A Experiment settings

A.1 Hyper-parameter

Adam (Kingma and Ba, 2015), with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-6}$, and L2 regularization factor 0.01, was used as the optimizer. We use linear warmup over the first 6% of the training steps and linear decayed. The dropout rate was 0.1, a batch size was 32, and label smoothing (Szegedy et al., 2016) was 0.1.

Summarization task The learning rate was set to 2×10^{-5} for both models. However, since the weights of the newly added special tokens were trained from the initialization state, the learning rate for the word embedding weights was set higher, namely, 1×10^{-3}. Number of epochs was set to 10. During inference, the summary text was generated by using a beam search. For the CNN/DM dataset, the number of beams was 4, and the length penalty was 2.0. For the XSum dataset, the number of beams was 6, and the length penalty was 1.0. Maximum number of tokens for the source document was set to 1024, and maximum number of tokens for the summary text was set to 128. Any text with a higher number of tokens was truncated at the end.

Story generation task The learning rate was set to 2×10^{-5}, and the learning rate for word embedding weights was set to 1×10^{-3}. Number of epochs is 30. During generation, texts were generated by top-p sampling (where $p = 0.95$). The temperature was set to 0.1. Maximum number of tokens for the target text was set to 128, and any text with a higher number of tokens was truncated at the end.

Hyper-parameter search The batch size was selected from $\{32, 64, 128, 256\}$. The learning rate was selected from $\{1 \times 10^{-5}, 2 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, and 2 \times 10^{-4}\}$. Number of epochs was selected from $\{3, 5, 10, 15, 20\}$ for summarization and $\{10, 20, 30, 60, 100\}$ for story generation. Due to the large search area, we used the manual search to determine parameters. To determine the hyper-parameters, we used the accuracy of the keyword-position control listed in Table 1.

A.2 Environment

We used $4\times$ V100 GPUs for training and $1\times$ V100 GPU for inference. Training on the CNN/DM, XSum, and ROCStories datasets took 30, 40, and 4 hours, respectively, and inference took less than 1 hour for all datasets.

A.3 Dataset and Control tokens

The statistics of the dataset are shown in Table 6. For the CNN/DM dataset, we followed the data split proposed by Yoon et al. (2020). For the XSum dataset, we followed the data split proposed by Narayan et al. (2018). For ROCStories, we split the data into 8:1:1 for training, development, and test sets. Control tokens (keywords, each keyword position, and text length) were extracted from the target text and given to the model for training. Word tokenization of the text was done by NLTK library[6] to obtain keywords and text length. Note that the model receives the text tokenized into subwords. Therefore, the number of tokens the model receives differs from the pre-calculated text length.

In training, the model was provided with all control tokens (keywords, keyword position, and text length) each with a certain probability. The trained model was then used to generate texts under four different settings (Keyword, +Len, +Pos, and +Len+Pos in Table 1). In this way, a single model can perform inference experiments in multiple settings in a manner that lowers the cost of the experiments. Preliminary experiments showed that performance of each model is almost the same as when each model is trained for each inference setting. However, for regular text generation that does not use control tokens ("w/o Control" in Table 1), model training and inference were performed separately from the model described above.

Keywords A sequence consisting of one to three consecutive words was obtained from the target text as keyword candidates. Phrases whose first word is a stop word or a frequent word were excluded from the keyword candidates because they are considered unlikely to be given as keywords by the user. During training, no more than three keywords were randomly selected from the keyword candidates for each epoch and given to the model. During inference, one to three keywords were given to the model in the experiment of Table 1, and one keyword was given to the model in the experiment of Table 2 and Table 3. In both training and inference, when multiple keywords are selected from candidate keywords, they are se-

[6] https://github.com/nltk/nltk/blob/develop/nltk/tokenize/__init__.py

Datasets	#train	#dev	#test	#source document tokens	#target text tokens
CNN/DM	287,227	13,368	11,490	777.6 (399.5)	57.9 (24.6)
XSum	204,045	11,332	11,334	433.1 (355.8)	23.2 (5.8)
ROCStories	78,528	9,816	9,817	–	49.8 (9.4)

Table 6: Dataset statistics: number of training data, number of development data, number of test data, number of words in source document and its standard deviation, and number of words in target text and its standard deviation.

Control	CNN/DM S-BLEU↓	XSum S-BLEU↓	ROCStories S-BLEU↓
w/o Control	96.2	92.2	–
Keyword	96.6	92.8	71.2
+Pos+Len	**85.9**	**83.1**	**56.9**

Table 7: Evaluation of diversity of generated texts using the Self-BLEU metric

Control	CNN/DM		XSum		ROCStories	
	Include	Pos	Include	Pos	Include	Pos
Specifying oracle Position						
+Pos+Len	85.8	48.8	91.8	64.1	58.8	29.1
Specifying random Position						
+Pos+Len	84.2	36.6	89.0	43.1	58.7	23.8

Table 8: Comparison between specifying oracle keyword positions and random keyword positions. The keywords given to the model are single oracle keywords extracted from the target text, and they are the same in both settings.

lected so that one keyword is not part of another (e.g., "looking" vs "looking forward").

Keyword Position Relative positions of the above keywords in the target text were obtained and given to the model. However, for each keyword with a probability of 10%, the keyword position was not given to the model, and only the keyword was given during training.

Text Length Word length of the target text was obtained and given to the model as text length. Also, with a probability of 10%, text length was not given to the model during training.

B Generating diverse texts

Here, we show that by controlling keyword positions, it is possible to generate a variety of texts from a specific keyword. The ability to generate diverse texts will enable users to select their intended text from multiple generated texts. When normal generation ("w/o Control") and with keyword generation, 10 different texts are generated from a single input[7]. When keyword position is used, the model is given 10 different keyword positions for one particular keyword, and 10 different texts are generated. For each generated text, the diversity is evaluated with the Self-BLEU (Zhu et al., 2018) score. The results in Table 7 show that the generated text diversity is improved by providing a variety of keyword positions. In particular, beam search was used to generate multiple texts in summarization, and it resulted in very low diversity in the generated texts. This reduction in diversity, however, was mitigated by controlling the keyword position. The examples of generated texts in Table 4 show that controlling keyword position produces a variety of valid texts from a particular keyword.

[7] As indicated in Appendix A.1, beam search was used for summarization, and top-p sampling was used for story generation.

C Generation specifying random positions

In section 3.2, we showed that it is possible to generate texts by specifying oracle keyword positions extracted from target texts. We also show that it is possible to generate text by specifying arbitrary keyword positions. Specifically, we compare the accuracy of position control when oracle keyword positions are given and that when randomly selected keyword positions are given. Note that the keywords given to the model are single oracle keywords extracted from the target text, and they are same in both settings.

As shown by the results in Table 8, position control is still possible when a random position is specified. However, when random positions are specified, the accuracy of keyword inclusion is slightly lower, and the accuracy of keyword-position control is significantly lower. That result can be explained by the fact that keyword positions that are difficult to satisfied can be specified. For example, a keyword originally used at the end of the text is difficult to place at the beginning of the text.

D Generated samples

Examples of texts generated by the proposed method as given in Table 9, Table 10, and Table 11.

E Limitations

Using oracle information In this paper, text generation is controlled by providing the model with control tokens extracted from the target text. The improved accuracy of keyword inclusion and position control shown in our experiments is due to this additional information, not to improved performance of the model itself. Because the goal of this paper is to enable users to control the model by providing additional information such as keywords and positions, this experiment design is not a mistake. However, selecting appropriate keywords and placing them in the appropriate positions without relying on oracle information is one of the challenges for future work.

Using length information The proposed method requires that the model be given a target text length, which may impose an extra burden on the user in practical terms. Experimental results showed that length information itself is not essential for controlling relative position, but it is one key to improving performance. Lee et al. (2018) proposed a method for predicting target-text length from the source document used in machine translation. By incorporating this method, it may be possible to control the relative positions of keywords without providing additional length information.

Insufficient performance The experiment results in Table 1 show that accuracy of keyword inclusion and keyword-position control is low, especially in the case of story generation. The reason for this low accuracy may be that the model does not generate the appropriate context for the inclusion of keywords because the source document is not given. In the summarization, accuracy of the keyword-position control is also far from 100% control accuracy. Since it is possible to extract only desirable text from the generated text, the control need not necessarily succeed 100% of the time. However, if success rate of the control is improved, efficiency of text generation will improve. A deeper investigation of the cause of poor performance and improvement in control accuracy are challenges for future work. One idea to improve performance of story generation is to give the model several words at the beginning of the text, and this approach may make it easier for the model to generate the appropriate context.

<table>
<tr><th colspan="2">Source document</th></tr>
<tr><td colspan="2">Never mind cats having nine lives. A stray pooch in Washington State has used up at least three of her own after being hit by a car, apparently whacked on the head with a hammer in a misguided mercy killing and then buried in a field – only to survive. Thatś according to Washington State University, where the dog – a friendly white-and-black bully breed mix now named Theia – has been receiving care at the Veterinary Teaching Hospital. Four days after her apparent death, the dog managed to stagger to a nearby farm, dirt-covered and emaciated, where she was found by a worker who took her to a vet for help. She was taken in by Moses Lake, Washington, resident Sara Mellado. “Considering everything that sheś been through, sheś incredibly gentle and loving,” Mellado said, according to WSU News. “Sheś a true miracle dog and she deserves a good life.” Theia is only one year old but the dogś brush with death did not leave her unscathed. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity – and still requires surgery to help her breathe. The veterinary hospitalś Good Samaritan Fund committee awarded some money to help pay for the dogś treatment, but Mellado has set up a fundraising page to help meet the remaining cost of the dogś care. Sheś also created a Facebook page to keep supporters updated. Donors have already surpassed the $10,000 target, inspired by Theiaś tale of survival against the odds. On the fundraising page, Mellado writes, “She is in desperate need of extensive medical procedures to fix her nasal damage and reset her jaw. I agreed to foster her until she finally found a loving home.” She is dedicated to making sure Theia gets the medical attention she needs, Mellado adds, and wants to “make sure she gets placed in a family where this will never happen to her again!” Any additional funds raised will be “paid forward” to help other animals. Theia is not the only animal to apparently rise from the grave in recent weeks. A cat in Tampa, Florida, found seemingly dead after he was hit by a car in January, showed up alive in a neighborś yard five days after he was buried by his owner. The cat was in bad shape, with maggots covering open wounds on his body and a ruined left eye, but remarkably survived with the help of treatment from the Humane Society.</td></tr>
<tr><th colspan="2">Gold text</th></tr>
<tr><td colspan="2">Theia, a bully breed mix, was apparently hit by a car, whacked with a hammer and buried in a field. “Sheś a true miracle dog and she deserves a good life,” says Sara Mellado, who is looking for a home for Theia.</td></tr>
<tr><th>Keyword & Position</th><th>Generated text</th></tr>
<tr><td>w/o Control</td><td>A dog apparently hit by a car and buried in a field appears to have survived. Four days after her apparent death, the dog stagger to a nearby farm. She was found by a worker who took her to a vet for help. A fundraising page has raised more than $10,000 for the dog’s care.</td></tr>
<tr><td>true miracle dog
(w/o Position Control)</td><td>Stray dog apparently hit by a car and buried in a field survives. She was found emaciated and dirt-covered by a farm worker four days later. The dog, named Theia, has a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than $10,000.</td></tr>
<tr><td>true miracle dog (0%)</td><td>“She’s a **true miracle dog** and she deserves a good life,” foster mother says. Theia was apparently hit by a car and buried in a field. Four days later, she was found emaciated and dirt-covered by a farm worker. A fundraising page has raised more than $10,000.</td></tr>
<tr><td>true miracle dog (40%)</td><td>A stray dog apparently hit by a car and buried in a field survives. The dog, named Theia, is “a **true miracle dog**,” her foster owner says. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than $10,000.</td></tr>
<tr><td>true miracle dog (90%)</td><td>Stray dog apparently hit by a car and buried in a field survives. Four days later, she is found emaciated and dirt-covered on a farm. She has a dislocated jaw, leg injuries and a caved-in sinus cavity. A good Samaritan has set up a fundraising page for “the **true miracle dog**“</td></tr>
<tr><td>hit by (40%)
bully breed mix (60%)</td><td>A stray dog in Washington State apparently died after being **hit by** a car. Four days later, she was found emaciated and dirt-covered in a field. Theia, a **bully breed mix**, has a dislocated jaw, leg injuries and a caved-in sinus cavity.</td></tr>
<tr><td>hit by (10%)
bully breed mix (0%)</td><td>Theia, a **bully breed mix**, was apparently **hit by** a car and buried in a field. Four days later, she was found emaciated and dirt-covered by a farm worker. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than $10,000.</td></tr>
<tr><td>apparently (10%)
field (40%)
miracle dog (60%)</td><td>A stray dog in Washington State **apparently** died after being hit by a car. Four days later, the dog was found in a **field**, dirt-covered and emaciated. “She’s a true **miracle dog** and she deserves a good life,” the dog’s foster owner says.</td></tr>
<tr><th>Keyword & Position</th><th>Generated text (Error cases)</th></tr>
<tr><td>buried in (80%)</td><td>Stray dog apparently hit by a car, apparently killed and **buried in** a field. Four days later, the dog manages to stagger to a nearby farm. She was found by a farm worker, who took her to a vet for help. Theia is now receiving treatment.</td></tr>
<tr><td>whacked (70%)</td><td>A stray dog in Washington State apparently died after being hit by a car. Four days later, she was found emaciated and dirt-covered by a farm worker. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than $10,000.</td></tr>
</table>

Table 9: Examples of generated texts from the CNN/DM dataset. This table is the complete version of Table 4 with source document, gold summary, and additional examples.

<table>
<tr><th colspan="2">Source document</th></tr>
<tr><td colspan="2">“Iḿ really looking forward to it - the home of Scottish football,” said Rodgers ahead of his maiden visit. “I hear the pitch is good, a nice big pitch suits the speed in our team and our intensity. “The technical area goes right out to the end of the pitch, but you might need a taxi to get back to your staff.” This will be Rodgersśecond taste of the Old Firm derby and his experience of the fixture got off to a great start with a 5-1 league victory at Celtic Park last month. “It was a brilliant performance by the players in every aspect,” he recalled. “Obviously this one is on a neutral ground, but weĺl be looking to have a similar performance. “Weĺl be prepared and focused. We know itś going to be a tough game. We anticipated that the last time.” Rodgers is also aware Celticś visit to Hampden last season in the Scottish Cup semi-final against Rangers ended in defeat. “The last experience there wasnt́ so good for Celtic,” he said. “But itś a different team with a different mentality and weĺl look to take that mentality into the game.” Rodgers lost two semi-finals in his time at Liverpool and is aiming to make it third time lucky at the club he joined in the summer. “You have to perform - thatś whatś important and if you get little bit of luck along the way then hopefully youĺl get the result that you want.” said the Northern Irishman. “So, for us, itś really looking at our performance level, which in the main has been at a really high level. “My focus is always on my own team and, if we impose our style of play, it can give a lot of problems to opponents.” Unlike predecessor Ronny Deila, Rodgers would not be drawn on this semi-final being a step on the way to a potential domestic treble. “Itś very, very difficult to achieve,” he said. “Thereś been great managers here in the past that havent́ achieved that. “But itś the first competition where wevé a chance to win a trophy and Ivé always said that itś a priority for us because itś the first one. “Our idea was to qualify for the Champions League, which we did, to make a good start in the league, which wevé done, and then look to get to the League Cup final. “Unless you can do that then you dont́ have the chance to win the trophies.</td></tr>
<tr><th colspan="2">Gold text</th></tr>
<tr><td colspan="2">Manager Brendan Rodgers is sure Celtic can exploit the wide open spaces of Hampden when they meet Rangers in Sunday’s League Cup semi-final.</td></tr>
<tr><th>Keyword & Position</th><th>Generated text</th></tr>
<tr><td>w/o Control</td><td>Celtic manager Brendan Rodgers is relishing his first visit to Hampden Park as he prepares for Saturday’s Scottish Cup semi-final against Rangers.</td></tr>
<tr><td>Hampden
(w/o Position Control)</td><td>Celtic manager Brendan Rodgers is relishing the prospect of playing his first Scottish Cup semi-final against Rangers at **Hampden** Park on Saturday.</td></tr>
<tr><td>Hampden (0%)</td><td>**Hampden** Park will be the ‘home of Scottish football’ when Celtic take on Rangers in the Scottish Cup semi-final, says manager Brendan Rodgers.</td></tr>
<tr><td>Hampden (40%)</td><td>New Celtic manager Brendan Rodgers is relishing the prospect of playing at **Hampden** Park for the first time in the Scottish Cup semi-final against Rangers.</td></tr>
<tr><td>Hampden (80%)</td><td>Celtic manager Brendan Rodgers is relishing the prospect of playing his first Scottish Cup semi-final against Rangers at **Hampden** Park.”</td></tr>
<tr><td>Brendan (20%)
meet Rangers (50%)</td><td>Celtic manager **Brendan** Rodgers is relishing the opportunity to **meet Rangers** in the Scottish Cup semi-final at Hampden.</td></tr>
<tr><td>Brendan (80%)
meet Rangers (0%)</td><td>Celtic will **meet Rangers** in the Scottish Cup semi-final at Hampden Park for the first time on Saturday, but manager **Brendan** Rodgers is relishing the occasion.</td></tr>
<tr><td>Sunday’s League (0%)
Hampden when they (50%)
exploit the (80%)</td><td>Celtic manager Brendan Rodgers says **Sunday’s League** Cup semi-final against Rangers at **Hampden when they** meet will be a chance to “**exploit the** big pitch”.</td></tr>
<tr><td>Sunday’s League (40%)
Hampden when they (30%)
exploit the (10%)</td><td>Celtic will look to **exploit the** atmosphere at **Hampden when they** face Rangers in **Sunday’s League** Cup semi-final, says manager Brendan Rodgers.</td></tr>
<tr><th>Keyword & Position</th><th>Generated text (Error cases)</th></tr>
<tr><td>Manager Brendan (80%)
meet (50%)</td><td>Brendan Rodgers is relishing his first visit to Hampden Park as Celtic manager when his side meet Rangers in the Scottish Cup semi-final on Wednesday, 7 May.</td></tr>
<tr><td>semi-final. (40%)</td><td>New Celtic manager Brendan Rodgers is relishing the prospect of making his debut at Hampden Park in Saturday’s Scottish Cup semi-final.</td></tr>
</table>

Table 10: Examples of generated texts from the XSum dataset. In the second error case, “semi-final.” (the last word is a period) is given as a keyword. In such case, when the position where the keyword should be placed is obvious, the model tends to ignore the positional control token and place the keyword in the appropriate position.

Gold text	
Bob wanted a car of his own to drive to work. Bob began saving money from his paycheck to buy a car. Eventually Bob saved enough money to buy his very own car. Bob went to a local car dealership and browsed the inventory. Bob found a car he liked and bought it from the dealership.	
Keyword & Position	**Generated text**
dealership (w/o Position Control)	I went to the **dealership** to buy a new car. I was very nervous because I didn't know what I wanted. I went to the dealer and asked them what I wanted. They told me that I wanted a convertible with a leg drive. I drove the car home and it was a great car.
dealership (0%)	I went to the **dealership** to look at a car. I was very nervous about buying a car. I asked the salesman if I could try out a certain car. The salesman told me that I could try out a blue car. I drove the blue car home and loved it so much, I bought it.
dealership (30%)	I went to the **dealership** to buy a new car. I was very nervous about the car and the price. I went to the **dealership** and looked at the price. I decided to buy the car and I was very happy with it. I am glad I went to the **dealership** because it was a great deal.
dealership (80%)	I went to the **dealership** to look at a car. I was very nervous because I didn't know what I wanted. I asked the salesman if I could pick out a car. The salesman told me that I could get a new car if I paid $40,000. I drove to the **dealership** and bought a new car.
saving money (10%)	I was **saving money** for a new car. I had been **saving money** for a long time and was ready to buy it. I went to the car dealership and looked at all the models. I decided on a Ford Mustang and immediately signed up. I drove it home and was so happy with my purchase.
saving money (60%)	I was in a hurry to get to work. I had to hurry because I had to hurry so fast. I was going fifteen miles an hour faster than I was. I was **saving money** so I could get to work. I was glad I was able to make it to work on time.
saving money (80%)	I was in a hurry to get to work. I had to hurry because I didn't have my car keys. I looked everywhere for my keys, but couldn't find them. I finally found them under my coat hanger, in my coat pocket. I was relieved that I had **saved money** on my car keys."
drive to work (50%) saved enough (70%)	I was driving to work one day when I saw a car in the road. I pulled over and asked if I could **drive to work**. The driver told me that he had **saved enough** money to buy a new car. I drove to work and paid him back. I drove to work and paid him back and he was very happy.
drive to work (20%) saved enough (0%)	I **saved enough** money to buy a new car. I went to the car dealership to test drive my new car. I drove the car for a few hours before I left. When I got home, I realized I had forgotten my wallet. I had to **drive to work** to get my wallet back, but I was happy.
Keyword & Position	**Generated text (Error cases)**
began saving money (90%)	I was in a hurry to get to work. I had to hurry because I didn't have my car keys. I looked everywhere for my keys, but couldn't find them. I finally found them under my coat, and I was relieved. I was able to grab my keys and walk to work without losing my car keys.
local (0%) Bob (40%) enough (60%)	**Bob** was a **local** handyman. He was hired to fix up a leaky roof on his home. **Bob** was very handy and did a good job at it. Unfortunately, the roof was too deep and the water would not come out. **Bob** had to call a **local** handyman to fix the leaky roof.

Table 11: Examples of generated texts from the ROCStories dataset. Examples of generated texts from the CNN/DM dataset. This table is the complete version of Table 5 with gold story and additional examples. As can be seen, even if the keyword "saving money" is given to the model, for example, it may be used as another expression, such as "saved money". Note that in the quantitative evaluation of this study, such cases where the keyword was paraphrased are classified as "the keyword was not included in the text". In some cases, the given keyword is included in the text more than once.