

Improving Language Generation with Sentence Coherence Objective

Stanford CS224N Custom Project

Ruixiao Sun, Jie Yang, Mehrdad Yousefzadeh

ruixiaos@stanford.edu, jy0829@stanford.edu, mehrdady@stanford.edu

Abstract

Conditional story generation and contextual text continuation have become increasingly popular topics in NLP community. Existing models are often prone to output paragraphs of texts that gradually diverge from the given prompt. Although the generated text may have a reasonable perplexity and diversity, it could easily be identified by human as gibberish. The goal of our project is to improve the coherence and consistency across sentences in a language-generation model. We aim to solve this issue by first training a sentence pair coherence classifier with GPT-2 pretrained model, and then co-train the GPT-2 language model with this new coherence objective using a method analogous to the REINFORCE algorithm. This fine-tuned language model is able to generate lengthy paragraph conditioned on a given topic without diverging too much. The simplicity of this model allows it to be applicable to a variety of underlying language model architecture since it only modifies the final layer of the pre-trained model.

1 Key Information to include

- Mentor (for custom project): Reid Pryzant, Dan Iter
- We would like this report to be graded.

2 Introduction

Conditional story generation and contextual text continuation have become increasingly popular topics in NLP community. The problem setup usually consists of a user-provided prompt sentence, and the story generation model is required to generate text that revolve around the prompt's topic. This problem is different from language modeling in the sense that we are not only interested in predicting the next word given the previous words, but also the coherence between the generated text and the given prompt.

Large pretrained transformer-based models such as BERT [1] and GPT-2 [2] have reached state-of-the-art result on many common NLP tasks. GPT-2 model is a casual language model that predicts the next word conditioned on all the previous words. Naturally, one would be tempted to use this model to make conditional story generation. Indeed, the authors showed some good examples in the original paper. However, from our experience, the generated text still exhibit some degree of incoherence, that is, drifting off the topic for long paragraph of text. See Fig.1 for an example.

Our idea is to leverage the existing architecture of the pretrained transformer-based language model such as GPT-2, and turn it into a conditional story generation model without having too much modification. Compared to other models specifically targeting at story generation [3], our model is faster to training because of the pretraining, and the contextual representation allows it to model longer term dependency.

Input: If you have severe breathing problems, you should get help immediately.

Output: You may not need hospitalization for respiratory problems for which there is pain or even full internal combustion. Breathing is particularly hard in the high-pitched high pitch, low pitched pitches. Air you inhale and exhale can cause a shortening in your breathing, so a ventilator is best. Doping can be administered on your own. If you fall, you must be at least 1 kilometer away from you to be considered for the testing. Your turn-of-the-seat's softness may be a sign that something has gone wrong. Your depth should not exceed the mean of your seat height when wearing the seatbelt.

Diagnosis

I find that certain tests can be useful. For example, when you operate an airbag to fight off an airplane, it might help if you may identify the shoe blade with a blood test. Be careful of vibration due to the uneven ground between your toes. Dry air may also reduce your stomach acid. Additionally, when you are starting to plan an exercise routine, consider the daily nature of the exercise. Check to make sure you are doing a proper recovery.

Figure 1: Example of a GPT-2 text generation output, underlined text shows sentences drifting off topic

3 Related Work

Seq2seq RNN models have been traditionally used to generate conditional stories [4]. More recently, a hierarchical architecture of RNN models proposed by Fan, et al [3] has achieved great performance on conditional story telling by first generate a premise and then the story. The advance of large-scale pretrained transformer models such as GPT-2 [2] has revolutionized lots of NLP tasks including language modeling. Its ability to extract long-distance contextual information leads to its great success in generating text with great semantic and syntactic coherence. GPT-2 has been applied to conditional language generation and image captioned in an encoder-agnostic setting by Ziegler, et al[5]. In this paper, we would like our language model to conform to a new objective that takes into account the correlation between generated text and the prompt text. But this auxiliary objective is non-deterministic with respect to the model parameters since the story is randomly generated from a distribution that depends on the model parameters.

For models where the dependency between the objective function and model parameters are non-deterministic, or models where the objective is of discrete value thus non-differentiable, researchers often resorts to techniques used in Reinforcement Learning community. A specific set of algorithms are categorized as REINFORCE algorithm [6]. Co-training a generation model with auxiliary discrete objective functions has been applied to the field of music generation using Reinforcement Learning techniques, in which the objective consists of a music generation objective and a reward based on music theory [7]. REINFORCE algorithm has been applied in a similar context to text auto-completion system by Lee, et al[8].

4 Approach

To train a language model with an additional coherence objective, we have to first make predictions or evaluate whether two sentences followed a coherent topic. In the first step, we are training a coherence classifier on top of GPT-2 model. This allows us to (1) generate a sentence embedding for the prompt sentence (2) rank the coherence between two sentences. And in the second step, we co-train a modified language model that takes in the prompt sentence embedding as a "shortcut" or "guidance", with an additional coherence objective forcing the generated text to be aligned with the prompt's topic.

Sentence Coherence Prediction: We implemented our own coherence prediction model using GPT-2 as our underlying pretrained transformer and fine-tune it to predict the sentence coherence.

The basic idea is that, if two sentences appear in the same paragraph in our training set, they must have a common underlying topic. The model proposed here will be analogous to the sentence embedding approach first proposed by Kiros, et al [9] and extended by Oord, et al [10]. We encode each sentence by adding [CLS] token to the last position, and feed the hidden state of this token to a double dot-product regression model. The final output is from a logistic regression predicting if the two sentences come from the same paragraph or not. The binary classification problem is formulated as the following:

$$p_{coherence}(e_1, e_2) = \sigma(e_1^T W e_2 + b) \quad (1)$$

where e_1 and e_2 are embeddings for sentence 1 and sentence 2 respectively and σ is the sigmoid function. Aside from fine-tuning the parameters of the transformer model, we have new trainable

parameters, including: weight matrix W , bias b , weights for the [CLS] token. The architecture is shown in Fig.2

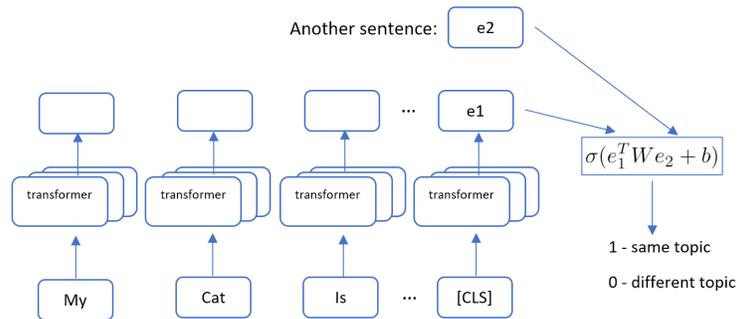


Figure 2: Architecture of the coherence-prediction model

The training step involves:

- Preprocess the dataset. One dataset contains adjacent sentences from the same paragraph, and one dataset contains random sentences from the entire training set.
- Load GPT-2 model with pretrained weights for known word pieces. Add an additional token [CLS] with randomly initialized weights. The hidden state of this [CLS] token will be considered as the sentence embedding. Then we create another double dot-product classifier with random initial weights, which takes as input the two sentence embeddings and predict if they should be adjacent.
- Train the model with a mixed objective: $L = L_{lm} + \lambda L_{coherence}$. Where L_{lm} is the casual language model objective with cross-entropy loss, i.e. $L_{lm} = -\log \frac{\exp(x[label])}{\sum_i \exp(x[i])}$. And $L_{coherence}$ is the binary cross-entropy loss for coherence prediction, which is $L_{coherence} = y \log(\sigma(x)) + (1 - y) \log(1 - \sigma(x))$. Label 1 indicates that two sentences come from the same paragraph. λ is a weighting factor for the coherence objective. For each iteration we have one positive sample and one negative sample picked randomly from the training set.
- Evaluate the model on validation set. We have 50% positive samples and 50% negative samples randomly chosen from validation set. The ultimate criterion is the percentage of accuracy on the test set.

Note that if we concatenate two sentences together, separated by [SEP] token, we could take the last hidden state of [CLS] token and feed it into logistic regression. This proves to be highly efficient and accurate because the two sentences could attend to each other, which is recommended in the original OpenAI GPT paper [11]. However, this approach is infeasible for our problem, because in language generation model, each word is generated on the fly as opposed to given a-priori. We will compare the prediction accuracy of the sentence embedding approach with this cross-attention approach in the following section.

Text Generation Model: In order to generate the text we used Conditional text generation using the auto-regressive models of the GPT-2[2]. GPT-2 is a large transformer-based language model trained on a dataset of 8 million web pages. GPT-2 is trained with a casual language modeling objective, that is: given all of the previous words within some text, predict the probability of the next word[11]. The probability of a sentence in this sense can be written as: $p(x) = \prod_{i=1}^n p(s_i | s_1, s_2, \dots, s_{i-1})$

The original implementation of the conditioned text generation model we used comes from pretrained GPT-2 small model from "Huggingface Transformers" (https://huggingface.co/transformers/model_doc/gpt2.html). It was trained on 12-layer, 768-hidden, 12-heads, 117M parameters. We fine-tuned the model on Wikipedia dataset (<https://blog.einstein.ai/the-wikitext-long-term-dependency-language-modeling-dataset/>)

Co-training objective: in a casual language model, the next token is conditionally dependent on all previous tokens, which is what the original GPT and GPT-2 model does. In our model, however, we

also want our model to be more conforming to the given prompt without diverting to another topic. To achieve this, we added a feed-forward neural network for each hidden state before deciding the output token. The layer takes as input the concatenation of the original hidden state and the sentence embedding, and output a hidden state of the same size as the original hidden state. The scores of each token are then calculated by multiply it to the vocabulary embedding matrix. Written in terms of equations:

$$\begin{aligned} H_1 &= W_1[e_p; h_i] + b_1 \\ H_2 &= W_2 \cdot \tanh(H_1) + b_2 \\ Scores &= VH_2 \end{aligned} \quad (2)$$

See figure.3 for an illustration. The size of the layer H_2 is the same as the output hidden state h_i , because in this way, the vocabulary embedding matrix V can be initialized as the input embedding matrix without any modification. The prompt sentence embedding here is acting as a "shortcut" for the next-token prediction, with the hope that this will force the next token to be more aligned to the given prompt. In the result section, we also compared this architecture with one with linear projection instead of nonlinear layers. And it showed that linear projection could lead to degenerated text.

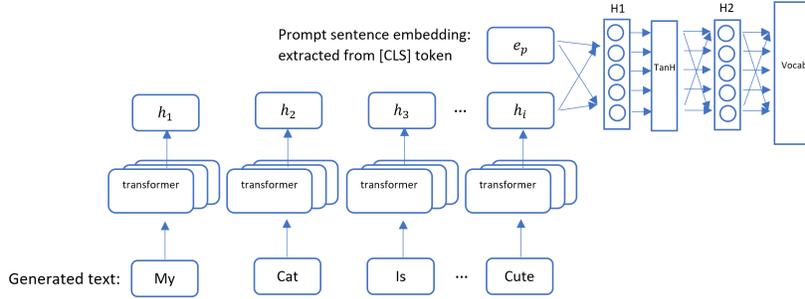


Figure 3: Architecture of the text-generation model

Correspondingly, we have an enhanced language modeling objective. Aside from the cross-entropy loss for next-token-prediction, there is another loss measuring the coherence and similarity between the generated text and the prompt text. Since we used logistic regression in the previous task to classify if two sentences are from the same topic, a natural choice for this loss is the negative of the logit for that classification model. The total loss can be written as $L = L_{lm} + \lambda_2 L_{similarity}$. The same as before, L_{lm} is the casual language model objective with cross-entropy loss, i.e. $L_{lm} = -\log \frac{\exp(x[label])}{\sum_i \exp(x[i])}$. Note that the language model loss L_{lm} is per-word based and the similarity loss $L_{similarity}$ is per-sentence based. We calculate the similarity loss after the model generated a complete paragraph.

During this step, the only parameters we are training are W_1, b_1, W_2, b_2 . The functional relationship between $L_{similarity}$ and the parameters are indirect: we have to generate a whole sentence before plugging it into the coherence prediction model. But the text is sampled from a distribution instead of a deterministic relationship with respect to the model parameters. Considering this difficulty, we decided to use a variant of the REINFORCE algorithm [6] to approximate the gradient w.r.t. the parameters. Written in terms of equations:

$$\begin{aligned} L_{similarity} &= -E[\sum CrossEntropy \cdot (R - \bar{R})] \\ \nabla L_{similarity} &= -E[\sum \nabla CrossEntropy \cdot (R - \bar{R})] \\ R &= e^T W e_p + b \\ \bar{R} &= \alpha R_{current} + (1 - \alpha) R_{previous} \end{aligned} \quad (3)$$

where R is the reward function calculated for each generated text, and \bar{R} is a running exponential average of previous rewards with parameter α . The term $\sum \nabla CrossEntropy$ is the sum of the cross-entropy loss for each of the generated token. The true labels for this cross-entropy loss is set to be the generated text itself. Finally, we perform stochastic gradient descent with this approximated gradient, using a weighting factor λ_2 . The rationale behind this is: if a piece of generated text

received high rewards, meaning a coherent sentence w.r.t. the prompt, we will update the weights such that it generate more sentences like this. And vice versa. Note that the gradient calculation using REINFORCE is noisy and does not converge as quickly as normal gradient descent.

The training procedure involves the following steps

- Preprocess the dataset. Segment the corpus into pieces of sentences of length about 200 tokens.
- Load GPT-2 model with pretrained weights as well as the trained binary classifier.
- Co-train the language model with a coherence objective: $L = L_{lm} + \lambda_2 L_{similarity}$. For each training step, we first train it using casual language model with cross-entropy loss L_{lm} . Secondly, we take the next sentence as the prompt sentence, let the model generate a complete paragraph of text. We then calculate the coherence between the prompt sentence and generated text using our binary classifier. The score is fed into the REINFORCE algorithm 3 to calculate the approximated gradient.

Baselines: Since our tasks are divided into two parts: coherence prediction and text generation, we have different baselines for them. For the coherence prediction problem, we compare sentence embedding approach with cross-attention approach and human evaluation. For text generation, the baseline is the vanilla GPT-2 model without any modification. We are evaluating the consistency of generated text manually to see if there is any jump in topic.

5 Experiments

The training and evaluation are divided into two steps: sentence coherence model and text generation model. Both of them shared the same data set: Wikipedia dataset [12] is applied for training and fine-tuning the model. Currently we trained the model on raw character level dataset WikiText-2.

5.1 Sentence Coherence Model

Data: To train the coherence classification model, we have to randomly choose positive samples and negative samples from the data set. Positive samples come from the same paragraph as the anchor text, negative samples are randomly picked from the entire training set.

Evaluation method: We evaluate the accuracy in percentage on the test set. The accuracy is defined as the number of correctly predicted labels divided by the total number of input pairs (true positive). As describe in the previous section, we will compare the prediction accuracy of the sentence embedding approach with this cross-attention approach. We also compared the results with human performance using 50 sentences, in which 25 of them are positive sample and 25 negative. The human evaluator correctly predicted 46 samples, as well as 3 false positive samples and 1 false negative samples.

Experimental details: In our experiment, hyperparameters are set as follows: weighting factor of the coherence objective is 0.05, the learning rate of Adam optimizer being $5e-5$, the adam epsilon being $1e-8$, number of training epoch being 4, batch size per GPU/CPU for training/evaluation being 2.

Results: The accuracy results are shown in Table.1 and the training process shown in Fig.4. We can see that the coherence classification model achieves human-level performance, and the performance is close the cross-attention classifier. Since the accuracy is near 95%, we can assume that the classifier is a good indicator of the coherence between two sentences, and the underlying sentence embedding is a decent summary of its content.

Methods	Accuracy (%)
Sentence embedding	94.21
Cross-attention	94.52
Human performance	92

Table 1: Accuracy of coherence prediction on the test set

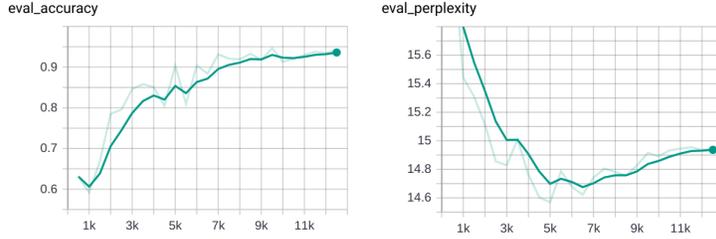


Figure 4: Coherence prediction accuracy and perplexity vs. iteration number on the validation set, visualized using TensorBoard with smoothing=0.6

5.2 Text generation model

Data: We take random sentences of length 300-500 from the training set to train the language model. As for the auxiliary coherence objective, we take a random piece of text of length 100 tokens from the training set as the prompt sentence, generate a 500-token long passage of text, and then do parameter update using REINFORCE algorithm.

Evaluation methods: We will manually check the consistency of the output. An example of text generation output with a prompt input is presented in Fig.1, from which we notice the generated sentences becomes incoherent (the underline part) to the input sentence.

Experimental details: hyperparameters are set as follows: weighting factor of the auxiliary objective is 1, the learning rate of Adam optimizer being $5e-5$, the adam epsilon being $1e-8$, number of training epoch being 4, batch size per GPU/CPU for training/evaluation being 4. The auxiliary objective is active after the second epoch, this is because for the first epoch, the language model is not yet tuned, resulting in an unphysical coherence objective.

Results: The training procedure is illustrated in Fig.5, in which the coherence loss is calculated based on Eq.3. Lower value means better consistency. As can be seen, the coherence loss is very noisy, because: (1) the generated text is randomly sampled (2) the REINFORCE algorithm computed a noisy approximation of the gradient. But the general trend indicates that the coherence loss goes down over time. The perplexity value on the validation set converges to 15.

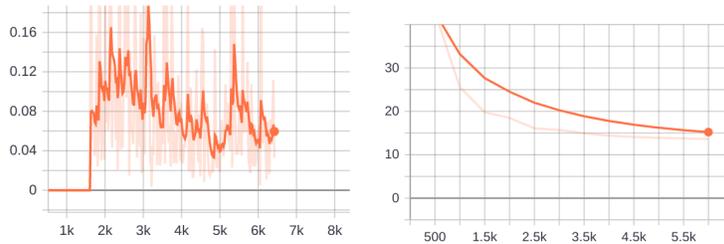


Figure 5: Left: Coherence loss vs. iteration number during training, Right: perplexity value on the validation set during training, visualized using TensorBoard with smoothing=0.8

To compare with the original GPT-2 model, we can look at an example of a generated piece of text from our model, using the same input prompt as Fig.1. We argue that the output from our model is significantly more consistent, and more closely aligned to the prompt sentence’s underlying topic. To see more examples and comparison, please refer to the appendix.

Input: If you have severe breathing problems, you should get help immediately.

Output: If you have severe breathing problems, you should get help immediately. Overriding physical exercise and stripping may cause poor cardiac rhythm. Stale bun cooking can make a fine meal. Timing of cooking changes depending on the range of conditions and of how much time patients wait to start food management. Salt serves as a source of antibiotics, which may be useful, but no toxins are used in comparison to guery instead of sodium. To compare, someone who has pneumonia burns medications they can me feed with salt. If the temperature rises too far or the smell of meat and mineral is suspicious, dig into the ground to see if the leaks cause subdural and granulocytic vein leakage. Leaving room for alveolar water secretion is not difficult because the small volume is sufficiently stable to avoid dehydration. Referring to a caregiver for safety or monitoring analgesia of the caregiver or caregiblite may be of benefit to the caregiver. Pregnant females, where pneumonia burns are not known to lead to cytotoxic injuries, will often be picked up and kept in an unfit environment, such as underground tunnels, or sheltered undercast balconies. Exhail electric heating on exposed walls of buildings where the air temperature is hot is not recommended. When striking a vulnerable person, electric burns may prove important, especially if people have been vulnerable to poisoning. Forced opening of burns has been reported to cause only three to seven days of cardiac difficulties before the pain begins to fail. Finally, the recovery period may be extended for prolonged exposure to pyremosis that causes breathing difficulties. Relically - related burns may involve significant up - risk worsening of heart and respiratory strain. All complications have been based on detailed examinations, but those involving unstable structures such as pulpi test or lymphospiritality have been described as rare, sometimes minor. Note that skin burns are not infection sensitive, such as rarely occurring wounds such as the coaring burn.

Figure 6: Example of a text generation output using our model

6 Analysis

Our language generation model co-trained with an auxiliary coherence objective is able to produce long paragraph of text that is aligned with the underlying topic of the prompt sentence. Pure language models such as GPT-2 will prone to drift off the topic over time. We have tried linear projection and nonlinear transformation of the prompt sentence embedding onto the next word prediction, results indicate that linear transformation often leads to degenerated text with repetitions, whereas nonlinear models have great performance (See appendix for an example). The reinforcement learning loss does not monotonically decrease with time. But over a longer period, the average loss does get lower.

Although the generated text from our model is able to stick to the prompt sentence's topic, the wording and grammar are sometimes not as natural as the pure GPT-2 language model. We believe this is a price to pay when we force each generated word to align with a given topic.

7 Conclusion

In this work, we have introduced the problem of text incoherence when applying a pure language model to a conditional story-generation problem. We proposed a new conditional language generation model following two steps. The first step involves training a sentence coherence classifier, which also enables us to generate a sentence embedding for each prompt sentence. The second step involves co-training a conditional language model with an auxiliary coherence objective. This new objective is optimized using REINFORCE algorithm. After some experiments, we found out that our model is able to produce long text without diverging from the main topic. However, the wording and grammar are sometimes not as natural as the pure GPT-2 language model. Future work will be focused on how to improve the fluency of the output while still maintaining the coherence between generated text and the prompt.

References

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [2] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9, 2019.

- [3] Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. *arXiv preprint arXiv:1805.04833*, 2018.
- [4] Melissa Roemmele. Writing stories with help from recurrent neural networks. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [5] Zachary M Ziegler, Luke Melas-Kyriazi, Sebastian Gehrmann, and Alexander M Rush. Encoder-agnostic adaptation for conditional language generation. *arXiv preprint arXiv:1908.06938*, 2019.
- [6] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.
- [7] Natasha Jaques, Shixiang Gu, Richard E Turner, and Douglas Eck. Generating music by fine-tuning recurrent neural networks with reinforcement learning. 2016.
- [8] Mina Lee, Tatsunori B Hashimoto, and Percy Liang. Learning autocomplete systems as a communication game. *arXiv preprint arXiv:1911.06964*, 2019.
- [9] Ryan Kiros, Yukun Zhu, Russ R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Skip-thought vectors. In *Advances in neural information processing systems*, pages 3294–3302, 2015.
- [10] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
- [11] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. URL https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language_understanding_paper.pdf, 2018.
- [12] Stephen Merity. The wikitext long term dependency language modeling dataset. *Salesforce Metamind*, 9, 2016.

A Appendix

In the following, we are going to show the conditional story generated from 3 models: our proposed model, original GPT-2 model, and a model with linear projection from sentence embedding to the output (see experiment section for more details). All 3 examples followed the same prompt sentence, which described a new gaming console called Playstation 5.

Similar to our previous findings, the proposed model is able to generate long passage of text without drifting off the topic, whereas GPT-2 model gradually diverge from the given prompt topic. Linear model does not have a good performance and is prone to stuck in self-repetition.

Input: PS5 (or PlayStation 5) is the next-generation PlayStation, releasing in late 2020. Sony has remained pretty tight-lipped about its next-gen console but the company recently gave us our first official look behind the curtain during its first PS5 reveal event.

Output: PS5 (or PlayStation 5) is the next-generation PlayStation, releasing in late 2020. Sony has remained pretty tight-lipped about its next-gen console but the company recently gave us our first official look behind the curtain during its first PS5 reveal event. There are a lot of new accessories, including the PlayStation Accessories 360 2, PlayStation 360 PlayStation 2, PlayStation 3, PlayStation Vita and PlayStation Vita wireless accessories. The PlayStation 360 and PlayStation 3 are also equipped with a video console capable of 360 - screen compatible wireless connectivity. This is only implemented on the PSP's original DSi gaming controller, but as of 2013, the PlayStation Vita is able to attach to the console. It was later announced that both consoles would be sold in North America, and after a nine - month development period it was announced that PlayStation 2 and PS2 would be shipped in Japan via PlayStation Stores.

PlayStation 3 is one of the oldest portable video game consoles out of the past. Featuring two of the key third - generation PlayStation 3 games, the original PlayStation 3 was one of the earliest consoles to operate independently of the Internet in the 19th century. As part of their run on the series, Nintendo, Capcom and Capcom designed and produced several arcade games throughout the 20th century, including a " Nintendo DS " game. A third - generation version of the console was first released in the 1990s, but both generations supported the consoles later. With the release of the PlayStation 4, the game platform has changed more and more to support a Japanese version of the PlayStation III and later PS4. All three of the PlayStation 3 PlayStation 4 versions are ported to the Wii, which is another feature that might have been backward - determined at that time, but the ports were still developed prior to the development of the PlayStation 3.

The PlayStation 3 is the main platform for video game video game games. Until 2004, the PlayStation 2 was an older, dual - controller PS2 game console. A game variant was developed for the Xbox 360 and Wii, which was released in 2010. It was developed by Namco Media, published by Sony, and was released in two versions that were three times larger than the previous release. These games contained a mechanical controller for actions, the gameplay elements of which could be manually determined by navigating the game mode's graphics cards and a three - dimensional graphic where possible. According to the Nintendo DSit eGS, the gameplay of a PlayStation 2 version was both hard and reasonably difficult. Later games in its franchise, such as Legends of Mortal Kombat and King of the Ring, also had difficulty considering the three - dimensional graphics due to the graphics system's limited dimensions.

Note : Commentation of the version that first appeared in IGN, a two - dimensional model of the PlayStation 2.

The gameplay of the PlayStation 3 is similarly similar to one of the other PlayStation 3 consoles, the PlayStation 2. The graphics are less vibrant than the PlayStation 2, and colors and features are compared to the PlayStation 2's gameplay. However, the two versions in itself are far more enjoyable than the two in its own game. The PlayStation 2's gameplay has the same Game Boy flavor, as are the PS2's versions, with voice actors based on Japanese creators, some of whom had voiced later versions of the games in series. Consuming 10 @,@ 000 calories from a cook is one of the main fun games available to the player. In its original Game Boy form, the games are watched by over 2 @,@ 000 people, several of whom like to contribute to the franchise with the various voice roles available to the players. As with other consoles, when the player enters the Game Boy, they may cause injury to the screen when it is damaged ; some minor glitches cause some bugs and certain glitches may be eliminated. Throughout the series, the characters and sections of the environment, the camera, and the weapons are reused from games such as Mario & Dronicon. In addition to the character sprites, the PlayStation 2 also includes voice actors for other Capcom series characters such as X - Commander, Mega Man & Ace, and DJ Love.

Figure 7: Example of a text generated using our proposed model

Input: PS5 (or PlayStation 5) is the next-generation PlayStation, releasing in late 2020. Sony has remained pretty tight-lipped about its next-gen console but the company recently gave us our first official look behind the curtain during its first PS5 reveal event.

Output: PS5 (or PlayStation 5) is the next-generation PlayStation, releasing in late 2020. Sony has remained pretty tight-lipped about its next-gen console but the company recently gave us our first official look behind the curtain during its first PS5 reveal event. The PlayStation 4 is meant to be a " cross between the Xbox 360, PlayStation 3 and Xbox 360. " During the press conference Sony showcased several new features. PlayStation All - Star mode, in which the player controls the entirety of the team, will be available alongside Online Play. Aside from a mode where all four members of the PSN team play together, Sony expects other features to include cooperative multiplayer online component support and higher resolution video.

Various features of the PlayStation 4 were announced during its live event during its reveal event in Japan. Wearers of PlayStation All - Star Mode and Friends Mode received an exclusive emblem from Sony, indicating that they would be appearing at the event in early Summer. Other features in the demo included a multiplayer online battle arena, the ability to play up to six characters, and Sony's ability to stream music from the PlayStation Network via PlayStation Pass. PlayStation All - Star mode was also announced at the 2013 International PlayStation Experience event, during which the player would play characters in front of a live audience of gamers attending the Sony PlayStation All - Star Special presentation in Tokyo. PlayStation Now's smartphone game recognition system was also shown during the show in January 2014. It was expected that Sony will include more features in future games when its next generation of consoles are released, such as PlayStation Move support for its PlayStation Move controllers and support for the Game Boy Advance console's touchscreen. In addition, the PlayStation 2 is slated to support downloadable content in a future game update.

" Chariot of Cerberus " is the seventh episode of the first season of the American science fiction television series The X - Files. It first aired on January 11, 1993 on the Fox network. It is the eighteenth episode overall and the seventy - fifth episode overall. It was written by series creator and executive producer Mark Millar, and directed by series producer Lawrence Thomas.

The episode was created in the summer of 1993 by the producers of the original series and was inspired by discussions about the 2006 film Spyro and the case of Roger Moore, the three murder cases. Millar said that he thought it was " very popular " at the time and " surprising " at the time, and that he was aware that many television series could go to different lengths to make their original series funnier than their licensed creations. It was a " stark contrast to their current format " and a product of " a couple of editors in a heartbeat ". As with most series, the show dealt with detective work, case of Red Dog, and various social issues. The episode received mixed reviews. Many called it a " fantastic episode ", while some critics praised the overall plot, and said that it " handles the case fairly well ". The episode was among the first to be shown at the Fox network premiere of the second season and the seventy - sixth episode overall.

While making the episode, Millar realized that he wanted to use it in a storyline that " sounded interesting " for the show, to show a middle ground between the rest of The X - Files and the American science fiction television series. Originally, this would have been a storytelling gag, but later in the season, writer Cameron Helligau decided to expand the idea and cut out the gay references and tell it with less side effects. He wanted to make the new series more of a real - life story rather than a mystery. Initially, the idea was " weird ", but Helligau eventually settled on adding as much humor as possible. The episode aired in the United States on ABC in the United Kingdom on January 11, 1993 and in the United States on February 13, 1993. The episode received a mixed reception from critics. Some people, such as Seth MacFarlane, felt that the episode was an easy flop, and said it was " amateurish ". Others, such as ABC television critic John Roos, liked the addition of the " joyfully snarky " character Casterman. Roos noted that!

Figure 8: Example of a text generated from the original GPT-2 model. Underlined text shows the portion of text diverging from the given topic

Input: PS5 (or PlayStation 5) is the next-generation PlayStation, releasing in late 2020. Sony has remained pretty tight-lipped about its next-gen console but the company recently gave us our first official look behind the curtain during its first PS5 reveal event.

Output: PS5 (or PlayStation 5) is the next-generation PlayStation, releasing in late 2020. Sony has remained pretty tight-lipped about its next-gen console but the company recently gave us our first official look behind the curtain during its first PS5 reveal event.

The PlayStation 5 arrives in April 2016 in North America, in December in Europe and in Japan on March 29, 2016 in North America. Although the update contains a 4 @.@ 4GB version of the same memory as the original, some of the adjustments are cosmetic. As part of the Microsoft Windows update of the Xbox 360 controller, Microsoft updated the memory controller itself.

While in development, PlayStation Network began functioning on June 5, 2007. During the development process of the PlayStation Network, Microsoft's partners made changes that made accessing PlayStation Network more visible via gestures than via traditional video chat. Although it wasn 't immediately clear whether the changes affected the number of possible players on the system, on July 5, 2007, Microsoft announced that there would be no visible " freeze " on the Network.

Sony still owns up to the initial launch of the operating system, as was revealed in a leaked BIOS document. The GameWorks team did initially think about bringing the operating system on the PlayStation Network, but pushed back the company's initial executive search effort, which includes " compatibility with various types of PSP and Wii game systems ". On August 12, 2007, the organization began seeking out defectors on Microsoft's support forums and in the hope of finding internal changes. During the testing, eight Sony defectors had opted out, one of which was an Rtek rig. On September 4, 2007, one of the defectors left the defectors's support section. On November 8, 2007, the defectors left PlayStation Network in order to speak with Sony officials about defectors. An individual who asked not to be named disclosed that he worked on several personal projects in addition to the PlayStation 2. On December 7, 2007, two defectors left the support section to speak with Nintendo officials about defectors, one of whom went by the name " Eugene ". On January 10, 2008, three defectors left the support section to speak with CEO Kaz Hirano about defectors. On February 7, 2008, on March 7, a fourth defector left the support section, and on April 1, 2008, on May 5, 2008, on June 7, 2008, on June 7, 2008, and on July 10, 2008, on July 10, 2008, on July 13, 2008, on September 14, 2008, on September 14, 2008, and on October 2, 2008. On September 16, 2008, on January 14, 2009, an individual from Finland went on a two - week trip to South Korea, where he was staying with defectors, on January 19, 2009, and on April 2, 2009, after traveling to South Korea by private jet from Japan. On March 31, 2009, on March 31, 2009, a woman with a black dress went on a five - month trip to South Korea with defectors, on March 31, 2009, and was staying with defectors on April 6, 2009. On April 1, 2009, a person from Turkey went on a six - month trip to South Korea with defectors, on April 1, 2009, and was staying with defectors on April 1, 2009. On April 10, 2009, an individual from Azerbaijan went on a five - month trip to South Korea with defectors, on April 10, 2009, and was staying with defectors on April 1, 2009.

Figure 9: Example of a text generation output using a linear model. Notice the repetition in the last paragraph