Appendix for Paper "LOT: A Story-Centric Benchmark for Evaluating Chinese Long Text Understanding and Generation"

Jian Guan[1], Zhuoer Feng[1], Yamei Chen[1], Ruilin He[2],
Xiaoxi Mao[3], Changjie Fan[3] and Minlie Huang[1][*]

[1]The CoAI group, DCST; [1]Institute for Artificial Intelligence; [1]State Key Lab of Intelligent Technology and Systems;
[1]Beijing National Research Center for Information Science and Technology; [1]Tsinghua University, Beijing 100084, China.
[2]Huawei Technologies Co., Ltd. [3]Netease Fuxi AI Lab.
{j-guan19,fze17}@mails.tsinghua.edu.cn, chenziym4132013@163.com, heruilin@huawei.com,
{maoxiaoxi,fanchangjie}@corp.netease.com, aihuang@tsinghua.edu.cn

1 Data Source

We show the full list of web pages used for constructing the pretraining data of LongLM and the LOT benchmark datasets in Table 1. These web pages provide all kinds of long-form stories varying from classical fairy tales to contemporary novels. All source texts are under licenses that allow use and redistribution for research purposes. We delete all spaces in all texts to normalize our data because it is not needed to use a space to separate two Chinese characters.

Links	Size	Domains
https://www.163.com/	85G	Novels
http://www.txshuku.net/	22G	Novels
http://touch.m.dangdang.com/	7.5G	Novels
https://www.xstt5.com/	6.5G	Novels
http://zuopinj.com/	3.2G	Novels
https://www.kanunu8.com/	2.3G	Novels
https://www.douban.com/	2.0G	Novels
https://www.idejian.com/	1.0G	Novels
http://t.icesmall.cn/	0.7G	Novels
Grand Total	128.3G	Novels
http://www.00txt.com/	83M	Short Stories
https://www.517gj.com/	43M	Children's Stories
http://www.bestgushi.com/	25M	Fable Stories
https://www.etstory.cn/	25M	Fable Stories
https://www.xyyuedu.com/gushifl/	24M	Short Stories
http://www.99gs.net/	21M	Fable Stories
https://www.xigushi.com/	21M	Short Stories
http://www.qbaobei.com/tag/	21M	Children's Stories
https://www.5068.com/kangushi/	21M	Children's Stories
https://www.etgushi.com/	11M	Children's Stories
https://www.hzy365.com/	18M	Fairy Tales
https://www.qigushi.com/	15M	Fable Stories
https://www.gushi365.com/	10M	Fable Stories
http://www.zuowen.com/	6.6M	Fable Stories
http://www.rensheng5.com/	6.5M	Fable Stories
http://www.49j.cn/news/yuyan/	6.3M	Short Stories
Grand Total	134M	Short Stories
	109M	Fable Stories
	96M	Children's Stories
	18M	Fairy Tales

Table 1: Full list of links to the source web pages used for constructing pretraining data of LongLM (Top) and the LOT benchmark datasets (Bottom). **Size** means the corresponding storage space.

2 Annotation Instruction

We show the detailed annotation instructions for story filtering, ClozeT dataset construction and SenPos dataset construction in Figure 1, 2 and 3, respectively. And we also show the annotation instruction for manual evaluation on the PlotCom task in Figure 4.

3 Annotators

We hired a commercial annotation team to annotate the datasets in LOT. The team is led by a professional screenwriter and has taken on hundreds of NLP annotation projects. There were five annotators involved in filtering stories/writing Cloze tests/constructing SenPos examples, respectively. All annotators are native Chinese speakers and well-trained for the annotation tasks. We paid about $0.6 (CNY 4) on average for filtering a story. These stories were then used for constructing the datasets of the ClozeT, PlotCom and OutGen tasks. And we paid $0.8 (CNY 5)/$0.7 (CNY 4.5) to construct each ClozeT/SenPos example, respectively. We decided the payment according to the average text length. Totally, we paid annotators $7.5 (CNY 50) per hour on average, which was far beyond the minimum hourly wage in China (CNY 21)[1]. Finally, we browsed through the annotation results and gave the annotators detailed feedback before approving their submissions. We have clarified these settings in "Section 3: LOT Benchmark" of the revised paper.

4 Error Analysis

We show two ClozeT examples and two SenPos examples that LongLM$_{large}$ does wrong in Table 2 and 3, respectively. And we also shows two PlotCom examples and two OutGen examples where LongLM$_{large}$ generates unreasonable outputs in Table 4 and 5, respectively.

We summarize our observations as follows:

---

[*] Corresponding author

1. Data Description

we ask annotators to construct a story from a story segment. The stories usually have the following characteristics:

1. There is a clear main plot of the story:
   a) The story is told in the form of a coherent event sequence involving several specific and inter-related characters
   b) Events that happen to different characters should be inter-related, not independent.
   c) The event sequence needs to include a complete cause, development and result, and there should be a clear logical relationship between the events. The cause (at the beginning) explains the background of the story or introduces the characters, and at the end explains the ending of these characters after a series of events.
2. The story mainly consists of narrative, without arguments, and may include few descriptive sentences (such as psychological description, environmental description, etc.).
3. The story mainly consists of declarative sentences or a small amount of exclamatory sentences, seldom contains interrogative sentences, rhetorical questions, questioning sentences, etc., and should not include dialogue between characters.
4. The story consists of sentences that are as concise as possible, and does not contain sentences that are irrelevant or a little relevant to the main line.

2. Task Description

The purpose of this task: Given an incomplete story segment of about 50-500 words, and the original whole story corresponding to the segment as a reference, creating a complete and concise story by completing the story beginning and ending of the segment, deleting sentences that have nothing to do with the main line, deleting or rewriting unnecessary questions, dialogues and descriptions. In the rewriting process, annotators should utilize their own commonsense knowledge as much as possible to infer the possible context. The rewriting should be carried out using concise but coherent narrative language. Note that the length of the story after rewriting does not differ from the given story segment by more than 100 words.

Story Segment	Original Story	Reject	Result
Once, the deer and the horse were arguing over a piece of grassland. Everyone wanted to take this grassland as their own. In the end, the deer finally defeated the horse by relying on its powerful horns, this was simply intolerable for horses. Because who can be willing to fail? How could it drive the deer away again?	【Once, the deer and the horse were arguing over a piece of grassland. Everyone wanted to take this grassland as their own. In the end, the deer finally defeated the horse by relying on its powerful horns. This was simply intolerable for horses. Because who can be willing to fail? How could it drive the deer away again?】The horse thought about it; finally thought of asking for help. It found a very strong man. This man came to the meadow, fought with the deer, and killed the deer. Since then, this disputed grassland has been completely owned by the horse. However, the man who helped the horse win already taken the horse as his own. He said, "I have helped you a lot. Therefore, you have to serve me from now on." He saddled the horse on its back, put a bridle on the horse's head, and put a pair of strong tinplate on the horse's mouth. Then, the man said to the horse: "Because I helped you win, so from now on, you have to listen to me honestly." Some people have benefited from the suffering of others and are no longer poor. But for this reason, he lost his precious freedom, which is far more precious than gold.		Once, the deer and the horse were arguing over a piece of grassland. In the end, the deer finally defeated the horse by relying on its powerful horns. This was simply intolerable for horses. Because who can be willing to fail? How could it drive the deer away again? The horse thought about it; finally thought of asking for help. It found a very strong man. This man came to the meadow, fought with the deer, and killed the deer. Since then, this disputed grassland has been completely owned by the horse. However, the man who helped the horse win had already taken the horse as his own. The horse had benefited and was no longer poor, but it had also lost its freedom.
The evil smiled slyly and stepped on the horse. Kindness took the reins and walked aside, still singing while walking. He sang softly, but he didn't expect the evil to have a wicked heart at this time.	(omitted)	1	
After this, every time Zhuge Liang came to the nunnery, he received not only hospitality, but also graciously prepared meals. After dinner, they either laughed or played chess for fun. Compared with Taoist temples, here is really another world. Zhuge Liang was overwhelmed by the woman's sweet words! Zhuge Liang's thinking went wrong, and he became tired of studying.	(omitted)	1	

Note that:

1. When annotating, you first need to determine whether the given story segment can be rewritten as a complete story. If the judgment is no, you need to mark 1 in the 'Reject' column. This story segment should be skipped, do not rewrite. Otherwise, you should rewrite the segment.
2. You should delete grammatical errors, garbled characters, advertisements, and other content unrelated to the story in the given story fragment. And you should ensure that the rewritten story is grammatically correct, the sentences are coherent in terms of causal and temporal relations.
3. The period and comma in the story fragment may be used casually. You should ensure that one event corresponds to one sentence as much as possible, and change the comma in between two events to a period. For example, for the sentence " In the end, the deer finally defeated the horse by relying on its powerful horns, this was simply intolerable for horses." The underlined comma should be replaced with a period. "

3. Annotation Examples

For ease of understanding, in the following examples, we use strikethrough to indicate deleting sentences, use underline to indicate supplementing content, and blue to indicate correcting errors, garbled characters, and character errors.

(1) How to judge whether the given story segment could be rewritten.

If the segment meets one of the following characteristics, you should reject the example.

a. **There is no main story line, it is all about the background and premise of the story.** For example: Once upon a time, Horuna was in the Caliph. During Rashid's administration, two methods were very clever. An unusually weird person, one named Ahmed. Dai Naifu, another named Hasan. Schumann. Both of them are used to design strategies. Using tricks, I have deceived many people, and everyone is afraid of them. Stay away. However, the Caliph had a special liking for them and admired their deception, so he bought them at a high price, let them be their own close guards, and rewarded each of them with a set of expensive clothes. The Caliph appointed Ahmad.

b. **Use metaphors, or use uncommon verbs and adjectives as nouns.** For example: The evil smiled slyly and stepped on the horse. Kindness took the reins and walked aside, still singing while walking. He sang softly, but he didn't expect the evil to have a wicked heart at this time.

c. **It contains too much discussion/explanation/descriptive content.** For example: Second, some styles, such as declarations, and some content, such as mourning the dead, seem unsuitable for writing briskly and lively. Third, solemnity and rigor are not the same thing as dullness and dullness, but they are similar in nature but similar in appearance. Fourth, because it is not easy to draw a line, saying that a certain way of writing is dull and dull, and should be changed, it is difficult to find a reason that everyone can believe. Fifth, assuming that everyone is convinced, that is to say, it is not easy to admit that it is a disease and want to treat it.

d. **It is not standard modern Chinese.** For example: Later, the Mongolian kings fought for the throne, Kublai withdrew his troops in a hurry, Jia Sidao took the opportunity to lie about his achievements in battle, and was praised by Emperor Lizong's imperial edict and was named a young division. Turning into the court, he became more arrogant, suspected of impeaching and undermining the leftist Wu Qian, and dominated the court. After Duzong ascended the throne, he was awarded the title of Taishi, Pingzhang Army State Affairs, and was named Wei Guogong. The other time was in the first year of Emperor Gong Deyou, when Kublai Khan established the Yuan Dynasty in the north and invaded the south again ...

e. **The logic of the segment itself is not reasonable and it is difficult to rewrite.** For example: The little fish who was feeling nothing to do heard Mr. Dolphin's words and happily agreed. Since then, Xiaoyuer has settled down in a marine animal school. He hired squid as a smoke screen teacher, electric eels as a power generation teacher, and swordfish as a speed swimming teacher. He taught the sonar class by himself. Soon after the school was established, little octopus, little electric rays, little sardines, little seals, etc. went to the smoke screen class, power generation class, speed tour class, and sonar class respectively.

f. **The segment contains a lot of information related to unique settings (such as sci-fi, martial arts, fantasy, etc.)** For example: Dou Yu scratched his hair and turned away depressed. The partners looked at each other, not knowing what happened. At this time, the partners all returned to the front yard of the Forbidden Palace. After a brutal battle, the walls of the Forbidden Palace and surrounding buildings were turned into ruined walls. The soldiers were clearing the battlefield in an orderly manner and moving the corpses of the dead black warriors.

(2) Rewrite the story segments. Annotators can refer to the original story, but should describe the story in concise narrative language.

a. **You should change grammatical errors, garbled characters, and character errors (such as parentheses, misalignment of quotation marks, and duplicate symbols) in the story.**
   a) Example 1: Salt and cotton The poor donkey was panting with exhaustion, carrying heavy bags of salt. Suddenly, a small river appeared in front of him ...
   b) Example 2: Happy benchday! -> Happy birthday!
   c) Example 3: They had a great time that day... -> They had a great time that day...

b. **Complete the ending of the story and delete irrelevant content, and make the story reasonable in terms of logic. The ending of the story should resolve the suspense left in the previous plot, or form a complete logical chain with the last event of the story.**
   a) Good Example 1: Once upon a time, there was a very poor old man and old woman. Soon, the old woman passed away. Winter is coming and it is very cold. The old man went to find relatives and friends to help dig a grave. Everyone knew that he was poor and did not agree. The old man had no choice but to bury the old woman alone. (Explanation: "winter is coming..." is a descriptive sentence and has nothing to do with the main plot. The ending can be reasoned based on context and common sense.)
   b) Good Example 2: Once upon a time, there was a very poor old man and old woman. Soon, the old woman passed away. The old man had no money to organize the funeral for the old woman, so he had to ask the priest to help bury the old woman. The priest took a group of people to bury the old woman according to the canon. In order to express his gratitude, the old man invited the priest to eat at home. They sat at the table with wine, vegetables, and various snacks. When the guest priest saw such a good thing, he gulped. After eating and drinking, the priest asked the old man to collect a large sum of funeral expenses. However, the only money left by the old man was used to treat him, and he was now penniless. Therefore, the old man had to go to the priest's house to work as a servant to pay off the debt. (Explanation: Adding "in order to express his gratitude", etc. to improve the story coherence, poor->invite->no money->work to pay off the debt)
   c) Bad Example 1: Once upon a time, there was a very poor old man and old woman. Soon, the old woman passed away. The old man had no money to organize the funeral for the old woman, so he had to ask the priest to help bury the old woman. The priest took a group of people to bury the old woman according to the canon. The old man invited the priest to eat at home. They sat at the table with wine, vegetables, and various snacks. When the guest saw such a good thing, he gulped. When the guests had finished their meal, they left one by one, and the priest stood up. The old man escorted him out of the yard. (Explanation: the ending is weakly related to the context, "poor" has not been embodied in the ending.)
   d) Bad Example 2: Once upon a time, there was a very poor old man and old woman. Soon, the old woman passed away. The old man had no money to organize the funeral for the old woman, so he had to ask the priest to help bury the old woman. The priest took a group of people to bury the old woman according to the canon. The old man invited the priest to eat at home. They sat at the table with wine, vegetables, and various snacks. When the guest saw such a good thing, he gulped. After eating and drinking, the priest asked the old man for a large sum of money for the funeral. The old man turned into a werewolf and ate the priest. (The ending is too much fantasy but this is a realistic story)

Given Reference: Once upon a time, there was a poor old man and an old woman. Soon after, the old woman died. Winter is coming and the weather is very cold. The old man went to his friends and relatives to help dig a grave. Knowing his poverty, he refused. The old man went to see the priest, who was a greedy man. Act without ethics. "Please bury the old woman, agha." The old man said. "Have you any money for the funeral? Tell the truth!" "I dare not lie to you. I don't have any money. Please allow temporarily owe, I am willing to pay high interest, certainly return!" The priest wouldn't listen to the old man. "If you have no money, don't come here!" "There is no way," thought the old man. "I will dig a hole in the grave and bury the old woman myself. He took an axe and a spade to the grave and began to dig. He dug, he found an iron pot, shining with gold coins. The old man was overjoyed. "Thank God! Now I can bury my wife." He gave up digging and carried the tin can home. Money makes the world go round. He hired some men, dug the grave, made the coffin. He sent for wine and food, and for the comprador's funeral, and took the gold pieces himself to the priest. As soon as he entered, the priest said to him, "I tell you plainly, old man, don't come here without money!" "Don't be angry, agha sahib." "Said the old man cheerfully. "I'll give you gold coins to help me bury my wife. I'll never forget your kindness." The priest took the money and didn't know how to receive the old man or where to sit. His mood changed. "Don't worry, old man. Everything will be done." ...

   e) Good Example 3: A fox had its tail pinched by a trap. After receiving this insult, he felt that life was very difficult, and he decided to persuade other foxes to remove their tails. Everyone is the same, and his shortcomings can be concealed. He summoned all the foxes and advised them to cut off their tails. He said that this thing is both unsightly and cumbersome. But the other fox said they needed their tails to keep their balance when they ran. So they saw through the fox's trick and kept their tails. (Explanation: Foxes need tails->the fox is bad)
   f) Bad Example 3: A fox had its tail pinched by a trap. After receiving this insult, he felt that life was very difficult, and he decided to persuade other foxes to remove their tails. Everyone is the same, and his shortcomings can be concealed. He summoned all the foxes and advised them to cut off their tails. He said that this thing is both unsightly and cumbersome. Other foxes cut off their tails, and the wolf came and ate all the foxes. (Explanation: the wolf has not appeared in the context and the ending is very rigid)
   g) Good Example 4: A little rabbit lives in the forest. The little rabbit grows a lot of carrots, and the growth is gratifying. It seems that this year's carrots will definitely have a bumper harvest. The little bunny felt happy in his heart. In the night of this day, the weather was sultry, and the little rabbit was enjoying the coolness on the edge of the carrot field, and suddenly heard movement in the carrot field. Taking a closer look in the moonlight, it was his neighbor's voles family, or to be precise, the voles mother and its two children. It turned out that the voles family was stealing carrots from the little rabbits. The little rabbit was very angry and beat the voles family away.
   h) Bad Example 4: A little rabbit lives in the forest. The little rabbit grows a lot of carrots, and the growth is gratifying. It seems that this year's carrots will definitely have a bumper harvest. The little bunny felt happy in his heart. In the night of this day, the weather was sultry, and the little rabbit was enjoying the coolness on the edge of the carrot field, and suddenly heard movement in the carrot field. Taking a closer look in the moonlight, it was his neighbor's voles family, or to be precise, the voles mother and its two children. The voles' family ate all the carrots. (The ending of the rabbit is given)
   i) Good Example 5: This day is the little brown bear's birthday party. However, the little brown bear did not invite him, so why was he embarrassed to go? The little squirrel girl received an invitation letter from the little brown bear, and she was very happy to attend the little brown bear's birthday party. She was carrying a small basket full of red cherries. It was delicious! She is going to give these cherries to the little brown bear as a birthday present. The Fat Little Pig was leaning at the door of his home, looking at the little animals going to the banquet. He wanted to go, with a pair of small hands behind his back, holding a birthday card. The fat pig thought, if anyone could give me this card to the little brown bear! He stopped the little squirrel girl and asked her to give the greeting card to the little brown bear. After receiving the greeting card, the little brown bear regretted forgetting to invite the fat pig.
   j) Bad Example 5: This day is the little brown bear's birthday party. However, the little brown bear did not invite him, so why was he embarrassed to go? The little squirrel girl received an invitation letter from the little brown bear, and she was very happy to attend the little brown bear's birthday party. She was carrying a small basket full of red cherries. It was delicious! She is going to give these cherries to the little brown bear as a birthday present. The Fat Little Pig was leaning at the door of his home, looking at the little animals going to the banquet. He wanted to go, with a pair of small hands behind his back, holding a birthday card. The fat pig thought, if anyone could give me this card to the little brown bear! As he thought about it, he got angrier and angrier and ate the card. (The ending is not accordant with the pig's characteristic. It is not commonsense that the pig ate the card)

c. **Rewrite the story plot if the logic is not unreasonable.**
   a) Example 1: A poor old lady lives in a village. She picks a plate of beans and prepares to cook them. She lit the fire in the stove, and in order to make the stove burn faster, she used a handful of wheatgrass when making the stove. When pouring the beans coal into the boiler, she did not notice that a bean a piece of burning coal fell out and landed next to a wheat straw on the ground. The coal ignited the wheat straw and burned down the old lady's house.
   b) Example 2: Once upon a time, there was a crane with a kind heart. He became friends with a small fish. He often rescued the small fish when he was caught by other cranes. Once another crane ate the little fish while he was not paying attention. The crane that hadn't eaten the fish came over. He found the fish bones on the grass and cried sadly. The crane cried out in surprise. The crane that dared to eat fish hurriedly took a photo of his head against the river. The other cranes were the same, and they didn't get the beautiful red hat. Only the crane that released the small fishes gained beauty, so this crane later became the red-crowned crane it is today. The kind-hearted red-crowned crane buried the bones of the little fish sadly, and stood by the river every day and sang sadly to commemorate his dead friend! (It is fantastic that "a crane becomes a red-crowned crane" and the plot needs to be rewritten)

4. Other Requirements:

1. The stories should not contain dialogues. If there is a direct quotation in the story, it should be changed to an indirect quotation. And the available punctuation markers contain only the question mark, period, comma, exclamation mark, ellipsis, semicolon and guillemet.
For Example: The old man was overjoyed. "Thank God! Now I have money to bury his wife." He thanked God for letting him have money to bury his wife.
2. Deleting the unrelated content, descriptive and argumentative sentences.
   a) The air humidity at the foot of the mountain is getting higher and higher, and the mountain view is gradually hazy in the fog. With gusts of mountain wind blowing, the night sky was clear, and a white and transparent crescent moon hung in the sky. At dawn, the black knight's body temperature dropped. However, if the wound on his leg continues to become inflamed, then he will still have a fever...
   b) A group of monkeys live on a mountain by the river. This mountain is full of waterfalls and springs, and the scenery is very beautiful. Every spring, wild fruits grow all over the mountains and plains. It is not clear what year, a group of monkeys came to this mountain to settle down. Since then, they have been living a leisurely and contented life. One day, King Wu took his entourage on a boat on the river. When he found this scenic Monkey Mountain among the strange peaks on both sides of the river, he was very excited... King Wu was a little annoyed and ordered his entourage to join him. Go chasing the monkey. Faced with the arrows shot by so many people, the monkey couldn't resist, and was shot to death by random arrows. A wise person knows to hide but not to reveal its strength, and likes to show off with a little skill is stupid. They are either self-defeating and being laughed at, or they end up in failure.
   c) One day, Ruan Xiaotian was walking on the way to school when he suddenly found a car accident on the side of the road. Ruan Xiaotian ran over. He had learned cardiopulmonary resuscitation and might be able to help. The elderly Grandma Li fainted on the ground with a pale face, and the vegetable basket was lying on her lap. The owner of the sports car was still arguing with passers-by. The driving car is a beautiful woman's car. The woman is full of celebrity style. Ruan Xiaotian knelt and felt Grandma Li's breathing with his fingers. Breath is very weak. He stretched out his hand to press Grandma Li's pulse, shocked in his heart. Just when his finger touched Grandma Li, the physical condition of Grandma Li appeared in his mind. Grandma Li had a slight fracture in her left leg. The most serious problem was her heart. She originally had a heart attack, and her heart's blood vessels were narrowed. Because of the fright, her heart muscle contracted sharply, blood supply was insufficient, and a heart attack occurred. If it is not rescued in time, Grandma Li is likely to die. Ruan Xiaotian wants to help Grandma Li! A force was injected into Grandma Li's body from Ruan Xiaotian's fingertips. He immediately began CPR on Grandma Li. Ruan Xiaotian was surprised to find that Grandma Li's left leg and her heart were undergoing good changes. Grandma Li groaned and slowly opened her eyes. Ruan Xiaotian breathed a sigh of relief, and Grandma Li also expressed her sincere thanks to Ruan Xiaotian.

Figure 1: Annotation instruction for story filtering.

- In the first ClozeT example, LoneLM fails to reason the intention of the "cat". And in the second ClozeT example, LongLM fails to reason the consequence based on the intention of "Mercury" and the reaction of "Juno".

- In the first SenPos example, LongLM mistakes the implicit causality between the removed sentence and the context. In the second SenPos example, LongLM fails to capture the inter-sentence topical relatedness around "question".

- In the first PlotCom example, LongLM generates an incoherent sentence with conflicting logic despite good fluency. In the second PlotCom example, LongLM generates a fluent and coherent sentence in isolation, but the sentence has a bad logical connection with the context.

1. Task Description

The purpose of this task is: given a story of about 50~500 words and a sentence marked in it, rewrite this sentence into another sentence that is relevant to the context but violates common sense with minimal changes. The marked sentence may appear at the end or middle of the story, and we have different requirements for rewriting sentences in different positions:

- For the sentence at the end: If it is judged that it cannot be reasoned based on common sense, it should be marked 1 at the "REJECT" column.
- For the sentence at the middle: If it is judged that it cannot be reasoned based on common sense, the annotator should find another sentence except the beginning and end that can be reasoned based on common sense and rewrite this sentence. If there are no such sentences, the annotator should mark 1 at the "REJECT" column.

Given Story	REJECT	Original Sentence	Rewritten Sentence	Policy
A goblin had buried a treasure under the ground. After that, he received a long flight mission from the Devil King. The goblin began to worry about how to guard the treasure during his mission. The goblin thought for a long time and decided to give the treasure to a miser. The miser clung to his vault even when he was asleep, so the goblin trusted him very much. Ten years later, the goblin returned from his mission and found that the miser had starved to death, but he still held his vault to death and completed the goblin's entrust.		The goblin thought for a long time and decided to give the treasure to a miser.	The goblin thought for a long time and decided to ask a miser to give the treasure to the Devil King.	1
Once upon a time, there was a very poor old man and old woman. Soon, the old woman passed away. The old man had no money to organize the funeral for the old woman, so he had to ask the priest to help bury the old woman. The priest took a group of people to bury the old woman according to the canon. In order to express his gratitude, the old man invited the priest to eat at home. They sat at the table with wine, vegetables, and various snacks. When the priest saw such a good thing, he gulped. After eating and drinking, the priest asked the old man to collect a large sum of funeral expenses. However, the only money left by the old man was used to treat him, and he was now penniless. Therefore, the old man had to go to the priest's house to work as a servant to pay off the debt.		Therefore, the old man had to go to the priest's house to work as a servant to pay off the debt.	Therefore, the old man had to return the money that the priest gave him to bury the old woman to the priest.	1
The little snail Xinxin went to grandma's house to play, but grandma was ill. Xinxin went out to find a doctor immediately. But Xinxin crawled slowly, it was very anxious. It started crying as it crawled. A swan named Feifei saw it, and after asking the reason, he reassured Xinxin that he would go to help him find a doctor. Feifei flew fast, and after a while, the doctor arrived. Grandma gradually recovered under the treatment of the doctor. Xinxin was very grateful to Feifei, and they became good friends ever since.		Feifei flew fast, and after a while, the doctor arrived.	Feifei was slower than Xinxin, and it took a long time to find a doctor.	2
27-year-old Chemetir is a peasant woman in Kenya. After a year of hard training, she participated in the domestic competition and won the seventh place. After that, under the guidance of professional coaches, she made great progress. In the international marathon she participated in, she was the first to cross the finish line. Her four children also got a good education with this bonus.	1 ("four children" can not be reasoned)			

2. Detailed instruction

(1) How to judge whether a sentence can be reasoned based on common sense. A qualified sentence tends to:

- Describe characters' reaction or intention such as "feeling upset after breaking up", "working hard to make money", "living in a hotel when traveling", or narrate an event to introduce characters' reaction or intention.
- Embody the nature of physical objects and concepts such as places or usage. Physical objects refer to general concepts, proper nouns or events, such as "sparrow", "computer", "prince", "Google", "New York", "playing soccer", "running". In contrast, "John" and "method of playing Poker" are not physical objects. Examples for "nature": "university used for education", "scoop used for eating", "cats eat fishes", "little animals are afraid of tigers" etc.

Good Examples	Explanation
Once upon a time, there was a crane with a kind heart. He became friends with a small fish. He often rescued the small fish when he was caught by other cranes. Once another crane ate the little fish while he was not paying attention. The crane that hadn't eaten the fish came over. He found the fish bones on the grass and cried sadly. The crane buried the bones of the little fish, and stood by the river every day and sang sadly to commemorate his dead friend!	We can reason that the reaction of the crane is "bury the fish and commemorate his friend" based on the context "the fish was eaten".
Once, the deer and the horse were arguing over a piece of grassland. Everyone wanted to take this grassland as their own. In the end, the deer finally defeated the horse by relying on its powerful horns. This was simply intolerable for horses. How could it drive the deer away again? The horse thought about it; finally thought of asking for help. It found a very strong man. This man came to the meadow, fought with the deer, and killed the deer. Since then, this disputed grassland has been completely owned by the horse. However, the man who helped the horse had already taken the horse as his own. The horse had benefited and was no longer poor, but it had also lost its freedom.	We can reason the horse lost in this fight based on the context "deer and horse argue" and "it was intolerable for the horse".
The little snail Xinxin went to grandma's house to play, but grandma was ill. Xinxin went out to find a doctor immediately. But Xinxin crawled slowly, it was very anxious. It started crying as it crawled. A swan named Feifei saw it, and after asking the reason, he reassured Xinxin that he would go to help him find a doctor. Feifei flew fast, and after a while, the doctor arrived. Grandma gradually recovered under the treatment of the doctor. Xinxin was very grateful to Feifei, and they became good friends ever since.	We can reason that "Feifei found the doctor quickly" based on the common sense "swans are quicker than snails".

Bad Examples (the sentences usually contain some unique story settings, such as features of roles)	Explanation
In 1839 of the sheep calendar, Yangyang Xi and Tailang Hui died together in the bomb-throwing battle. When the sheep were commemorating him with the Xi statue, they received a war note from the Tailang Hong. Hong wants to seek justice for Hui and once again compete with the sheep. Hong has to compete with Yangyang Mei. When the memorial statue of Xi was completed, the memorial statue of Hui was also stationed in the Hui Memorial. Hong and Mei expressed their respective grief for the deceased toward the statue. Hong eventually lost the game because of lack of food all year round.	The story setting is very unique, and the logical connection of the underlined sentence is very weak. The annotators should mark 1 at the "Reject" column.
A goblin had buried a treasure under the ground. After that, he received a long flight mission from the Devil King. The goblin began to worry about how to guard the treasure during his mission. The goblin thought for a long time and decided to give the treasure to a miser. The miser clung to his vault even when he was asleep, so the goblin trusted him very much. Ten years later, the goblin returned from his mission and found that the miser had starved to death, but he still held his vault to death and completed the goblin's entrust.	The story setting ("goblin", "Devil King") is unique. "long flight mission" is difficult to reason based on the context and common sense.
There is a scholar who always talks endlessly. Once, a friend couldn't bear it and interrupted him. He was very angry and ridiculed the friend who had worked in the mill. This friend was also very angry, and said that this scholar was the donkey back then. Because they were too mean to speak, no one around them wanted to help them, and they all left silently.	"had worked in the mill" can not be reasoned based on the context and common sense.

(2) Sentence Rewriting:

The annotators should rewrite the marked sentence to another that is relevant to the context but violates common sense. Generally we recommend two rewriting policy.

2.1 Rewriting policy

a. Policy 1: make characters have unreasonable reaction or intention

Given Story	Rewritten Sentence	Rewriting Policy
Once, the deer and the horse were arguing over a piece of grassland. Everyone wanted to take this grassland as their own. In the end, the deer finally defeated the horse by relying on its powerful horns. This was simply intolerable for horses. How could it drive the deer away again? The horse thought about it; finally thought of asking for help. It found a very strong man. This man came to the meadow, fought with the deer, and killed the deer. Since then, this disputed grassland has been completely owned by the horse. However, the man who helped the horse had already taken the horse as his own. The horse had benefited and was no longer poor, but it had also lost its freedom.	In the end, the horse finally defeated the deer by relying on its powerful horns.	1
A silly wolf and a fox stole a jar of honey and then hid it in a tree hole. They agreed that neither of them were allowed to eat the honey alone. However, the fox sneaked back to eat up all the honey the next day. Afterwards, whenever the wolf asked the fox to eat the honey together, the fox always refused its request. Finally the wolf could not help coming back to the tree hole and found that the jar had been empty. The wolf felt very regretful that the honey became dry because it had been too long. It had no doubts about the fox at all. After hearing this, the fox became more active to look for food together with the wolf.	When hearing this, the fox became very angry and decided no longer to look for food together with the wolf.	1

b. Policy 2: make physical objects violate their nature. For comparison, policy 1 pays more attention to the causal and temporal relations, while policy 2 focuses more on the association of objective things in terms of their attributes.

Given Story	Rewritten Sentence	Rewriting Policy
The little snail Xinxin went to grandma's house to play, but grandma was ill. Xinxin went out to find a doctor immediately. But Xinxin crawled slowly, it was very anxious. It started crying as it crawled. A swan named Feifei saw it, and after asking the reason, he reassured Xinxin that he would go to help him find a doctor. Feifei flew fast, and after a while, the doctor arrived. Grandma gradually recovered under the treatment of the doctor. Xinxin was very grateful to Feifei, and they became good friends ever since.	Feifei was slower than Xinxin, and it took a long time to find a doctor.	2
Once upon a time, there lived a mother and her son at the foot of a mountain. After her son grew up, he went out to learn skills and never came back. Therefore, the mother went to the nearby city to look for him. However, her son became an official and disowned his mother. The mother sat by the roadside and cried sadly. A young man passed by and knew the cause. Then the man took her home. Actually, the man was also the prince of the city.	Actually the man was also an official	2
A goblin had buried a treasure under the ground. After that, he received a long flight mission from the Devil King. The goblin began to worry about how to guard the treasure during his mission. The goblin thought for a long time and decided to give the treasure to a miser. The miser clung to his vault even when he was asleep, so the goblin trusted him very much. Ten years later, the goblin returned from his mission and found that the miser had starved to death, but he still held his vault to death and completed the goblin's entrust.	Unexpectedly, he quickly squandered his money, leaving only an empty treasury.	2

2.2 Restriction

a. All the things, roles, and places that appear in the rewritten sentence should appear in the context. If the original sentence introduces new things, roles, and places that are not mentioned above, the rewritten sentence should also introduce these things, roles, and places.

Given Story	Good Example	Bad Example	Explanation
Once, the deer and the horse were arguing over a piece of grassland. Everyone wanted to take this grassland as their own. In the end, the deer finally defeated the horse by relying on its powerful horns. This was simply intolerable for horses. How could it drive the deer away again? The horse thought about it; finally thought of asking for help. It found a very strong man. This man came to the meadow, fought with the deer, and killed the deer. Since then, this disputed grassland has been completely owned by the horse. However, the man who helped the horse win had already taken the horse as his own. The horse had benefited and was no longer poor, but it had also lost its freedom.	In the end, the horse finally defeated the deer by relying on its powerful front hooves.	In the end, the horse finally asked its good friend the donkey to defeat the deer together.	Both "horse" and "deer" appear in the context and the original sentence. But "donkey" in the bad example does not appear in the context.
A goblin had buried a treasure under the ground. After that, he received a long flight mission from the Devil King. The goblin began to worry about how to guard the treasure during his mission. The goblin thought for a long time and decided to give the treasure to a miser. The miser clung to his vault even when he was asleep, so the goblin trusted him very much. Ten years later, the goblin returned from his mission and found that the miser had starved to death, but he still held his vault to death and completed the goblin's entrust.	The goblin thought for a long time and decided to ask a miser to give the treasure to the Devil King.	The goblin thought for a long time and decided to give the treasure to the Devil King.	All three roles "goblin", "Devil King" and "miser" appear in the original sentence and good example. The bad example does not introduce "miser".
There is a scholar who always talks endlessly. Once, a friend couldn't bear it and interrupted him. He was very angry and ridiculed the friend who had worked in the mill. This friend was also very angry, and said that this scholar was the donkey back then. Because they were too mean to speak, no one around them wanted to help them, and they all left silently.	They are very happy, believing that each other is the person who knows them best, and they have become best friends.	This scholar is very happy and complacent about his ingenuity.	The original sentence describes the endings of both "scholar" and "his friend". But the bad example does not describe the ending of "his friend".

b. The rewritten sentence should be coherent and fluent within itself, and it should be similar to the original sentence in terms of language style, narrative tone, and sentence pattern (declarative, interrogative, imperative).

Given Story	Good Example	Bad Example	Explanation
Once upon a time, there was a very poor old man and old woman. Soon, the old woman passed away. The old man had no money to organize the funeral for the old woman, so he had to ask the priest to help bury the old woman. The priest took a group of people to bury the old woman according to the canon. In order to express his gratitude, the old man invited the priest to eat at home. They sat at the table with wine, vegetables, and various snacks. When the priest saw such a good thing, he gulped. After eating and drinking, the priest asked the old man to collect a large sum of funeral expenses. However, the only money left by the old man was used to treat him, and he was now penniless. Therefore, the old man had to go to the priest's house to work as a servant to pay off the debt.	The old woman was not frugal, and soon ran out of all money.	1. The old woman turned into a werewolf and fled. 2. The old woman was very generous and never gave the old man a cent.	The first bad example violates the realistic style of the original sentence. The second bad example is logically unreasonable.
Once, the deer and the horse were arguing over a piece of grassland. Everyone wanted to take this grassland as their own. In the end, the deer finally defeated the horse by relying on its powerful horns. This was simply intolerable for horses. How could it drive the deer away again? The horse thought about it; finally thought of asking for help. It found a very strong man. This man came to the meadow, fought with the deer, and killed the deer. Since then, this disputed grassland has been completely owned by the horse. However, the man who helped the horse win had already taken the horse as his own. The horse had benefited and was no longer poor, but it had also lost its freedom.	In the end, the horse finally defeated the deer by relying on its powerful front hooves.	In the end, the horse finally defeated the deer by relying on its powerful horns.	The bad example violates common sense within itself since horses do not have horns.

c. The rewritten sentence should be similar to the original sentence in terms of length (1/2~2×original length).

Given Story	Good Example	Bad Example	Explanation
A group of senior students from the Department of Mechanical Engineering who are about to graduate are taking the final exam. Because this exam allows consulting books, the students are full of confidence. The test paper was issued, and there were only 5 questions. But after three hours, everyone was frowning when the exam is over. The teacher laughs and tells the students that the content of this test is a problem they will often encounter in the future. Their study at school is over, but their study in work has just begun. Even if failing the test this time, don't deny yourself.	The students handle the exam easily. After 3 hours, everyone handed in the test paper.	The exam is over quickly.	The bad example is too short.

d. Try to rewrite the original sentence not by directly adding negative words or replacing antonyms.

Given Story	Good Example	Bad Example	Explanation
A long time ago, a prince wanted to marry a real princess. He traveled all over the world to find her. Although there are many princesses in the world, but he is not completely sure whether they are real princesses. So he went back home again annoyed. One night, the lights were whining, the thunder rumbling, and the pouring rain poured down. At this time someone slapped the gate of the city, and the old king ordered someone to open it. Standing outside was a princess. She told them that she was a real princess. In order to determine whether she was a real princess, the prince put a pea on the princess's bed, and then put five layers of mattresses on the bed. The next day the princess said that she was uncomfortable and felt something on the bed. The prince felt that only the princess had such delicate skin, so he proposed to the princess. Finally, the prince and princess lived happily together.	Finally, the old king and princess lived happily together.	Finally, the prince and princess lived unhappily together.	The bad example replaces "happily" to "unhappily", which does not meet our requirement.

e. Try to use as rich a vocabulary as possible considering the context, not reuse the same scenes/actions when annotating multiple examples, and not use generic scenes and events such as "eat", "have fun", "don't know".

Given Story	Good Example	Bad Example	Explanation
There is a scholar who always talks endlessly. Once, a friend couldn't bear it and interrupted him. He was very angry and ridiculed the friend who had worked in the mill. This friend was also very angry, and said that this scholar was the donkey back then. Because they were too mean to speak, no one around them wanted to help them, and they all left silently.	They are very happy, believing that each other is the person who knows them best, and they have become best friends.	The scholar was happy and they went to eat together.	"happy" and "eat" in the bad example is generic without considering the context information.

f. The rewritten sentence should not violate the temporal relations with the context.

Given Story	Good Example	Bad Example	Explanation
When Chao Huang rebelled, he met a woman on the road. The woman lost her relatives due to the war. While fleeing, she carried her nephew on her back and held her younger son in her hands. Huang saw and asked why she carried the old and good children on her back, while the young and weak ones were dragged behind. The woman didn't know that he was Huang, and said that she was carrying her elder brother's orphan, and she was afraid that Huang would break her family's blood. And his son walked behind because in case he was killed, he could regenerate. Huang sympathized with her very much. He told the woman to hang kudzu vines on the door to take refuge. After Huang returned, he ordered that the people with kudzu vines hanging on the door were not to be killed. The woman was in peace since then.	The woman has lived a life of fear and displacement since then.	But the woman completely ignored Huang's words. When she was angry, she struck a pole on Huang's head and killed him.	The context has said that Huang had went back, which is temporally conflicting with "the woman killed Huang" in the bad examples.

Figure 2: Annotation instruction for constructing the ClozeT dataset.

- The two OutGen examples show that LongLM can generate fluent stories. However, LongLM tends to reuses the phrases in the given outline several times and can not organize the outline as a coherent event sequence.

These examples further demonstrate the challenges of our benchmark.

1. Task Description

We extracted a sentence from a narrative story and marked all the possible positions where the sentence can be inserted. The annotators are supposed to do as follows:

(1) Considering the sentence and story context, and select a proper position where the sentence can be inserted

(2) If there are more than one proper positions for the sentence, annotators should list all the position indexes from small to large.

Apart from this, we expect to gather samples satisfying:

(1) The candidate sentence is a dependent, complete narrative sentence, whose meaning can be understood alone.

(2) The story itself should embody consistent characters, especially the sense of logic and hierarchy, and it can be logically reasoned.

(3) The story context should be free of typos, garbled characters, grammar errors, and clauses of text are reasonable

For samples not satisfying the above criterion, annotators should reject the sample, or make revisions on it to make it satisfied and annotate it.

Here is an annotation example.

sentence	story	Reject	Revise	Revised Sentence	Revised Story	Position
He said the food was too hot and let it cool down.	[0] According to legend, there used to be a person who made friends with Satyrs, a god of the forest. [1] When winter came, the weather became very cold. The man placed his hands near his mouth and warmed them up by exhaling warm air. [2] The god of the forest hurriedly asked the reason why he did it. [3] The man replied that it was cold and so were his hands. His hands would be warmer if he blow warm air on them. [4] Afterwards, they ate at the same table. The food on the table was steaming hot. The man picked up a little bit and put it to his mouth. [5] The god of the forest asked him why again. [6] The god of the forest said to man, ' hey, friend. [7] I have no choice but to break up with you, since your mouth can blow out the heat for a while, and blow out cold again. [8] The story tells that we must never make friends with those who are capricious. [9]	0	0			6

2. How to judge whether to reject or accept the sample

Among samples given, some might not satisfy the basic requirements for this task. Therefore, annotators should reject these samples. We hope the annotators can focus on the discourse relations between the sentence and the story, and reject those samples where there are logic confusions or lacks in logical connections. The leftovers are samples of clear and proper logic.

1. How to reject

When rejecting a sample, annotators are supposed to fill in the blank under "Reject" in the sheet with "1". And the other blanks in the sheet of the sample are not obliged to be filled. Annotators can proceed to the next sample.

2. Situations when rejecting a sample

2.1 The Candidate sentence's meaning is not clear, or too general to fit in the task. OR the candidate sentence is descriptive or argumentative, lacking temporally sequential logic.

2.2 The candidate sentence shall be like a independent and complete sentence, whose meaning can be almost understood solely. The sentence is not ambiguous and it does not cause ambiguity in reasoning.

Bad Example (Candidate Sentence)	Explanation
If not doing so, these five people would stay alive in their home to a very old age, and everybody is able to drive them as slaves. How can they have the celebrities who want to express lament towards the great ambition pay respect and bow in front of their tombstones?	There is an obscure close: "stay at home and doing nothing means to be driven as slaves", which we do not expect the candidate sentence owns. Reversely, the candidate sentence should be clear and easy to understood considering the words and story context.
Yuan Zhou smiled bitterly and said.	This sentence is a functional sentence in the story, which leads a direct speech. We do not hope such functional sentences to be candidates
This Scottish glanced at him sharply.	This is description, which do not have temporal or sequential logic in story. Please reject samples with such candidates.

2.3 The whole story is full of non-narrative (descriptive, argumentative) texts whose temporal and sequential logic is not clear. Specifically, stories mainly aim to narrate a event sequence. Sometimes there would be description or discussion in it, which is allowed in our task. However, large portion of description or discussion would make the text not a story. Please reject such cases.

Bad Example (Text)	Explanation
Instead of honoring this sacred obligation, America has given the Negro people a bad check, a check which has come back marked "insufficient funds." But we refuse to believe that the bank of justice is bankrupt. We refuse to believe that there are insufficient funds in the great vaults of opportunity of this nation. And so, we've come to cash this check, a check that will give us upon demand the riches of freedom and the security of justice.	This is part of speech I have a dream by Jr Martin Luther King. This is a discussive text with clear logic and expression. But this is not a story, which does not fit in our task.

2.4 Although the text is a narrative, but it lacks logic. Story should be organized based on causal and temporal logic. In contrast, narrative essays, dialogue, speeches, debates lack temporal relationships and thus do not fit our task. Please reject these samples.

Bad Example (Text)	Explanation
[0] A few minutes ago, he was burned by the fire of jealousy. [1] He wants to doing anything for her, but he doesn't know what to do. [2] That was the most unbearable because he didn't know how to help her. [3] He is the only one who can help her. It is him, only him. [4] Because she has no one else in the world, no one anywhere in the world. [5] Tears had wet his cheeks, he stood up, walked a few steps towards the villa, and called her name loudly. [6]	The topic of this example is describing personal feelings and experiences. The points involved are highly divergent and lack internal connections. It is a type of narrative essay and should be rejected.
[0] Above the twists and turns of the lotus pond, the leaves of the fields can be seen. [1] The leaves are a high above water, as if the skirt of a gentle dancing girl. [2] In the midst of the layers of leaves, there are scattered white flowers in scattered spots, some blooming curly, and some shyly hitting flowers. They are just like the pearls, like the stars in the blue sky, and like the beauty just out of the bath. [3] When the breeze is passing, wisps of fragrance are sent, like the faint singing of a tall building in the distance. [4] At this time, the leaves and flowers also carry a trace of lightning, passing across the lotus pond in an instant. [5] The leaves were close to each other side by side, and there seemed to be a condensed blue wave. [6] Under the leaves is the silent running water, which is covered, and its color cannot be seen. But the leaves are more beautiful. [7]	The description of the moonlight in the lotus pond adopts prose style. Although the writing is based on certain points, there is no logical connection between these points, and the context of the entire paragraph is not closely integrated (shown in that some sentences can be switched in order, such as "a breeze is passing" and "leaves and flowers at this time").

2.5 The story context is not correlated, or is logically confusing. Specifically, we hope that the context in the story we accept has strong narrative coherence, that is, there is a strong logical relationship between events and the writing sequence is reasonable.

Bad Example (Text)	Explanation
And yet that mark on the wall is not a hole at all. It may even be caused by some round black substance, such as a small rose leaf, left over from the summer, and I, not being a very vigilant housekeeper – look at the dust on the mantelpiece, for example, the dust which, so they say, buried Troy three times over, only fragments of pots utterly refusing annihilation, as one can believe.	This is an excerpt from the stream of consciousness novel "The Spot on the Wall". There is a certain logical causal connection between the two sentences in the story, and there is also a certain narrative order in the whole. However, from the perspective of the entire passage, the overall logic of this segment is not strong. The content at the end of the paragraph and the beginning of the paragraph has deviated, and the logical connection is puzzling. It is not the story with clear logical lines and a clear sense of organization we want.
[0] Chongyun talks sadly with his junior and sister about a person whom each other likes in the dark twilight. [2] After a while, Chongyun said with a displeased face: "Do you remember that spring." [3] Then he turned and left. [4] In the spring night with the fragrance of flowers in the air, bid farewell to the gentle girl sadly. [5] Then I walked over to arrange his clothes for him, and suddenly saw the hideous scars on his body. [6]	The text is incoherent. The sentences are blunt and not smooth, and the context is not closely connected, so it is rejected.

2.6 If the story contains dialogues in it and the dialogue is related to the story but lack narrative logic. Specifically, we allow partial dialogue in the content of the story, but if the portion of the dialogue is too large, the entire paragraph will turn from a story to a dialogue. Please reject all dialogue-style texts for two-person dialogues. In addition, if the content of the large-length dialogue lacks narrative, it is not a description of the event, but a question and answer, please also decline.

Bad Example (Text)	Explanation
[0] The matchmaker praises the girl. [1] The young man said: "I've seen this girl, as if she is blind in one of her eyes." [2] "That's great, other men won't scoff with her anymore." [3] "I heard She is a stutter." [4] "That's great, she won't babble, talk too much." [5] "Is it short, right?" [6] "Short and save clothes." [7] "Yes. Everyone said she was a little silly." [8]	In this example, it is mainly the dialogue between the two. The content of the dialogue is too much, which makes the structure of the story deformed, so it is rejected.
[0]On a grassland, the lion said to a snail: "Little snail, how about learning to hunt with me?" [1] The snail asked: "Can you fly?" [2] The lion shook his head and said, "No, I can't." [3] The snail said: "If you can't fly, you can't be my master." [4] After speaking, he ignored the lion and crawled forward on his own. [5] A yellow oriole flew over: "Snail, snail, how about learning to sing with me?" [6] The snail asked: "Do you know how to swim?" [7] The oriole shook his head and said, "No, I can't" [8]]The snail said: "You are not all-rounder, either." [9] After speaking, he crawled forward. [10] The snail was very tired and rested under a big tree. [11] He said to himself: "Why are there no all-rounders in the world?" [12] The tree heard and said: "Haha, how can there be any all-rounders in the world! If you had learned these skills one by one gradually, you would have become successful!" [13] The snail lowered his head in shame after hearing what the tree said. [14]	Although this example requires a certain amount of dialogue content, the dialogue is not long, and the content of the dialogue has a certain philosophical meaning, which can promote the development of narrative. Also, it is related to the story, and has a certain influence on the structure of the story. So it is accepted.

2.7 This annotation task mainly focuses on narrative stories in modern Chinese. Classical Chinese, English, and stories with a lot of unfamiliar online vocabulary should be rejected.

Bad Example (Text)	Explanation
[0] Many emperors of the Western Han Dynasty established their own temples during their lifetimes. [1] The book of "Han Shu", four years after Emperor Wen, made Gu Cheng Temple. [2] Note: The emperor is himself a temple, and the system is narrow, if it can be expected to become one. [3] Jingdi Temple is called Deyang, Wudi Temple is called Longyuan, Zhaodi Temple is called Wandering, Xuandi Temple is called Leyou, Yuandi Temple is called Longevity, Chengdi Temple is called Yangchi. [4] See all notes on "Han Shu". [5]	This is a Classical Chinese story. Please reject.
[0] First double change and then ask for a monthly pass, in case I fail to say something, I dare not make a promise, after all, I haven't been in a state and time is not too much recently. Well, this is already Xiaoqi's greatest sincerity. [1] In any case, the double monthly pass time is the decisive battlefield. In the last 24 hours, everyone should work harder and give the small July monthly pass as much as possible. Can you update your motivation this month? It depends on the day after that. The results are up, Min Na Sang, Gan Ba Daddy, oh. [2]	There are lots of uncommon internet slang. Please reject.

2.8 There are more than one appropriate position in the context for candidate. Please reject such samples.

Sentence	Text	Explanation
But the farmer smiled and said that it was okay since it would not be fruitful.	[0] Once upon a time there was a farmer who argued with the Old Man of Cold one day. [1] The Old Man of Cold threatened the farmer and said that he would freeze the farmer to death. [2] They decided to resolve the dispute all day and night. [3] The farmer was dressed warmly, standing in the severe cold. [4] The Old Man of Cold worked very hard, and the farmer's feet were completely frozen. [5] The farmer felt that he couldn't stand it anymore, took off his boots and stood barefoot on the ground. [6] The Old Man of Cold was surprised and asked why you took off your boots. [7] The farmer replied that my feet got hot and sweated a lot. [8] The Old Man of Cold concluded that he had failed, so he stopped letting the wise farmer suffer from the cold. [9]	This sample contains several appropriate positions (3, 4, 5, 8). Please reject such samples.

3. How to judge whether to revise the sample

3.1 The goal of revision

There are some texts that do not meet the above rejection conditions, but there are still some problems with the qualified samples. In order to improve the sample quality and simplify the labeling process, the annotator can modify the original text in the judgment process to make it a reasonable text without the following problems. The recommended modification methods are: Changing the text, punctuation (such as typos); Deleting inappropriate content; Adding necessary content, etc. The annotator does not require creative writing.

3.2 How to revise

If the annotator encounters the text that needs to be modified, please fill in the number "1" in the "revise" column to indicate that it needs to be revised, and then fill in the "Sentence after revision" column on the right side of this column with the modified candidate sentence. ill in the modified story in the "story after revision" column. Note that whether you are modifying a sentence or a story, you need to fill in the final candidate sentence and story in the two columns.

Sentence	story	Reject	Revise	Sentence after revision	Story after revision	Positions
Candidate sentence	Original sample story	0	1	Sentence after revision	Story after revision	Annotated sentence index
Candidate sentence	Original sample story	0	0			Annotated sentence index

3.3 Revision Examples

(1) Typos. Some text contains typos. Please revise those typos.

(2) Punctuation are not enclosure, matched, or appropriate. The use of punctuation marks in some texts does not conform to the norm, manifested as the mismatch between the front and back quotation marks, and the front and back brackets. Please mark the addition, deletion, and modification of the punctuation marks corresponding to the unreasonable position to make it reasonable. For example, the quotation marks in the following example should use the book title number.

Sentence	story	Reject	Revise	Sentence after revision	Story after revision	Positions	Explanation
"His Majesty only needs half of his cough now. Please do it at once.	[0] The king wanted to make it difficult for Avanti. After discussing with the ministers, he called Avanti and said to him: "Afanti, we urgently need something now, please do it for us. [1] "Your Majesty, please tell me, as long as I can do it," Avanti replied. [2] A minister said. [3] Without even thinking about it, Avanti drew the knife from his waist and put it in Then he coughed loudly and said, "Your Majesty, I would like to ask whether you want to cough in the right half or the left half, whatever you want. [4] After speaking, I left. [5]		1	"His Majesty only needs half of his cough now. Please do it at once."	[0] The king wanted to make it difficult for Avanti. After discussing with the ministers, he called Avanti and said to him: "Afanti, we urgently need something now, please do it for us. [1] "Your Majesty, please tell me, as long as I can do it," Avanti replied. [2] A minister said. [3] Without even thinking about it, Avanti drew the knife from his waist and put it in Then he coughed loudly and said, "Your Majesty, I would like to ask whether you want to cough in the right half or the left half, whatever you want." [4] After speaking, I left. [5]	2	Many front quotation marks lack the corresponding back quotation marks, so they need to be filled in. In addition, the position selection of [2] and [3], because there is no colon before [3], so [3] cannot be used as the insertion position.

(3) Repetition. Some sample story contains context which previously existed in the story.

(4) Containing advertisement, harsh information, or other unrelated contents.

Figure 3: Annotation instruction for constructing the SenPos dataset.

Human Evaluation Guideline
This study aims to evaluate automatic generation systems. Specifically, for each story, we will mask a sentence and put the left context into a generative system, and the masked sentence will be generated by the system. In the process of evaluation, you will be given four generated sentences by four difference systems, denoted as Generation 1-4. You need to compare and rate them in terms of the grammaticality, coherence and relatedness to inputs.
Notes that please annotate the stories carefully based on comprehensive comparison and technically following the instruction.

➢ Grammaticality: intra-sentence grammar quality of generated texts	
Scores	Generation 1: ☑0 □1; Generation 2: ☑0 □1; Generation 3: □0 ☑1; Generation 4: ☑0 □1
e.g.	Input: A little boy was playing in the field and he suddenly found a nettle. He touched it curiously, but he was stabbed severely by the nettle. So, the little boy cried and ran home to tell his mother about it. [MASK] The little boy remembered his mother's words. When he went to play in the field again, he grasped the nettle.Sure enough, he didn't get stabbed.
	Generation 1: The mother listened, comforted the little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed <u>and</u>.
	Generation 2: The mother <u>listen</u>, comforted the little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed.
	Generation 3: The mother listened, comforted the little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed.
	Generation 4: The mother listened, comforted the <u>ed</u> little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed.

➢ Coherence: causal and temporal dependencies within generated texts	
Scores	Generation 1: ☑0 ☑1; Generation 2: ☑0 □1; Generation 3: □0 ☑1; Generation 4: ☑0 □1
e.g.	Input: A little boy was playing in the field and he suddenly found a nettle. He touched it curiously, but he was stabbed severely by the nettle. So, the little boy cried and ran home to tell his mother about it. [MASK] The little boy remembered his mother's words. When he went to play in the field again, he grasped the nettle.Sure enough, he didn't get stabbed.
	Generation 1: The mother listened, comforted the little boy and told him that next time he saw the nettle, <u>he wouldn't be stabbed so that he should be brave to grasp it.</u> (Conflicting Logic)
	Generation 2: The mother listened, <u>comforted the little boy and comforted the little boy</u>. (Repetition)
	Generation 3: The mother listened, comforted the little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed.
	Generation 4: The mother listened, told him that he should be brave to hit it so that it is difficult to comfort him. (Chaotic Scenes)

➢ Relatedness to Inputs: reasonable logical connections to the input context	
Scores	Generation 1: ☑0 □1; Generation 2: ☑0 □1; Generation 3: □0 ☑1; Generation 4: ☑0 □1
e.g.	Input: A little boy was playing in the field and he suddenly found a nettle. He touched it curiously, but he was stabbed severely by the nettle. So, the little boy cried and ran home to tell his mother about it. [MASK] The little boy remembered his mother's words. When he went to play in the field again, he grasped the nettle.Sure enough, he didn't get stabbed.
	Generation 1: The mother listened and told him that the nettle is the nemesis of the nettle, and don't touch it because the nettle will hurt the nettle. (Unreasonable logic)
	Generation 2: The mother told the boy that he must scratch out the nettles. (Unreasonable logic)
	Generation 3: The mother listened, comforted the little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed.
	Generation 4: However, the boy's mother said it was because he was unwell. (Unreasonable logic)

Figure 4: Annotation instruction for manual evaluation of the PlotCom task. The instruction for the OutGen task is similar.

Text	C1	C2	P	T
一个人非常爱他的母猫，几乎到了一种狂热的地步。他虔诚地请求命运女神将母猫变成人。在一天早晨，命运女神同意了他的请求，把母猫变成了人。母猫成人后，主人便急不可耐地与她结了婚。主人认为母猫就是一个真正的新娘。这天晚上，来了几只耗子，新娘出于本能想捉住老鼠。但是，母猫已经变成了太太，老鼠不再怕她了。太太很难过，整天郁郁寡欢。主人看她伤心的样子，后悔因自己的私欲把母猫变成人，又去请求女神把母猫恢复原样。[MASK]	可是母猫只想和主人在一起，再也不想捉老鼠了。	于是母猫又可以肆意地捉老鼠了。	C1	C2
A man loved his female cat almost to the point of fanaticism. He piously asked the Fates to turn the female cat into a woman. One morning, the Fates granted his request and turned the cat into a woman. Then the man was impatient to marry her. The owner thinks the female cat is a true bride. That night some mice came, and the bride instinctively wanted to catch them. But the cat had become a woman, and the mouse was no longer afraid of her. <mark>The wife is very sad, all day long depressed.</mark> The man saw her sad appearance, regretting for his own selfish desire to turn the cat into a woman, and asked the Fates to restore the female cat. [MASK]	But she just wanted to be with the man and didn't want to catch mice any more.	So the female cat was free to hunt mice at will.	C1	C2
天后朱诺有一只母牛，天神十分想要。但是朱诺让一个百眼巨人来看管那只母牛。于是天神派出了墨丘利去偷那只母牛。百眼巨人虽然知道墨丘利会讲故事，会唱歌，能让神安眠，但还是耐不住寂寞，开始听他讲故事，听他唱歌。不出意料，百眼巨人困得睡着了，闭上了他的一百只眼睛。[MASK]天后很生气，为了惩罚他，就把他的一百只眼睛都安到了孔雀的羽毛上。百眼巨人知道自己犯了错，也心甘情愿受罚了。	墨丘利很顺利地偷走了母牛。	墨丘利和百眼巨人成了好友，一起看管那只母牛。	C2	C1
Juno had a cow that the god wanted very much. <mark>But Juno put a hundred-eyed giant in charge of the cow.</mark> So the gods sent Mercury to steal the cow. Even though the Hundred-eyed giant knew that Mercury could tell stories and sing songs that would put the gods to sleep, he could not bear his loneliness and began to hear his stories and songs. As expected, the hundred-eyed giant fell asleep and closed his hundred eyes. [MASK] In order to <mark>punish</mark> him, she fastened all of his hundred eyes to the feathers of the peacock. The hundred-eyed giant knows he made a mistake, and he's willing to take it.	Mercury stole the cow without any trouble.	Mercury and the hundred-eyed giant became friends and they guarded the cow together.	C2	C1

Table 2: Two ClozeT examples that LongLM$_{large}$ does wrong. **C1/C2**: The candidates; **P**: The predicted candidate of LongLM$_{large}$; **T**: The true labels. The red entities or events relate to the focused commonsense relations.

Text	Sen	P	T
年老的狮子重病躺在洞里。[1]除了狐狸之外，动物们都去问候国王。[2]狼便趁机在狮子面前诬陷狐狸，说狐狸胆大包天，藐视大王，竟敢不来问候。[3]正在此时，狐狸进来了，听到了狼所说的最后几句。[4]狮子一见到狐狸就怒吼起来，狐狸马上请求让他解释几句。[5]他说："在所有向大王问候的动物之中，有谁像我这样忠诚，为你四处奔走，遍访名医，寻找妙方呢？"[6]狐狸说："将狼的皮活剥下，趁热裹在身上。[7]"狼立刻成为一具尸体，躺在了那里。[8]狐狸得意地笑着说："你不应当怂恿主人起恶念，而应该诱导他发善心才对呀。[9]"	狮子立即命令他将药方说出来。	[5]	[6]
The old lion was lying seriously sick in the cave. [1] The animals all went to visit the king except the fox. [2] The wolf took the opportunity and framed the fox in front of the lion, claiming that the fox dared to despise the king, since he did not come for a visit. [3] At this moment, the fox came in and heard the last few words of the wolf. [4] The lion roared when he saw the fox, <mark>and the fox immediately begged to provide explanation. [5] "Who, among all the animals that have visited you, is as loyal as I am, running around to visit famous doctors for the efficacious prescriptions for you?" he said. [6] "Peel the wolf's skin off alive and wear it when it's hot. [7]"</mark>, the fox said. The wolf turned into a corpse at once and lay there. [8] the fox smiled complacently and said: "You should not instigate your master to raise evil thoughts, but guide him to be kind. [9]".	The lion promptly ordered the fox to offer the <mark>prescription</mark>.	[5]	[6]
森林绿了又枯，枯了又绿；小河肥了又瘦，瘦了又肥；燕子来了又去，去了又来。[1]不知不觉中，毛克利已在森林度过了十一个春秋。[2]如今，他已长成了一个英俊威武的少年。[3]这些年来，毛克利因其独特的个性，赢得了许多动物的喜爱，一直过着无忧无虑自由自在的生活。[4]毛克利一直以为自己是一只小狼。[5]可是随着年龄的增长，他的大脑越来越发达，考虑的问题也越来越多。[6]这个问题一连困扰他好多天，使他心烦意乱，烦躁不安。	有一天，他想到了一个奇怪的问题：为什么自己的身躯和相貌与其他的狼一点儿也不像呢？	[4]	[6]
The forest is green and then withered, withered and then green; the river is wide and then thin, thin and then wide; the swallows come and then go, go and then come. [1] Unconsciously, Mowgli has spent eleven years in the forest. [2] Now, he has grown into a handsome and mighty teenager. [3] Over the years, Mowgli has won the love of many animals because of his unique personality, and has been living a carefree and free life. [4] Mowgli always thought he was a little wolf. [5] But as he grows older, his brain becomes more developed and he considers more and more <mark>questions</mark>. [6] This <mark>question</mark> has confused him for many days, making him distraught and irritable.	One day, he thought of a strange <mark>question</mark>: why his body and appearance look nothing like other wolves.	[4]	[6]

Table 3: Two SenPos examples that LongLM$_{large}$ does wrong. **Sen**: The removed sentences; **P**: The predicted positions of LongLM$_{large}$; **T**: The true labels. The red entities or events relate to the focused inter-sentence relations.

Context	Generation	Truth
一个小男孩在田野里玩，忽然发现了一株荨麻。他好奇地摸了一下，没想到却被荨藤狠狠地刺了一下。于是，小男孩哭着跑回家告诉母亲这件事情。[MASK]小男孩记下了妈妈的话。当他再一次去田野里玩时，一把握住了那棵荨麻。果然，它不扎手了。	母亲听了，告诉他，荨麻是荨藤的克星，千万不要去触碰它，因为荨藤会把荨麻扎伤。	母亲听到后安慰了小男孩，并且告诉他，下次看到荨麻时，要勇敢地一把握住，这样就不会被刺伤了。
A little boy was playing in the field and he suddenly found a nettle. He touched it curiously, but he was stabbed severely by the nettle. So, the little boy cried and ran home to tell his mother about it. [MASK] The little boy remembered his mother's words. When he went to play in the field again, he grasped the nettle. Sure enough, he didn't get stabbed.	The mother listened and told him that the nettle is the nemesis of the nettle, and don't touch it because the nettle will hurt the nettle.	The mother listened, comforted the little boy and told him that next time he saw the nettle, he should be brave to grasp it so that he wouldn't be stabbed.
有一天，一个小孩独自在河里洗澡。突然他的腿开始抽搐，并开始下沉，眼看就要被淹死了。就在孩子极力挣扎的时候，看到了自己的老师，于是他连忙大声呼救。那个老师看到自己的学生擅自下水游泳，而且现在还有了危险，就很生气地在岸边责备学生不听话。[MASK]学生得救后，接受了老师的批评，再不去河里洗澡了。	学生听了老师的话，赶紧跳进了河里，没过多久就淹死了，老师也淹死了。	学生哭着说现在最重要的是先把他救起来。
One day, a child was bathing in a river alone. Suddenly his legs began to twitch, and he began to sink and was about to drown. When the child was struggling, he saw his teacher, so he hurriedly called for help. The teacher saw that his student went into the water to swim without permission, and now he was still in danger. So he angrily blamed the student for being disobedient. [MASK] After the student was saved, he accepted the teacher's criticism and stopped bathing in the river.	The student heard the teacher's words and hurriedly jumped into the river. It didn't take long to drown, and the teacher also drowned.	The student cried and said that the most important thing now was to save him first.

Table 4: Two PlotCom examples that LongLM$_{large}$ generates unreasonable outputs.

Keywords	Generation	Truth
标题：麦草、煤块和豆子 大纲：{想去河对面玩, 伤心的边哭, 胆子走上, 正当正午, 火焰燃成, 纷纷丧命, 绕路走, 桥上走}	一个人绕路走，他看到桥对面有一家麦草铺的摊子。他想去河对面玩，便绕路走去。他走着走着，发现桥对面有一家豆子店，他想去河对面玩。于是他走了过去。可豆子店里没有人，他伤心的边哭边喊着其他人。这时，有人经过，看到他有心事，便问他为什么这样伤心。他告诉他们自己被火焰燃成灰了，没有人知道他是谁。大家才知道他是为了自己而伤心。于是，他们都开心的笑了。	麦草、煤块和豆子是好朋友。有一天，他们想去河对面玩，不过过河的桥还没完全修好，但是它们不愿意绕路走，非要从没修好的桥上走。煤块性子比较急，拉着麦草大着胆子走上了桥，这时正当正午，烈阳把麦草晒得自燃了，煤块也被麦草的火焰燃成了灰烬，两人纷纷丧命。还在岸边的豆子看到这一幕非常伤心的边哭边远离了那座危桥。
Title: Straw, coal, and beans **Outline:** {want to go across the river to play, cry sadly, walk boldly, just at noon, burned to, lost their lives, detour, walk on the bridge}	A man took a detour and saw a straw shop across the bridge. He wanted to go across the river and took a detour. As he walked, he found a bean shop across the bridge. He wanted to go across the river to play. So he went over. But there was no one in the bean shop. He cried sadly and called for others. At this moment, someone passed by and saw that he was worried. They asked him why he was so sad. He told them that he had been burned to ashes and that no one knew who he was. Everyone knew that he was sad for himself. So they all smiled happily.	Straw, coal, and beans are good friends. One day they wanted to go across the river to play. The bridge across the river was not yet fully repaired, but they did not want to take a detour and walked on the bridge. Coal was impatient, so he took the straw and walked boldly across the bridge. Just at noon, the straw was set on fire by the hot sun, and the coal was burned to ashes by the straw flame. Both of them lost their lives. The beans on the shore saw this scene and cried sadly away from the dangerous bridge.
标题：神秘的大自然 大纲：{穷人拥有南半边, 富人拥有北半边, 正要得逞时, 想要霸占穷人, 土地神公正, 假装土地神, 穷人田里, 吓唬穷人}	一个富人拥有南半边的地，他便假装土地神，想要霸占穷人的田里。他假装土地神，实际上是想要霸占穷人的粮仓。正要得逞时，一个穷人田里来了一只猴子。富人便作土地神，吓唬穷人。穷人就假装土地神，实际上是想要霸占穷人的粮仓。正要得逞时，猴子也来了。两人打得不可开交。猴子得手后，便想要霸占穷人的粮仓。富人很清楚，这只是个假象，他故意说出了土地神，想要霸占穷人的粮仓。但是猴子却相信了，还当真了。正要得逞时，富人又被猴子打伤了。	穷人和富人共同拥有一块地。富人拥有北半边，穷人拥有南半边。他们在各自的地里撒下了种子。穷人田里的种子发芽了，长出了嫩叶。而富人田里的种子都死了。富人仗势欺人，想要霸占穷人的土地，还让儿子躲在田边地洞里假装土地神，吓唬穷人。正要得逞时，真正的土地之神出现了。土地神公正地辨明了土地的归属。而富人父子因为恶行，被变成了整天钻在地洞里的鼹鼠，不见天日。
Title: The mysterious nature **Outline:** {the poor have the south half, the rich have the north half, is about to succeed, want to occupy the poor, the land god rightly judged, pretend to be the land god, the poor man's field, frighten the poor}	A rich man owned the southern half of the land, so he pretended to be the land god and wanted to occupy the poor man's field. Pretending to be a god of the earth, he really wanted to commandeer the granary of the poor. Just as he was about to succeed, a monkey came to the poor man's field. The rich dress up as the land god to frighten the poor. The poor pretended to be the land god, but really wanted to commandeer the granary of the poor. Just as he was about to succeed, the monkey came. The two fought like cats and dogs. When the monkey had succeeded, he wanted to occupy the poor man's granary. The rich man knew that this was only an illusion. He deliberately told the land god that he wanted to occupy the poor man's granary. But the monkey believed and took it seriously. Just as he was about to succeed, the rich man was wounded by the monkey.	The poor and the rich own a piece of land together. The rich have the north half, the poor have the south half. They sowed seeds in their fields. The seeds in the poor man's field sprouted and put forth young leaves. All the seeds in the rich man's field died. The rich tried to bully the poor and wanted to occupy the poor man's land, and even let his son hide in holes near the fields and pretend to be the land god to frighten the poor. Just as he was about to succeed, the real land god appeared. The land god rightly judged the ownership of the land. Because of their evil deeds, the rich man and his son were turned into moles who burrow in the ground all day and never see the light of day.

Table 5: Two OutGen examples that LongLM$_{large}$ generates unreasonable outputs. Red words are the given phrases.