happen to read more stories and actively engage in them when the stories are in their native language, and they have improvement of the use of their language over the cause of time reading.
- Parents happen to encourage children to read more English stories and keep them engaged with educational activities without considering the pandemic.
- The age group of 5 to 8-year-olds' parents tend to read stories to children than children read them by themselves

The researcher was able to draw meaningful conclusions from the observations stated above, having a substantial influence on the study endeavour. The event planning of the event has been aided by the children's and the kid with autism's recognized behaviour. The modelling process needs to create a universal application that can be utilized by any youngster, whether or not they have tech autism, who belongs to the target audience of users.

As per the conclusion of the observations and the questionnaire, it was able to form a model to cater to the target audience's requirements. A high-level abstraction, a genre/theme-wise story generating an application with the ability to generate stories based on starting statements. And to support the Sinhala language to satisfy the native language readers and listeners in Sri Lanka.

4. Design and Implementation

With the completion of the entire project on python3, react was used as the frontend framework to produce the graphical user interface for generic users. GPT-2 model was trained on amazon web services to deploy the trained model on google drive in order to fetch it with google colab to present the proof-of-concept for state-of-the-art text generation.

A. Overall System Architecture

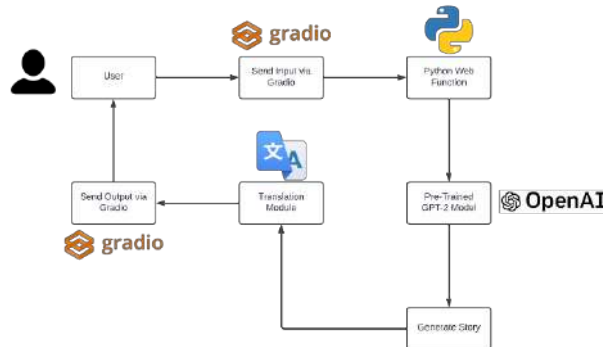


Figure IV: Overall System Architecture

The proof-of-concept application was built on google colab environment. When utilizing the Gradio online interface, the user is given the option to choose a genre, topic, or an opening statement. The chosen input feed is then sent to the trained model, which was trained using Amazon Web Services, to produce a series of tales that are then returned to the output environment. The created tale is placed into a collection of English language stories, and using the Google Translator API, it is translated into Sinhalese. Both tales may be produced and sent to the user for recitation.

B. Implementation Plan

To achieve the aims and objectives of the development by sprints, implementation was carried out in an agile way, and the research component took the bulk of the time for project preparation.

The research process is broken down into seven steps.

Step 1: Form up a suitable dataset

Step 2: Finetune the GPT-2 model into children's stories

Step 3: Train the fine-tuned model in the AWS environment

Step 4: Generate children's stories based on the dataset formed up

Step 5: Translate the generated stories into Sinhala using Google translator API

Step 6: Enabling text-to-speech synthesis for special-abled children

Step 7: Evaluate the generated text output

The formation of the dataset was the most critical task of the study to validate the proof-of-concept. When gathering the dataset, it is important to pay attention to the details of the context and the vocabulary that has been used in the datasets. Length, credibility, and the language vocabulary of a story are the key points to focus on.

After that, it was able to go through some other datasets that are available outside which have been already used in similar studies. Facebook researcher dataset, Shakespeare stories dataset, Rick and

Morty stories dataset, hugging face story merge dataset, and fellow researchers' dataset were thoroughly studied.

Later, a few datasets were dropped due to the reasonings of:

Dataset	Removed /Selected	Reason
Facebook researcher	Removed	Volent and racist words
Shakespeare stories	Removed	Contains high profile words which are not easily understandable by a child
Rick and Morty Stories	Removed	Contains mismatch data in between
Hugging face story merge	Selected	Continue to proceed with
Fellow researchers' dataset	Selected	Continue to proceed with

Table 2: Datasets

```

1 The first time I saw her she was sitting on the ground, her hair was long and dark, and she was wearing a simple, dark dress.
2 She looked up at me and smiled, her eyes were bright and full of life.
3 I had never seen her before, but she seemed so familiar, as if I had known her for a long time.
4 She was sitting on the ground, her back to me, and she was looking at something in her hands.
5 I walked over to her and saw that she was holding a small, round object in her hands.
6 It was a small, round object, and it was made of a dark, smooth material.
7 I had never seen anything like it before, and I was curious to know what it was.
8 She looked up at me and smiled, her eyes were bright and full of life.
9 I had never seen her before, but she seemed so familiar, as if I had known her for a long time.
10 She was sitting on the ground, her back to me, and she was looking at something in her hands.
11 I walked over to her and saw that she was holding a small, round object in her hands.
12 It was a small, round object, and it was made of a dark, smooth material.
13 I had never seen anything like it before, and I was curious to know what it was.
14 She looked up at me and smiled, her eyes were bright and full of life.
15 I had never seen her before, but she seemed so familiar, as if I had known her for a long time.
16 She was sitting on the ground, her back to me, and she was looking at something in her hands.
17 I walked over to her and saw that she was holding a small, round object in her hands.
18 It was a small, round object, and it was made of a dark, smooth material.
19 I had never seen anything like it before, and I was curious to know what it was.
20 She looked up at me and smiled, her eyes were bright and full of life.
  
```

Figure V: Facebook Dataset

```

GoldilocksAndTheThreeBears animal social 1837 RobertSouthey Once upon a time, there was a little girl named Goldilocks.
1 She was sitting on the ground, her back to me, and she was looking at something in her hands.
2 I walked over to her and saw that she was holding a small, round object in her hands.
3 It was a small, round object, and it was made of a dark, smooth material.
4 I had never seen anything like it before, and I was curious to know what it was.
5 She looked up at me and smiled, her eyes were bright and full of life.
6 I had never seen her before, but she seemed so familiar, as if I had known her for a long time.
7 She was sitting on the ground, her back to me, and she was looking at something in her hands.
8 I walked over to her and saw that she was holding a small, round object in her hands.
9 It was a small, round object, and it was made of a dark, smooth material.
10 I had never seen anything like it before, and I was curious to know what it was.
11 She looked up at me and smiled, her eyes were bright and full of life.
12 I had never seen her before, but she seemed so familiar, as if I had known her for a long time.
13 She was sitting on the ground, her back to me, and she was looking at something in her hands.
14 I walked over to her and saw that she was holding a small, round object in her hands.
15 It was a small, round object, and it was made of a dark, smooth material.
16 I had never seen anything like it before, and I was curious to know what it was.
17 She looked up at me and smiled, her eyes were bright and full of life.
18 I had never seen her before, but she seemed so familiar, as if I had known her for a long time.
19 She was sitting on the ground, her back to me, and she was looking at something in her hands.
20 I walked over to her and saw that she was holding a small, round object in her hands.
  
```

Figure VI: Fellow Researcher Dataset

First Citizen:
Before we proceed any further, hear me speak.

All:
Speak, speak.

First Citizen:
You are all resolved rather to die than to famish?

All:
Resolved. resolved.

First Citizen:
First, you know Caius Marcius is chief enemy to the people.

All:
We know't, we know't.

First Citizen:
Let us kill him, and we'll have corn at our own price.
Is't a verdict?

All:
No more talking on't; let it be done: away, away!

Second Citizen:
One word, good citizens.

First Citizen:
We are accounted poor citizens, the patricians good.
What authority surfeits on would relieve us: if they
would yield us but the superfluity, while it were
wholesome, we might guess they relieved us humanely;
but they think we are too dear: the leanness that
afflicts us, the object of our misery, is as an
inventory to particularise their abundance; our
sufferance is a gain to them: Let us revenge this with
our pikes, ere we become rakes: for the gods know I
speak this in hunger for bread, not in thirst for revenge.

Second Citizen:
Would you proceed especially against Caius Marcius?

All:
Against him first: he's a very dog to the commonalty.

Second Citizen:
Consider you what services he has done for his country?

Figure VII: Shakespeare Dataset

And the following observations have been carried out regarding the selected datasets.

- Stories have no same length
- Storied happen to be untrue sometimes
- Stories happen to have different characters on the ongoing
- Stories happen to have genres that are quite out of the league for the kids
- Generating a native translation of the stories could be challenging

C. Setting the Development Environment & Finetuning the Model

To carry out the development procedure, Google Colab notebook environment was selected as it requires no prior installation or configurations and provides GPU, CPU, and TPU runtime environment free of charge to proceed with the coding. And also, it supports Python, and its libraries and faster runtime environment access are needed to host the Jupyter notebook to carry out heavy coding without crashing the browser.

The development process included a few more libraries along with Python which are: transformers, GPT-2 simple, Gradio, and Google trans. And the gradio was used for the development of the Python web interface.

GPT-2 model was trained by feeding the selected datasets to get better results from the state-of-art text generation. The final model

was 1.32GB in size when the finetune was the end. Amazon web services provided by The AI (Artificial Intelligence) Team (Pvt) Ltd were given the ability to train the model without causing any challenges on the computing power.

Finally, the python web interface was developed using Gradio library.

5. Results

The transformer module was evaluated to see if it was successful in optimizing for the generated table. To test the transformer-based GPT-2 module, multiple values were assigned to the parameters, which were then fine-tuned numerous times to produce cohesive tale material.

The output material had several pointless phrases of information plagiarized from the internet as well as numerous variations on the story that was being taught. The effectiveness of the deep learning module in preventing overfitting of the input training data was also examined.

After some tweaking and adjusting the dataset filed to see how the GPT-2 module behaved, the goal of a cogent creative narrative construction without underfitting or overfitting was attained. This was caused by the GPT-2 transformer model's auto-regressive function. As a result, it was able to extract two tidy instances of stories from the dataset that it had learned.

A. Translating Into Sinhala

The process of translating the story from English to the Native Sinhala language was tested and evaluated.

	English Text	Sinhala Text
Case 01	Once upon a time there was a little prince who loved adventurous travel. But one day his father said, "Son, you must not go anywhere without your father's permission." He sadly fled in search of something. But after a while, he started to get into trouble. The king sent his army and ordered them to return home. So, he went back home and promised never to disobey him.	ඉ සජනක ගමන් බිමන් වලට ප්‍රිය කරන පුංචි ගමාරයෙක් එක කායලක හිටියා. නමුත් දිනක් ඔහුගේ පරියා පැවහගේ "පතා, ඔබ ඔගේ පරියාගේ අවසරයකින් යතාරව යකාය ටේවන් යන්න එපා" යනුවෙන්. ඔහු දුටුයකින් යමක් යසායමරින් පලා ගියේ ය. ඒත් ටික දවසක් යන යකාට එයා අමාරුයටේ වැයටන්න ගතේකා. රජු තම මුද්‍රාව යවා ඔවුන්ට ආපහ යගදර යන යලස අණ කය ය. ඒ නිසා, ඔහු ආපහ යගදර
Case 02	Cats and dogs. "The dogs were catching and playing," he said. Then she caught a scorpion in her claws, and she ran away. "" What did she do to you? "I asked." She bit you with her teeth. "" Did she bite you with her	බලුන් ස බල්ලන්. "බල්ලෝ අල්ල යගන යසල්ලම් කා," ඔහු පැවසය. එවිට ඇය යගෝණුසේයසග ඇගේ නියයපාතු වලට හ වූ අතර ඇය පලා ගියා ය. "" ඇය ඔබට කය ය. "" ගමක් ද? "මම ඇහවා." ඇය ඔබව ඇය ඇගේ නියයපාතු

claws? " "Yes," I said. "She bit your clothes. She bit your teeth." He asked. "Yes," I said. "She bit them on your clothes. She bit your teeth	වලින් ඔබව සපා කැවාද? "ඔව්, " මම බයවමි. "ඇය ඔයේ ඇඳුම් පා කැවා ය. ඇය ඔයේ දත් පා කැවා ය." ඔහු ඇහයේ ය. "ඔව්, " මම බයවමි. "ඇය ඒවා ඔයේ ඇඳුම් මත සපා කැවා ය. ඇය ඔයේ දත් සපා කැවා ය.
--	--

Table 3: Translation Testing

According to the generated outcome, it shows that the translation has been successfully done in both cases yet few grammar issues along the context.

The integration testing was carried out after the transformer0based GPT-2 model was integrated into the Python web interface application. The values were entered into the system and the correlation between the user inputs (theme, prompt keywords) was assessed quantitatively. The pipeline ran successful measures showing the integration testing has passed for the implemented system.

6. Conclusion

The novel approach discussed in the research study depends on the dataset that is being used to implement the system.

To continue the further development, it is required a system with higher RAM capacity to finetune the GPT-2 model in advance. Children love interactive stories consisting of several imageries and lesser texts. Such as Disney movies, ice age, how to train your dragon, Moana, Garfield. Including imageries to the implemented system has been restricted due to GPT-3 model not being available for public usage.

Story generating system discussed in this research study evaluates the results through a web application using AI (Artificial Intelligence) based on RNN. For future development finding relevant Sinhala datasets and implementing the Sinhala web application, test and evaluate the sentences in the resulted story datasets to be grammatically corrected and include image generation in the stories that are generated.

In this study, GPT-2 model was used to generate the kid-centric stories. The use of the GPT-3 model in future work will stand better for the image generation along with the text generation. With the imagery generation the application will be more appealing and able to reach more audiences of children.

References

Author, S.B. (2011) 'A utomat e d Story G e n e rati on with M ultipl e Int e rnal Fo c alization', (August).

Chen, J. *et al.* (2021) 'Controllable and Editable Neural Story Plot Generation via Control-and-Edit Transformer', *IEEE Access*, 9, pp. 96692–96699. doi:10.1109/ACCESS.2021.3094263.

Ciolino, M., Kalin, J. and Noever, D. (2020) 'The Go Transformer: Natural Language Modeling for Game Play', *Proceedings - 2020 3rd International Conference on Artificial Intelligence for Industries, AI4I 2020*, pp. 23–26. doi:10.1109/AI4I49448.2020.00012.

Hoppe, S. and Toussaint, M. (2020) 'Qgraph-bounded Q-learning: Stabilizing Model-Free Off-Policy Deep Reinforcement Learning'. Available at: <http://arxiv.org/abs/2007.07582>.

Krukaset, W., Krukaset, N. and Khancome, C. (2018) 'Thai Sentence Generation Machine Employing Fixed Patterns', *Proceedings - 2017 IEEE International Conference on High Performance Computing and Communications Workshops, HPCCWS 2017 and Multicore and Multi-threaded Architectures and Algorithms, M2A2 2017*, 2018-Janua, pp. 70–73. doi:10.1109/HPCCWS.2017.00017.

Lin, J.W., Gao, Y.C. and Chang, R.G. (2019) 'Chinese Story Generation with FastText Transformer Network', *1st International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2019*, pp. 395–398. doi:10.1109/ICAIIIC.2019.8669087.

Qu, Y. *et al.* (2020) 'A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2', *ICEIEC 2020 - Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication*, pp. 323–326. doi:10.1109/ICEIEC49280.2020.9152352.

Rigsby, J.T. and Barbara, D. (2017) 'Automated Storytelling Evaluation and Story Chain Generation', *IEEE International Conference on Data Mining Workshops, ICDMW, 2017-Novem*, pp. 61–68. doi:10.1109/ICDMW.2017.15.

Yu, H. and Riedl, M.O. (2012) 'A sequential recommendation approach for interactive personalized story generation', *11th International Conference on Autonomous Agents and Multiagent Systems 2012, AAMAS 2012: Innovative Applications Track*, 1(June), pp. 248–255.

Zhang, Y., Tran, T.D. and Fangbemi, A. (2016) 'Intuitive Curve Drawing Based Folktale Generator', *Proceedings - 2015 International Conference on Culture and Computing, Culture and Computing 2015*, pp. 87–90. doi:10.1109/Culture.and.Computing.2015.20.

Abbreviations and Specific Symbols

DL	Deep Learning
RNN	Recurrent Neural Network
NLP	Natural Language Processing
AI	Artificial Intelligence
S2S	Sequence to Sequence
LSTM	Long Short-Term Memory
FGS	Folktale Generator System
NLG	Natural Language Generation
GPT-2	Generative Pretrained Transformer
SGF	Smart Game Format

Table 4: abbreviations

Acknowledgment

We would like to sincerely thank our supervisors, anonymous responders for their helpful and continuous support, patience, and motivation given on the case of study. And we thank the open-source community for their contribution to the libraries and software that were utilized for this research project.

Author Biography



VIDURA BANDARA is an alumnus of Ananda College & computer science undergraduate of General Sir John Kotelawala Defence University. He is a university colorsman in several sports. Currently, He is employed at “The AI Team”

as the Project and Operation Manager. He is the Section Student Representative of IEEE Sri Lanka Section. His current research interests include Natural Language Processing, Neural Networks, Artificial Intelligence and Psychology



ASINI SANJA, is a third-year software engineering undergraduate of General Sir John Kotelawala Defence University. She is a graphic designer and a student member of IEEE. Her current research interests include Natural Language Processing, Augmented

Reality and UX/UI.



BUDDITHA HETTIGE, is a Doctor of Philosophy at University of Moratuwa, working on Multi-Agent Technology for Automatic Machine Translation. His current research interests include Natural Language Processing, Sinhala Language Processing,

Computational Grammar for Sinhala, Multi-Agent Systems, Robotics