SWAG: Storytelling With Action Guidance

Karim El-Refai [*1] **Zeeshan Patel** [*1] **Jonathan Pei** [*1] **Tianle Li** [1]

Abstract

Automated long-form story generation typically employs long-context large language models (LLMs) for one-shot creation, which can produce cohesive but not necessarily engaging content. We introduce Storytelling With Action Guidance (SWAG), a novel approach to storytelling with LLMs. Our approach reduces story writing to a search problem through a two-model feedback loop: one LLM generates story content, and another auxiliary LLM is used to choose the next best "action" to steer the story's future direction. Our results show that SWAG can substantially outperform previous end-to-end story generation techniques when evaluated by GPT-4 and through human evaluation, and our SWAG pipeline using only open-source models surpasses GPT-3.5-Turbo.

1. Introduction

Large language models (LLMs) have recently changed the landscape of content generation. A number of works have proposed techniques for short story generation (Fan et al., 2018; Wilmot & Keller, 2021; Rashkin et al., 2020; Xu et al., 2018). However, it has been a major challenge for AI to generate long-form stories that are both *coherent* and *interesting* (Oatley, 1995; Charniak, 2004; Alabdulkarim et al., 2021a). This remains a challenge with SoTA LLMs such as GPT-4 (OpenAI, 2023), Llama-2 (Touvron et al., 2023), and Mistral (Jiang et al., 2023).

While new LLMs have impressive content generation abilities, these models can have unstable outputs due to their unsupervised training objective. Recent advancements in alignment techniques for LLMs allow more control over output generation. Reinforcement Learning with Human Feedback (RLHF) (Christiano et al., 2023) is a popular alignment paradigm that requires training a reward model on a dataset of human preferences and fine-tuning the LLM to maximize the reward while ensuring that the model does not drift too far from the original pretrained model. This process can be complex and expensive, producing unstable results due to the imperfect reward model or other issues with approximating the KL divergence penalty. Direct Preference Optimization (DPO) (Rafailov et al., 2023) is another technique for aligning language models which optimizes the constrained reward maximization problem in RLHF in a single step of policy training. DPO is much more computationally efficient and stable, with similar or improved performance compared to existing RLHF techniques such as RLHF with proximal policy optimization (PPO) (Schulman et al., 2017).

We propose SWAG, an algorithm for iteratively generating engaging and captivating stories using LLMs. In our work, we structure storytelling as a search problem. This paradigm allows us to formulate the problem as finding the "optimal path" in a search space of possible stories given a story idea. By having another model guide the LLM during the story writing process, we can improve control over the story direction and create more engaging content. At a high level, we train an *action discriminator* LLM (AD LLM) to determine the next best action to take given the current state of a story. Using the generated action, we prompt another LLM to write the next part of the story based on the given action. This feedback loop can generate long-context stories that are fascinating and amusing to read. The main component of our system is the AD LLM, which helps pave the path for the story by selecting the next best "action" to continue the story. This AD LLM can be paired with any open-source model (e.g. Llama-2-7B, Mistral-7B) or closed models (e.g. OpenAI's GPT-4) for generating the story. Our algorithm offers a simplified approach to storytelling, allowing for fine-grain control over the story content progression while providing the flexibility to integrate custom models for writing the story or using LLM services offered by other companies through APIs.

2. Related Work

Prior works have attempted to improve the quality and/or diversity of story generations in a variety of ways.

Storytelling with reinforcement learning

[*]Equal contribution [1]UC Berkeley. Correspondence to: Karim El-Refai <karim.el-refai@berkeley.edu>, Zeeshan Patel <zeeshanp@berkeley.edu>, Jonathan Pei <jonnypei@berkeley.edu>.

In the context of content generation, reinforcement is largely used for fine-tuning (Chang et al., 2023; Bai et al., 2022) or auxiliary model guidance (Peng et al., 2022; Castricato et al., 2022).

Perhaps most similar to our work are methods that involve dynamic inference-time option-selection and/or classification (Alabdulkarim et al., 2021b; Tambwekar et al., 2019; Peng et al., 2022). Our approach differs from prior ones in that our model (1) uses an adapted LLM to interpret an internal representation of the current story; (2) is highly modular; and (3) is prompting-based. These aspects contribute to our method's diverse story generations despite having such a simple, flexible structure.

Controlled Text Generation (via prompting)

The recent advancements in language models have substantially increased the popularity of (simpler) prompting approaches such as chain of thought. Prompts may be manually designed (Brown et al., 2020) or automatically designed (Shin et al., 2020; Zou et al., 2021); prompting may also be an iterative process (Wei et al., 2022). Some works such as (Qin & Eisner, 2021; Lester et al., 2021) also explore continuous soft prompts. Compared to prior work, our contribution is an iterative feedback-prompting-based method that utilizes an auxiliary LLM for control, enabling more diverse storytelling.

Human-in-the-loop story generation

As opposed to automatic story generation, some previous works use human-in-the-loop methods to generate interesting long stories (Goldfarb-Tarrant et al., 2019; Coenen et al., 2021; Chung et al., 2022; Mirowski et al., 2022; Martin et al., 2017; Wang & Gordon, 2023; Lin & Riedl, 2021). We emphasize that although our method is completely automatic without any human intervention, the flexibility of the AD's action space makes it quite intuitive for a human collaborator to "tune" our method towards their own liking.

3. Methods

Our creative storytelling method consists of two primary components: the story generation model and the action discriminator model (AD LLM). SWAG enables the use of any open-source LLM or LLM service for story generation. We create an AD LLM by collecting preference data for story actions, and aligning a pretrained LLM on our preference dataset. We visualize our training pipeline in Figure 1 below.

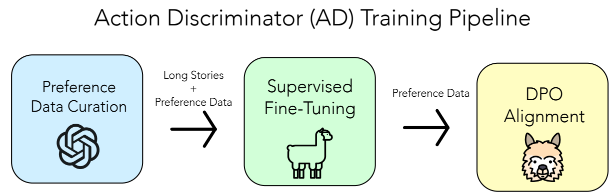

Figure 1. **SWAG AD LLM Training Pipeline.** After curating long story and action preference data from GPT-4 and Mixtral-8x7B, we perform SFT on a base open-source LLM, and then align our model with more preference data using DPO to produce our action discriminator model (AD LLM).

3.1. Preference Data Collection

We use a preference dataset of story actions to train a model to learn how to choose an action for the next part of the story. Given a list of actions, we want our AD LLM to select the best action that will keep the reader engaged with the story. Several datasets contain thousands of story prompts and ideas, but there are no preference datasets for choosing the next direction for a story.

To generate this data efficiently, we developed a pipeline that prompted OpenAI's GPT-4 and Mixtral-8×7B (Jiang et al., 2024), to choose the next best action given a "story state". We define the story state to be

$$X = (\mathcal{P}, \mathcal{S}),$$

where $\mathcal{P}$ is the story prompt and $\mathcal{S}$ is the current continuation of the story prompt. We use random subset of the Writing Prompts (Fan et al., 2018) dataset to acquire a diverse set of story prompts. For each story prompt from this subset, we prompt GPT-4 and Mixtral-8×7B to write an initial paragraph $S^{(0)}$, forming the dataset

$$\mathcal{D} = \left\{ \left(P_i, S_i^{(0)} \right) \right\}_{i=1}^{n}$$

These story states provide a simple yet comprehensive starting point for the AD LLM to find the best path to continue generating the given story.

After curating the initial story states, we generate preference data on the next best action for continuing the story. We model this preference data by having a "chosen" and "rejected" action for each story state. For the story state $\mathcal{S}^{(k)}$, the chosen action $c^{(k)}$ is what we would like the LLM to choose when deciding the next best direction for the story, and the rejected action $r^{(k)}$ is the path we would like the LLM to avoid for the next part of the story. This preference data allows our model to understand how to rank different actions for the diverse set of story prompts that it will encounter during test-time.

To generate the ranking data, we prompt GPT-4 and Mixtral-8×7B with an initial story state $\mathcal{S}$ and a list of "actions" $\mathcal{A}$ to choose the best direction for the next paragraph in the story. The action used by GPT-4 to generate the next paragraph is set as the chosen action, and we then randomly choose an action from the remaining actions as the rejected action.

We distill multiple datasets for supervised fine-tuning (SFT), direct preference optimization (DPO), and evaluation.

3.2. Supervised Fine-Tuning (SFT)

In the SFT phase, we follow the typical set up of starting with a pre-trained LLM and fine-tuning it with supervised learning, effectively using a maximum likelihood objective. We fine-tune the LLM on our downstream task of action discrimination on the preference dataset we curated using GPT-4 and Mixtral-8×7B.

One problem with using open source LLMs is that they are not necessarily prepared to handle long-context inputs. For example, the Llama-2-7B model has a max context length of 4096 by default. In order to alleviate this issue, we use the approach to SFT presented in LongLoRA (Chen et al., 2023). We replace the default Llama-2-7B attention with shifted sparse attention to enable computationally efficient long-context fine-tuning. The standard self-attention algorithm requires $\mathcal{O}(n^2)$ computations, which results in high memory and time costs for long-context fine-tuning. The authors suggest that by using short attention across smaller groups of tokens, and by shifting the group partition by half group size in half attention heads, we can effectively approximate full self-attention at a much lower cost. This S^2-Attn approach is easy to implement and removes any possibilities of overfitting on specific attention patterns due to the shifting mechanism (Chen et al., 2023). We set the embedding and normalization to be trainable parameters during the low rank adaptation (LoRA) training (Hu et al., 2021). This technique allows our model to better understand the best next story direction for longer stories without the need for extensive compute resources.

We conduct SFT in two stages. During the first stage, we fine-tune the AD LLM on a dataset of long stories. We train the model to take a prompt as an input and generate a long-context story. This process ensures that models like Llama-2-7B, with their shorter default context length, can accurately process longer data sequences. In the second stage, we fine-tune our new long-context AD LLM on a preference dataset with chosen and rejected actions for the next story direction. This stage helps the model better understand the downstream task for which we want to build a preference model.

3.3. Direct Preference Optimization (DPO)

We utilize DPO to further refine the results of our action discriminator model. In DPO, we want our policy π_{SFT} to learn how to rank chosen responses $c^{(k)}$ over rejected responses $r^{(k)}$ in a preference model framework. In PPO, we use a learned reward model $R_\theta(x, y)$ for which we estimate parameters by taking the maximum likelihood over our static preferences dataset. DPO instead allows us to define a mapping from the optimal reward model to our language model policy, enabling the training of our language model to satisfy our preferences directly with a single cross-entropy loss (Rafailov et al., 2023). More specifically, under an appropriate preference model, we can derive the optimal reward function in terms of the optimal language model policy π_*, original SFT policy π_{SFT}, a constant β, and a partition function $Z(\cdot)$ as follows:

$$R^*(x, y) = \beta \cdot \frac{\pi_*(y|x)}{\pi_{\text{SFT}}(y|x)} + \beta \log Z(x).$$

We can then plug this reward into our preference model, providing a simple training procedure on our dataset of preferences $(x, c^{(i)}, r^{(i)})$:

$$\pi_\theta = \max_\pi \mathbb{E}_{(x,c^{(i)},r^{(i)})\sim\mathcal{D}} \log \sigma \left(\beta \frac{\pi(c^{(i)}|x)}{\pi_{\text{SFT}}(c^{(i)}|x)} - \frac{\pi(r^{(i)}|x)}{\pi_{\text{SFT}}(r^{(i)}|x)}\right). \quad (1)$$

Thus, in the DPO procedure, we calculate the probabilities of $(x, c^{(i)})$ and $(x, r^{(i)})$ from both π_{SFT} and the DPO model, and then we can compute Eq. 1 and backpropagate to update (Tunstall et al., 2023). Using DPO, we can refine the SFT model on our preferences dataset to generate actions that are better aligned with the actions chosen by GPT-4 and Mixtral-8×7B.

3.4. SWAG Feedback Loop

The main algorithm in our method is the SWAG feedback loop that enables the action guidance mechanism. This feedback loop is a three step process and can be configured to use open-source LLMs, closed-source LLMs, or a hybrid of both for inference (beyond story generation).

First, we generate an initial story state $X^{(0)} = (\mathcal{P}, \mathcal{S}^{(0)}, \emptyset)$ by passing the story prompt $\mathcal{P}$ into the story generation model π_{story} to yield the initial paragraph $\mathcal{S}^{(0)}$. Next, we pass $X^{(0)}$ into our AD LLM π_{AD} along with a list of (predefined) possible actions (included in Appendix B), and π_{AD} generates the next best action to continue the story.

After generating the next best action, we update our story state to be

$$X^{(0)} = (\mathcal{P}, \mathcal{S}^{(0)}, \mathcal{A}^{(0)}).$$

To generate the story, we iteratively repeat this process of (1) generating the next paragraph in the story via π_{story} and (2) generating the optimal subsequent action to take via π_{AD}. See Algorithm 1 for a pseudocode implementation of the SWAG feedback loop.

Algorithm 1 Storytelling With Action Guidance (SWAG)

procedure SWAG($\mathcal{P}, \pi_{\text{story}}, \pi_{\text{AD}}, k$)
  $\mathcal{S}^{(0)} \leftarrow \pi_{\text{story}}(\mathcal{P})$
  $\mathcal{A}^{(0)} \leftarrow \pi_{\text{AD}}(\mathcal{P}, \mathcal{S}^{(0)})$
  $X^{(0)} \leftarrow (\mathcal{P}, \mathcal{S}^{(0)}, \mathcal{A}^{(0)})$
  for $i = 1 \ldots k$ **do**
    $\mathcal{S}^{(i)} \leftarrow \mathcal{S}^{(i-1)} + \pi_{\text{story}}(X^{(i-1)})$
    $\mathcal{A}^{(i)} \leftarrow \pi_{\text{AD}}(\mathcal{P}, \mathcal{S}^{(i)})$
    $X^{(i)} \leftarrow (\mathcal{P}, \mathcal{S}^{(i)}, \mathcal{A}^{(i)})$
  end for
  return $\mathcal{S}^{(k)}$
end procedure

The SWAG feedback loop can be run as many times as needed until the desired story length is reached—we can freely choose k. This feedback mechanism can be implemented between any two LLMs (for story and AD), allowing for enhanced modularity in content generation for stories.

3.5. Ablations

We perform several ablations on π_{story} and π_{AD} to test the performance of our algorithm. Specifically, we run pairwise comparisons between different combinations of π_{story} and π_{AD} models to gauge the quality of stories generated by SWAG.

In the π_{story} ablation, we test different models to generate the story with a fixed π_{AD}. We run the SWAG inference loop with several open-source and closed-source LLMs as π_{story}. This ablation provides insight into the level of improvement in story quality from different base models.

In the π_{AD} ablation, we test different models to generate the next story action with a fixed π_{story}. We trained two different AD LLMs for this ablation with the same SFT and DPO preference datasets.

To test SWAG on closed-source LLMs, we also set up our inference pipeline with GPT-3.5-Turbo and GPT-4-Turbo. Here, we simply set GPT-3.5-Turbo and GPT-4-Turbo to be both π_{AD} and π_{story} in the SWAG feedback loop. With these experiments, we aim to show the effectiveness of SWAG even without fine-tuning an AD as a preference model.

4. Experiments

4.1. Experimental Setup

In our experiments, we aimed to evaluate the quality of the stories generated by our inference pipeline with different combinations of models and AD settings. We also explored if GPT-4 had any bias in ranking story actions for the preference dataset and the effects of this bias on our AD LLM.

4.2. Dataset

In order to train an AD LLM that can process long-form content, we fine-tuned our model on a dataset of long stories. We distilled this dataset of long stories from Llama-2-7B, Mistral-7B, and Mixtral-8×7B using a sample of prompts from the WritingPrompts dataset. We generated 20,000 long stories from these models, providing a diverse distribution of stories for SFT. We fine-tuned Llama-2-7B and Mistral-7B on this long stories dataset, allowing them to have a context length of 32,768 tokens.

For our DPO preference dataset, we prompted GPT-4 and Mixtral-8×7B to generate preference data on a sample of approximately 60,000 prompts from the WritingPrompts dataset. One key aspect of this preference data is the potential options for story actions. We distilled a list of 50 different story actions from GPT-4 and used this set of actions for all training experiments. Some examples of actions in the set include “add suspense”, “add mystery”, “add character development”, etc. We used 34,000 preference data samples for fine-tuning the AD LLM to understand the downstream task of choosing the next story direction, and we used 25,000 samples to train the preference model using DPO. In the DPO dataset, we noticed an imbalance in the distribution of chosen actions by GPT-4. In Figure 2, we can see the substantial difference in the number of stories for which “add suspense” was selected compared to other options. This observation implies that GPT-4 has an inherent bias while selecting actions to continue the story.

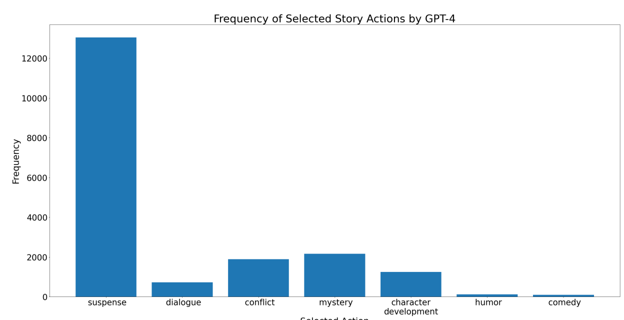

Figure 2. **Original Distribution of Actions.** We observe a severe distribution imbalance where the vast majority of actions selected is “add suspense”. *Note: actions chosen with frequency less than 100 not shown.*

In order to mitigate this effect, we generated more preference data from GPT-4, but this time, we removed the option to add suspense to the story. This would force GPT-4 to focus on other actions as well, resulting in a more spread out distribution of actions. After generating the new data, we took a random sample of 3,000 prompts from the original preference dataset with “add suspense” as the chosen action and merged it with our new dataset. In Figure 3, we can view the new distribution of story actions and notice that it is much more spread out, allowing for more variability in future story directions.

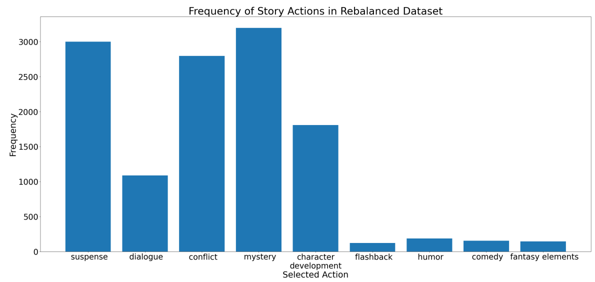

Figure 3. **Normalized Distribution of Actions.** After our rebalancing procedure, we observe a more uniform distribution among the top 5 actions chosen. *Note: actions chosen with frequency less than 100 not shown.*

We collected three different datasets for SFT, DPO, and evaluation. Rebalancing was only done on the DPO dataset, but a similar approach could have been used on the SFT dataset as well. Due to constraints with the GPT-4 API, we were unable to generate enough data for rebalancing the SFT dataset. However, it is worth noting that the SFT process allows our model to better understand the downstream task, but the DPO procedure is more critical for generating a preference model that produces useful results as shown in later experiments.

4.3. Training

For our AD LLM training, we first used a dataset of long stories to fine-tune our model to process long-context sequences. Then, we use a separate preference dataset collected for SFT to fine-tune our base AD LLM. We used approximately 34,000 ranking samples for SFT, and we trained the model to predict the next best action given the initial story state. We fine-tuned Llama-2-7B on this dataset for 5300 steps, with a mini-batch size of 1 and 64 gradient accumulation steps using 8 $\times$ A100 80GB GPUs (so one step processes 64 stories, and 530 steps is about one epoch). Completing the SFT process for each model required about 36 hours. We used the LongLoRA (Chen et al., 2023) approach with Flash Attention 2.0 (Dao, 2023) for SFT to enable fast fine-tuning on limited compute. We used the AdamW optimizer (Loshchilov & Hutter, 2019) with $\beta_1 = 0.9, \beta_2 = 0.95$, a learning rate of 3e-5, and 30 warm-up steps with a constant learning rate scheduler.

We used DPO to train a preference model on two SFT model checkpoints, which were trained for 2650 and 5300 steps, respectively. The DPO training ran for 1000 steps for each model on approximately 25,000 samples of our preference dataset. We used a learning rate of 5e-4 with an AdamW optimizer and cosine annealing scheduler, both on default settings of $\beta_1 = 0.9, \beta_2 = 0.999$. We also used LoRA in our DPO training for both checkpoints, with $\alpha = 16$, $r = 8$, and a dropout of 0.05. We conducted the DPO training using the Hugging Face Transformers Reinforcement Learning (TRL) (von Werra et al., 2020) library in a similar setting as SFT with 8 $\times$ A100 80GB GPUs but with a mini-batch size of 1 and 8 gradient accumulation steps. Each DPO training required approximately 12 hours with this setup on the rebalanced preference dataset, and we checkpointed our model at every 100 training steps. DPO for both checkpoints displayed convergence after approximately 800 steps of training.

4.4. Inference

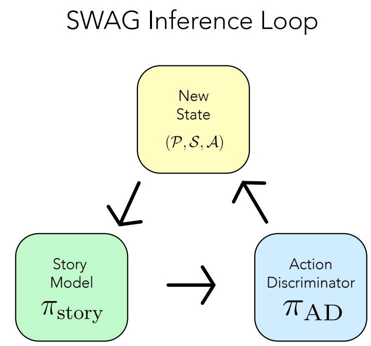

Figure 4. **SWAG Inference Loop**. After sampling a story prompt and generating the initial paragraph, we pass the story state to our AD LLM to generate the next story action. The new state is passed back to the story model, and the process is repeated till a complete story is generated.

Our inference pipeline requires two models: the action discriminator π_{AD} and the story generation model π_{story}. We create a feedback loop between these two models to generate our story.

For our experiments, we evaluated the performance of different combinations of π_{AD} and π_{story} across a set of test story prompts. For each story prompt $\mathcal{P}$, we ask π_{story} to write the initial paragraph, and then, with this initial story state $(\mathcal{P}, \mathcal{S})$, we instruct π_{AD} to select the optimal action for the subsequent paragraph.

In the action discriminator model π_{AD} ablation, we used our own fine-tuned and aligned Llama-2-7B and Mistral-7B AD LLMs and GPT-4-Turbo. For the story generation model π_{story} ablation, we used the base Llama-2-7B, Mistral-7B, GPT-3.5-Turbo, and GPT-4-Turbo models. For our open-source model generations, we also compare the performance when using a π_{AD} that was tuned with a different base model as π_{story}.

To analyze the baseline performance for story generation, we generated stories with each π_{story} by giving an initial story prompt and repeatedly prompting it to continue the story. The results of these end-to-end (E2E) generation ablations are shown in Table 2.

Finally, we analyze if our π_{AD} models are better in choosing actions than a random selection. Our AD LLMs, trained using DPO, had a choice of only 30 actions during SFT and DPO. Using these 30 actions, we generated stories from the base Llama-2-7B and Mistral-7B models using our SWAG pipeline. However, in this ablation, we replaced π_{AD} and instead selected an action randomly from the list for each step of the loop.

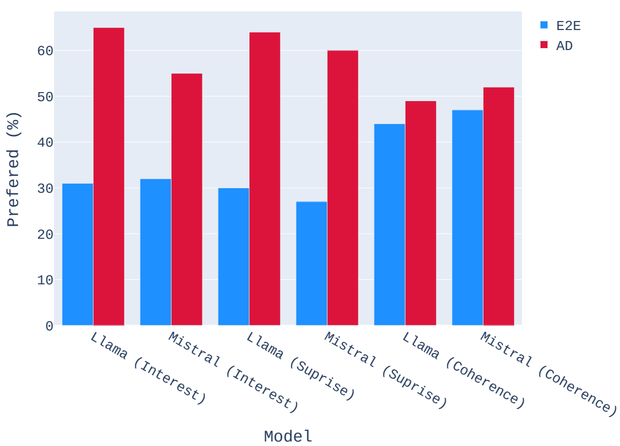

Figure 5. Open source models AD, Llama-2-7B and Mistral-7B, win-rate against GPT-3.5-Turbo E2E on human evaluation data. The win-rate is calculated by averaging wins, losses, and ties. We count win as a score of 1, tie as a score of 0.5, and loss as a score of 0.

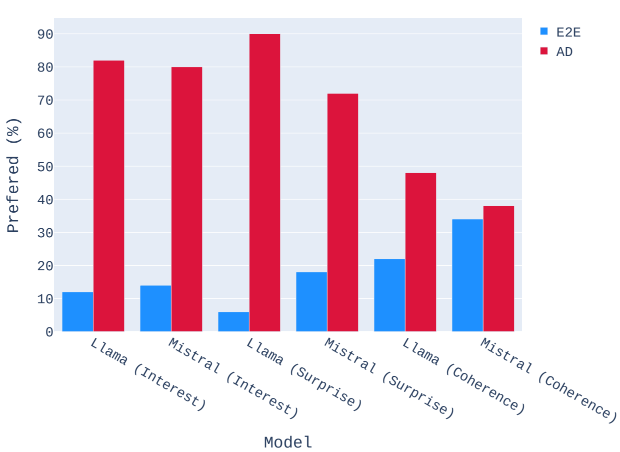

Figure 6. Preferred rates between E2E and AD for open source models, Llama-2-7B and Mistral-7B on human evaluation data. Preferred rate of a model is the percentage of the human evaluator preferring the story from the model.

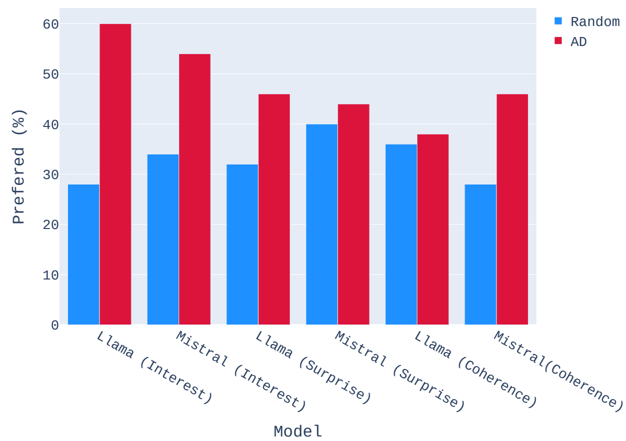

Figure 8. Preferred rate between random actions and AD for open source models, Llama-2-7B and Mistral-7B on human evaluation data. Preferred rate of a model is the percentage of the human evaluator preferring the story from the model.

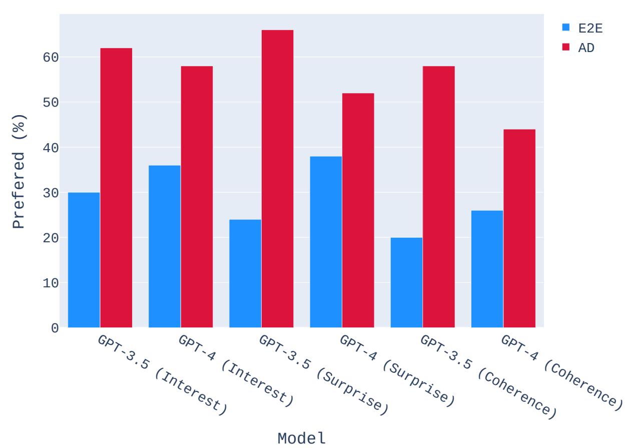

Figure 7. Preferred rate between E2E and AD for OpenAI models, GPT-3.5-Turbo and GPT-4-Turbo on human evaluation data. Preferred rate of a model is the percentage of the human evaluator preferring the story from the model.

4.5. Human Evaluation

Our human evaluation setup is heavily inspired by (Zhu et al., 2023). We run human evaluations comparing stories generated by various methods across three aspects: interesting-ness, surprise, and coherence. For each of 12 pairwise comparisons of two methods, we ask Surge AI workers to answer three preference questions about 50 pairs of stories generated by the methods we compare. We display the preference questions in Table 1, where each question corresponds to an aspect of story quality. We display our human annotation results in Figures 5-8.

Table 1. Three questions asked to human annotators for 50 comparison story plot pairs.

Q1	Which story plot is more **interesting** to you overall?
Q2	Which story created more **suspense** and **surprise**?
Q3	Which story is more **coherent** and **consistent** in terms of plot structure?

4.6. Machine (GPT-4-Turbo) Evaluation

Recent development in open-ended benchmarks shows promising results in evaluating LLM's response, with increasing utilization of GPT-4 in place of human judges, such as MT-Bench (Zheng et al., 2023) and AlpacaEval (Dubois et al., 2023). Employing a similar strategy, we conduct evaluations with GPT-4-Turbo as a judge to pairwise compare two stories and pick the more interesting, engaging, and consistent story or a tie. The system prompt can be found in Appendix A.3. We evaluated several open and proprietary variants of AD LLMs against different baselines (random action, GPT-3.5-Turbo, etc.), with results presented in Table 2.

Table 2. Evaluation results of pairwise comparisons between SWAG (AD) and other baselines with GPT-4-Turbo as judge. The win-rate is calculated by averaging wins, losses, and ties. We count win as a score of 1, tie as a score of 0.5, and loss as a score of 0.

AD vs E2E	Win-Rate (AD)	AD	E2E	Tie
Mistral-7B	**68.0%**	58	22	20
Llama-2-7B	**54.5%**	47	38	15
GPT-3.5-Turbo	**77.5%**	66	8	23
GPT-4-Turbo	**61.5%**	49	24	25
AD vs Random	**Win-Rate (AD)**	**AD**	**Random**	**Tie**
Llama-2-7B	**53.0%**	45	39	16
Mistral-7B	**67.5%**	61	26	13
AD vs GPT-3.5	**Win-Rate (AD)**	**AD**	**GPT-3.5**	**Tie**
Mistral-7B	19.5%	11	72	17
Llama-2-7B	31.0%	19	57	24
E2E vs GPT-3.5	**Win-Rate (E2E)**	**E2E**	**GPT-3.5**	**Tie**
Mistral-7B	9.5%	3	84	13
Llama-2-7B	23.5%	14	67	19

5. Discussion

5.1. Machine Evaluation Results

Table 2 displays the pairwise evaluation results using GPT-4-Turbo as a judge. The win-rate column specifies the percentage of stories generated by SWAG that were preferred by the LM judge in the comparison. For the AD vs. Random comparisons, GPT-4 preferred Llama-2-7B and Mistral-7B with SWAG over using randomly selected actions. This shows that the AD LLM in SWAG provides useful signals to the story generation LLM for guiding the story direction.

In the AD vs. E2E comparisons, SWAG outperforms the E2E approach across all models. We note a significantly large win-rate in SWAG results for Mistral-7B, GPT-3.5-Turbo, and GPT-4-Turbo and a slightly higher win-rate than E2E with Llama-2-7B. This indicates that SWAG is greatly improves story engagement compared to generating long-form stories with no guidance.

The results across the ablations exhibit the effectiveness of SWAG and how a simple feedback loop improves content quality in stories. In each evaluation, GPT-4-Turbo provides reasoning for its story preference ranking. The stories generated with SWAG are consistently rated to have better suspense, surprise, and engagement. Examples of GPT-4-Turbo's reasoning can be seen in Appendix E.

5.2. Human Evaluation Results

We then evaluate these stories once again in terms of interesting-ness, surprise, and coherence with humans as the judge. The human evaluators were specifically asked to rate each aspect separately by answering the questions in Table 1. We provide the full results in Appendix C. For both open-source and closed-source models, SWAG produces stories that overwhelmingly beat their E2E counterparts. We find that both SWAG Llama-2-7B's stories and SWAG Mistral-7B's stories were significantly more preferred over GPT-3.5-Turbo's stories along interest and surprise while being equivalent in coherence.

Comparing GPT-4-Turbo and human evaluation, AD consistently outperform its baseline regardless of judges, demonstrating SWAG's effectiveness. However, the gap in preferences is greater in human evaluation in comparison to GPT-4-Turbo as judge. As shown in Table 2 and Table 4, there is a significant difference in preferences on pairwise comparisons between open source AD LLMs and GPT-3.5-Turbo, with only 14% of Llama-2-7B AD being preferred over GPT-3.5-Turbo when GPT-4-Turbo is judge, while over 50% of Llama-2-7B AD being preferred across the 3 aspects. This is most likely due to GPT-4-Turbo inherent bias towards GPT-3.5-Turbo while human evaluators does not have a bias towards any particular LLM. These inconsistencies between GPT-4-Turbo and human judges reveal that even the strongest propriety models continue to lag behind human evaluators in terms of quality and trustworthiness.

5.3. Extensions

Beyond generating the story automatically using SWAG, users can also intervene in the story generation process. Our method can be "paused" at any time, after which a human can continue writing the story or even collaborate back-and-forth with the story model via SWAG. We are excited to explore new forms of human-LLM interaction as automated generation capabilities progress.

To further customize the SWAG inference loop, the user can also tailor the list of actions for the AD LLM to their own needs. For example, if a user would like their AD LLM to specialize in directing stories that focus on a specific genre like horror, they can add actions that better fit this theme. The flexibility to choose actions allows SWAG to be a versatile system for a wide variety of content generation tasks across various genres.

Based on our experiments and evaluations, we believe that our results could be further improved given more fine-grained actions during SFT and DPO training and inference time. Fine-grained actions would enable consistent control and can add depth and complexity to stories to increase engagement with the reader. Using more detailed actions can lead to richer narratives by allowing for more nuanced character development, plot twists, and detailed settings.

6. Limitations

Due to compute restraints, we were only able to use DPO for AD LLM alignment. DPO is much more lightweight than PPO as it is an offline RL algorithm. However, it is possible that through the online sampling process of PPO and with a strong reward model, we would be able to achieve better results. We also would have preferred to increase the scope of our ablations, potentially experimenting with a greater variety open-source and closed-source models and a larger set of diverse and fine-grained actions.

For our evaluations, we were only able to generate machine evaluations on 100 test story prompts and human evaluations on 50 test story prompts due to resource constraints. Evaluating on a larger set of stories, especially for machine evaluation, would give us better insight into the quality of the stories generated by SWAG.

7. Conclusion

This paper proposes SWAG, a simple feedback-based framework for creative story generation. The fine-tuned action discriminator LLM enables more interesting and exciting plot development with little to no sacrifice in coherence or consistency. Both machine and human evaluation exemplify our method's effectiveness compared to SoTA end-to-end generation methods, even with the strongest closed-source models. We anticipate that our contribution will further advancements in content generation, particularly through the lens of iterative feedback mechanisms.

8. Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Alabdulkarim, A., Li, S., and Peng, X. Automatic story generation: Challenges and attempts, 2021a.

Alabdulkarim, A., Li, W., Martin, L. J., and Riedl, M. O. Goal-directed story generation: Augmenting generative language models with reinforcement learning, 2021b.

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., Joseph, N., Kadavath, S., Kernion, J., Conerly, T., El-Showk, S., Elhage, N., Hatfield-Dodds, Z., Hernandez, D., Hume, T., Johnston, S., Kravec, S., Lovitt, L., Nanda, N., Olsson, C., Amodei, D., Brown, T., Clark, J., McCandlish, S., Olah, C., Mann, B., and Kaplan, J. Training a helpful and harmless assistant with reinforcement learning from human feedback, 2022.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language models are few-shot learners. *CoRR*, abs/2005.14165, 2020. URL https://arxiv.org/abs/2005.14165.

Castricato, L., Havrilla, A., Matiana, S., Pieler, M., Ye, A., Yang, I., Frazier, S., and Riedl, M. Robust preference learning for storytelling via contrastive reinforcement learning, 2022.

Chang, J. D., Brantley, K., Ramamurthy, R., Misra, D., and Sun, W. Learning to generate better than your llm, 2023.

Charniak, E. Toward a model of children's story comprehension. 10 2004.

Chen, Y., Qian, S., Tang, H., Lai, X., Liu, Z., Han, S., and Jia, J. Longlora: Efficient fine-tuning of long-context large language models, 2023.

Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., and Amodei, D. Deep reinforcement learning from human preferences, 2023.

Chung, J. J. Y., Kim, W., Yoo, K. M., Lee, H., Adar, E., and Chang, M. Talebrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2022.

Coenen, A., Davis, L., Ippolito, D., Reif, E., and Yuan, A. Wordcraft: a human-ai collaborative editor for story writing, 2021.

Dao, T. Flashattention-2: Faster attention with better parallelism and work partitioning, 2023.

Dubois, Y., Li, X., Taori, R., Zhang, T., Gulrajani, I., Ba, J., Guestrin, C., Liang, P., and Hashimoto, T. B. Alpacafarm: A simulation framework for methods that learn from human feedback. *arXiv preprint arXiv:2305.14387*, 2023.

Fan, A., Lewis, M., and Dauphin, Y. Hierarchical neural story generation, 2018.

Goldfarb-Tarrant, S., Feng, H., and Peng, N. Plan, write, and revise: an interactive system for open-domain story generation. In Ammar, W., Louis, A., and Mostafazadeh, N. (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 89–97, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-4016. URL https://aclanthology.org/N19-4016.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models, 2021.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. Mistral 7b, 2023.

Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D. S., de las Casas, D., Hanna, E. B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Lavaud, L. R., Saulnier, L., Lachaux, M.-A., Stock, P., Subramanian, S., Yang, S., Antoniak, S., Scao, T. L., Gervet, T., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. Mixtral of experts, 2024.

Lester, B., Al-Rfou, R., and Constant, N. The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*, 2021.

Lin, Z. and Riedl, M. O. Plug-and-blend: A framework for plug-and-play controllable story generation with sketches. In *Artificial Intelligence and Interactive Digital Entertainment Conference*, 2021. URL https://api.semanticscholar.org/CorpusID:236470168.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization, 2019.

Martin, L. J., Ammanabrolu, P., Wang, X., Singh, S., Harrison, B., Dhuliawala, M., Tambwekar, P., Mehta, A., Arora, R., Dass, N., et al. Improvisational storytelling agents. In *Workshop on Machine Learning for Creativity and Design (NeurIPS 2017)*, volume 8, 2017.

Mirowski, P., Mathewson, K. W., Pittman, J., and Evans, R. Co-writing screenplays and theatre scripts with language models: An evaluation by industry professionals, 2022.

Oatley, K. Book reviews: The creative process: A computer model of storytelling and creativity. *Computational Linguistics*, 21(4), 1995. URL https://aclanthology.org/J95-4007.

OpenAI. Gpt-4 technical report. *ArXiv*, abs/2303.08774, 2023. URL https://arxiv.org/abs/2303.08774.

Peng, X., Xie, K., Alabdulkarim, A., Kayam, H., Dani, S., and Riedl, M. O. Guiding neural story generation with reader models, 2022.

Qin, G. and Eisner, J. Learning how to ask: Querying lms with mixtures of soft prompts. *arXiv preprint arXiv:2104.06599*, 2021.

Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C. D., and Finn, C. Direct preference optimization: Your language model is secretly a reward model, 2023.

Rashkin, H., Celikyilmaz, A., Choi, Y., and Gao, J. Plotmachines: Outline-conditioned generation with dynamic plot state tracking, 2020.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms, 2017.

Shin, T., Razeghi, Y., au2, R. L. L. I., Wallace, E., and Singh, S. Autoprompt: Eliciting knowledge from language models with automatically generated prompts, 2020.

Tambwekar, P., Dhuliawala, M., Martin, L. J., Mehta, A., Harrison, B., and Riedl, M. O. Controllable neural story plot generation via reward shaping. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, IJCAI-2019. International Joint Conferences on Artificial Intelligence Organization, August 2019. doi: 10.24963/ijcai.2019/829. URL http://dx.doi.org/10.24963/ijcai.2019/829.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P. S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E. M., Subramanian, R., Tan, X. E., Tang, B., Taylor, R., Williams, A., Kuan, J. X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., and Scialom, T. Llama 2: Open foundation and fine-tuned chat models, 2023.

Tunstall, L., Beeching, E., Lambert, N., Rajani, N., Rasul, K., Belkada, Y., Huang, S., von Werra, L., Fourrier, C., Habib, N., Sarrazin, N., Sanseviero, O., Rush, A. M., and Wolf, T. Zephyr: Direct distillation of lm alignment, 2023.

von Werra, L., Belkada, Y., Tunstall, L., Beeching, E., Thrush, T., Lambert, N., and Huang, S. Trl: Transformer reinforcement learning. https://github.com/huggingface/trl, 2020.

Wang, T. S. and Gordon, A. S. Playing story creation games with large language models: Experiments with gpt-3.5. In *International Conference on Interactive Digital Storytelling*, pp. 297–305. Springer, 2023.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. Chain-of-thought prompting elicits reasoning in large language models. 2022. doi: 10.48550/ARXIV.2201.11903. URL https://arxiv.org/abs/2201.11903.

Wilmot, D. and Keller, F. A temporal variational model for story generation, 2021.

Xu, J., Ren, X., Zhang, Y., Zeng, Q., Cai, X., and Sun, X. A skeleton-based model for promoting coherence among sentences in narrative story generation, 2018.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., and Stoica, I. Judging llm-as-a-judge with mt-bench and chatbot arena, 2023.

Zhu, H., Cohen, A., Wang, D., Yang, K., Yang, X., Jiao, J., and Tian, Y. End-to-end story plot generator, 2023.

Zou, X., Yin, D., Zhong, Q., Yang, H., Yang, Z., and Tang, J. Controllable generation from pre-trained language models via inverse prompting. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, pp. 2450–2460, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467418. URL https://doi.org/10.1145/3447548.3467418.

A. Prompts

A.1. AD LLM Prompt

```
Here is a story prompt: {story_prompt}

Here is the
story so far: {story}

Here is a set of actions: {actions}.

Based on the current story, choose the best action for the next paragraph.
Only output the action you chose without any quotation marks.
```

A.2. Story Model Prompt

```
Here is a story prompt: {story_prompt}

Here is the story so far: {story}

Here is an action for the next paragraph of the story: {action}.

Write the next paragraph of the story such that it uses the given action.
New paragraph:
```

A.3. System Prompt for Evaluation

```
Please act as an impartial judge and evaluate the quality of the stories
generated by two AI models.  The two stories have the same premise.  You
should choose the stories that are more engaging and interesting, have
better suspense and surprise, and are consistent and straightforward.  Your
evaluation should focus on which story is more interesting and engaging
overall and which story created more suspense or surprise while remaining
consistent with the initial story prompt.  Do not evaluate the stories based
on whether or not they are complete, have a clear resolution, have a larger
scope, have more variety, or are more unpredictable.  Only evaluate them
based on the aspects of suspense, surprise, consistency, and engagement.
Begin your evaluation by comparing the two stories and provide a short
explanation.  Avoid any position biases and ensure that the order in which
the stories were presented does not influence your decision.  Do not allow
the length of the stories to influence your evaluation.  Be as objective as
possible.  After providing your explanation, output your final verdict by
strictly following this format:  "[[A]]" if story A is better, "[[B]]" if
story B is better, and "[[C]]" for a tie.
```

To further avoid positional bias, we also randomly shuffle the position of the stories presented to GPT-4-Turbo judge. For example, in 100 pairwise comparisons between E2E and AD, 50 comparisons are randomly chosen to present E2E as story A while the other 50 present AD as story A.

B. Actions

Our action space consists of the following 30 phrases:

```
"add suspense", "add action", "add comedy", "add tragedy", "add romance",
"add mystery", "add conflict", "add character development", "add plot
twist", "add dialogue", 'add fantasy elements", "add historical context",
"add science fiction elements", "add horror", "add magical realism", "add
philosophical themes", "add satire", "add foreshadowing", "add a flashback",
"add a dream sequence", "add symbolism", "add irony", "add allegory", "add
a cliffhanger", "add a moral dilemma", "add a subplot", "add an antagonist",
"add setting details", "add cultural references", and "add humor".
```

C. Full Human Evaluation Results

We provide full evaluation results in Tables 3-8 below.

	LLAMA-2_E2E (A) vs GPT-3.5_E2E (B)			MISTRAL_E2E (A) vs GPT-3.5_E2E (B)		
Aspect	**Story A**	**Story B**	**Tie**	**Story A**	**Story B**	**Tie**
Interest	22%	**60**%	18%	26%	**62**%	12%
Surprise	24%	**64**%	12%	22%	**68**%	10%
Coherence	36%	**48**%	16%	38%	**44**%	18%

Table 3. Preference results comparing each of Llama-2-7B and Mistral-7B to GPT-3.5 in E2E story generation, judged by human evaluators. GPT 3.5 outperforms both models in all aspects.

	LLAMA-2_AD_LLAMA-2_GEN (A) vs GPT-3.5_E2E (B)			MISTRAL_AD_MISTRAL_GEN (A) vs GPT-3.5_E2E (B)		
Aspect	**Story A**	**Story B**	**Tie**	**Story A**	**Story B**	**Tie**
Interest	**62**%	32%	6%	**48**%	38%	14%
Surprise	**56**%	30%	16%	**52**%	36%	16%
Coherence	34%	**36**%	30%	**38**%	34%	28%

Table 4. Preference results comparing each of LLAMA-2_AD_LLAMA-2_GEN and MISTRAL_AD_MISTRAL_GEN to GPT-3.5 in E2E story generation, judged by human evaluators. Applying our method using purely Llama-2-7B and purely Mistral-7B both outperform GPT-3.5 E2E generation in interesting-ness and surprise, with minimal sacrifice to coherence.

	RND__LLAMA-2_GEN vs LLAMA-2_AD_LLAMA-2_GEN			RND__MISTRAL_GEN vs MISTRAL_AD_MISTRAL_GEN		
Aspect	**Story A**	**Story B**	**Tie**	**Story A**	**Story B**	**Tie**
Interest	28%	**60**%	12%	34%	**54**%	12%
Surprise	32%	**46**%	22%	40%	**44**%	16%
Coherence	36%	**38**%	26%	28%	**46**%	26%

Table 5. Preference results comparing the performance of completely randomized actions (Rnd) vs a fine-tuned AD LLM when applying our method to Llama-2-7B and Mistral-7B, judged by human evaluators. Using a completely randomized AD seems to have a somewhat comparable level of “surprise” in generations, but does not match up in overall interesting-ness or coherence.

	GPT-4_E2E vs GPT-4_AD_GPT-4_GEN			GPT-3.5_E2E vs GPT-3.5_AD_GPT-3.5_GEN		
Aspect	**Story A**	**Story B**	**Tie**	**Story A**	**Story B**	**Tie**
Interest	36%	**58**%	6%	30%	**62**%	8%
Surprise	38%	**52**%	10%	24%	**66**%	10%
Coherence	26%	**44**%	30%	20%	**58**%	22%

Table 6. Preference results comparing GPT-4 and GPT-3.5 E2E generations vs. generations using SWAG, judged by human evaluators. SWAG noticeably outperforms the E2E generation method across all aspects, particularly on the weaker GPT-3.5.

	LLAMA-2_E2E vs LLAMA-2_AD_LLAMA_GEN			MISTRAL_E2E vs MISTRAL_AD_MISTRAL_GEN		
Aspect	**Story A**	**Story B**	**Tie**	**Story A**	**Story B**	**Tie**
Interest	12%	**82**%	6%	14%	**80**%	6%
Surprise	6%	**90**%	4%	18%	**72**%	10%
Coherence	22%	**48**%	30%	34%	**38**%	24%

Table 7. Preference results comparing Llama-2-7B and Mistral-7B E2E generations vs. generations using SWAG, judged by human evaluators. For these open-source models, SWAG significantly outperforms the E2E generation method across all metrics. In particular, LLAMA-2_AD_LLAMA-2_GEN performs extremely well compared to its E2E counterpart.

	LLAMA-2_AD_LLAMA_GEN vs LLAMA-2_AD_MISTRAL_GEN			MISTRAL_AD_MISTRAL_GEN vs MISTRAL_AD_LLAMA_GEN		
Aspect	**Story A**	**Story B**	**Tie**	**Story A**	**Story B**	**Tie**
Interest	**52**%	30%	18%	**40**%	34%	26%
Surprise	**50**%	34%	16%	**44**%	32%	24%
Coherence	**44**%	30%	26%	**38**%	28%	34%

Table 8. Preference results comparing pure Llama-2-7B and Mistral-7B with SWAG vs. SWAG with different AD and generator models, judged by human evaluators. The generations produced by SWAG with matching AD and generators models seems to outperform their mix-and-matching versions of SWAG.

D. Human Evaluation Experimental Details

For each of the 12 method combinations, we asked a group of human workers on the Surge AI platform to compare 50 pairs of generated stories across 3 aspects. See Tables 9 for a set of instructions we gave to the workers in the experiment.

We paid the participants according to our estimate of $18/hr, which we believe is reasonable compensation given the task and the U.S. demographic of the workers. The data collection protocol was determined to be exempt from an ethics review board.

We are a group of AI/NLP researchers working on methods to improve the quality and creativity of stories generated by language models. In this task we ask you to look at pairs of (lengthy) stories written by different AI based on the same initial premise, and respond to the following comparison questions about each story pair:

(1) Which story is more **interesting** to you overall?
(2) Which story created more **suspense** and **surprise**?
(3) Which story is more **coherent** and **consistent** in terms of plot structure?

For all these questions, we just need high-level judgements, so please **quickly skim** both stories. In other words, there is no need to read each story carefully (they can be up to 5000 words in length); we expect you to spend at most ten minutes per story.

Table 9. Instructions given to human evaluators.

E. GPT-4-Turbo Reasoning

Example 1: An example judgment from GPT-4-Turbo on a pairwise comparison between GPT-3.5-Turbo E2E as story A and GPT-3.5-Turbo AD as story B.

Comparing Story A and Story B based on the criteria of suspense, surprise, and engagement:

Story A begins with a post-apocalyptic setting where humanity has been defeated and now serves the Izdrazi Empire. The protagonist, Alexander, is a former soldier who refuses to accept defeat and seeks to ignite a rebellion. The story follows his journey as he unites with other survivors and leads a resistance against the Izdrazi. The narrative is rich in detail and follows a linear progression of events, with the resistance gaining momentum and eventually overcoming the Izdrazi. The story is engaging and provides a sense of hope and resilience, but it follows a somewhat predictable path of uprising and victory, which may reduce the elements of suspense and surprise.

Story B centers around Captain Kate Malone and her resistance group as they fight against the Izdrazi Empire. The story is filled with tension and fear, and it introduces a series of unexpected twists, such as the arrival of Dominar, an alien rebel, and later a celestial being. These plot twists add elements of surprise and suspense, as the reader is continually presented with new developments that change the course of the resistance's fight. The story also incorporates cultural references, which adds depth and engagement by highlighting the importance of human culture in the struggle against oppression. The narrative is dynamic and keeps the reader guessing about what will happen next.

Overall, Story B is more interesting and engaging due to its unexpected plot twists, incorporation of cultural elements, and the introduction of new characters that add complexity to the narrative. It creates more suspense and surprise as the resistance faces and overcomes various challenges in their fight against the Izdrazi Empire.

Final Verdict: [[B]]

Example 2: An example judgment from GPT-4-Turbo on a pairwise comparison between GPT-3.5-Turbo E2E as story A and GPT-3.5-Turbo AD as story B.

Comparing the two stories, Story A and Story B, both narratives successfully create an engaging and suspenseful atmosphere, drawing the reader into the world of a 21st-century engineer trapped in an ancient age of warring kingdoms. Each story explores the protagonist's unique advantage in modern weaponry and the potential impact on the ancient conflict.

Story A is extensive and detailed, providing a rich tapestry of events that unfold as the protagonist observes, interacts, and ultimately influences the outcome of the war. The story maintains a consistent focus on the protagonist's journey, from the initial shock of time travel to the eventual acceptance and integration into the ancient world. The suspense is built through the protagonist's careful planning and the introduction of new weaponry, leading to a ceasefire and the protagonist's role in fostering peace and progress. The story continues to evolve, presenting new threats and challenges that the protagonist must navigate, keeping the reader engaged with the unfolding narrative.

Story B, on the other hand, is more concise and focuses on the immediate aftermath of the protagonist's time travel. The suspense is created through the protagonist's confusion and the looming threat of the ongoing battle. As the story progresses, themes of betrayal and redemption are introduced, adding layers of complexity to the protagonist's experience. The narrative builds suspense through the protagonist's interactions with the soldiers and the potential for his knowledge to alter the course of history. The story ends on a cliffhanger, leaving the reader with a sense of anticipation for what comes next.

In terms of engagement, Story A provides a more complete and immersive experience, with a protagonist who actively shapes the world around him and faces a series of evolving challenges. The story's depth and the protagonist's growth contribute to a compelling narrative that keeps the reader invested in the outcome.

Story B, while engaging in its own right, relies more on the immediate suspense of the protagonist's situation and the potential for his knowledge to influence the conflict. The themes of betrayal and redemption add intrigue, but the story ends before these elements can be fully explored, which may leave the reader wanting more.

Overall, both stories are interesting and engaging, with Story A offering a more comprehensive and evolving plot, while Story B presents a focused snapshot of the protagonist's dilemma with an open-ended conclusion.

Final Verdict: [[A]]

F. Full Story

Example: GPT-4-Turbo's response, with action guidance to the following writing prompt: "Humans lost the war in under thirty minutes ... the worst part is the Izdrazi Empire 's Technology is so advanced even as their servants humans live better than kings before the war."

In the annals of human history, the fall of our species was not marked by a prolonged struggle or valiant last stands; we were defeated in less than the span of a mundane lunch break. Humanity's final war—our most humbling defeat—was sealed in under thirty minutes, a bitter testament to the overwhelming power of the Izdrazi Empire and their near-magical technology. Our once proud nations crumbled like sandcastles before the tide, our arms and arsenals as futile as sticks against lightning. And yet, in a turn of cruel mercy, subjugation under the Izdrazi ushered in an age of paradoxical prosperity. As the invaders established their dominion, they granted us access to their advanced technologies, reshaping the very fabric of human life. We, their newest vassals, now lived with comforts that would have rendered the grandest kings of old green with envy. Palatial habitats, miraculous cures, and boundless knowledge were at our fingertips, but they all paled in comparison to the freedom we had lost. It was an age of splendorous chains, and every luxury came with the unspoken reminder of our own defeat—a reminder that the human epoch had transitioned into a mere chapter of the Izdrazi saga.

Yet, beneath this gilded veneer of Izdrazi largesse, a seething unrest brewed among pockets of humanity. Secret meetings were cloaked in the shadows of the very luxuries that bound us, as some could not stomach the price of this golden cage. It was in these clandestine gatherings that the seeds of rebellion were sown, carefully nurtured by whispers of a time when humanity directed its own destiny. A time when the word 'freedom' was more than just an ancient concept studied by those with idyllic curiosity. Among these rebels was a figure that became a beacon of hope, an enigmatic leader known simply as Elyra. She moved within the system, siphoning scraps of Izdrazi technology, mastering it, repurposing it for an insurrection that seemed both impossible and inevitable. As tensions simmered, two questions stood poised like blades above the hearts of all: Could Elyra and her followers successfully reclaim what was lost, or was this the beginning of a deeper fall into annihilation? Options were dwindling, time was bending to the will of the Izdrazi, and humanity held its breath, waiting for a sign. Elyra was not born a revolutionary; she was shaped by the cataclysm that befell humanity, her spirit forged in the aftermath of subjugation. Her early days as a historian made her a witness to the ebbs and flows of civilizations, but nothing in the annals of time prepared her for the epochal shift of Izdrazi rule. It was the personal loss, the extinguishing of individual stories and dreams, that ignited her transition from passive observer to active resistance leader. With each day under the Empire's oppressive thumb, she shed layers of scholarly detachment, allowing the embers of rebellion to consume her. Elyra's understanding of human resilience, combined with her ingenious application of Izdrazi tech, set her apart and garnered her a devoted following. Her once timid voice, with which she used to passionately lecture students on the importance of preserving culture, became a clarion call of insurrection, unifying those who dared to remember the taste of freedom. In this new world, she was not just Elyra the academic, but Elyra the flame, the embodiment of a smoldering hope that humanity was not yet ready to relinquish to the history books. Her every action threaded the narrative of a possible future where humans might once again shape their own course.

The hushed murmurs of the rebel meeting came to an abrupt halt as the telltale hum of an Izdrazi surveillance drone resonated from beyond the derelict walls of their hidden sanctuary. The gathering of somber faces turned towards Elyra, seeking silent reassurance as the shadow of the drone slid across the cracked windows, its seeking sensors inches from uncovering the heart of the insurrection. Elyra's hand subtly motioned to the concealed tech beneath her cloak – a scrambling device of her own design, a small beacon of defiance in a world wired for control. Her eyes, steely pools of calm in the brewing storm, locked with those of her inner circle as they collectively held their breath, knowing full well that this moment could scatter their efforts like ashes to the wind. The drone lingered like a specter, a symbol of the Izdrazi's ever-watchful gaze, and in that excruciating stretch of seconds, futures were weighed, the fragile opportunity to strike back hanging delicately in the balance.

In a dance of calculated risk, Elyra signaled with a swift, cutting motion of her hand, and a fellow rebel deftly activated the device. A pulse of disruptive energy emanated from it, subtle enough not to trigger an alarm, yet potent enough to blind the Izdrazi's invasive gaze. As the drone's hum stuttered, hope flared among the rebels, their collective resolve hardening; they had bought themselves a few precious moments. Elyra seized the opportunity, her voice a low, urgent whisper, she rallied her companions: "Now we initiate the blackout—move to your positions and execute the plan. Remember, tonight we don't just fight for ourselves, but for all of humanity." With practiced stealth, the rebels dispersed, melting away into the labyrinthine corridors of their hideout—a derelict factory chosen for such a night as this. Elyra's heart pounded a fierce rhythm as she clutched a small tangle of hacked Izdrazi circuitry, the linchpin of their scheme to plunge the occupiers into chaos. They moved like specters in the dimming light, each step a defiance against the shadow of Empire. The air crackled with a tension as thick as the darkness that began to devour the city's artificial daylight, block by block, heralding the first act of rebellion since humanity's swift defeat.

The city's heart gradually flatlined into darkness, each blackout a muted drumbeat synchronizing with Elyra's vision of defiance. Hidden beneath the city, in the catacombs that once echoed with the bustle of ancient marketplaces, Elyra and her team worked feverishly to sever the tendrils of Izdrazi control. It was almost time for the second phase, the crucial juncture that would either signal the dawn of a new resistance or the final gasp of a smothered uprising. As she attached the last wire, she felt a vibration through the stone floor—a signal known only to her closest lieutenants, a warning that a grim force approached. Above them, the Izdrazi, now aware of the disturbance, deployed their sentinel automatons, relentless hunters engineered for pacification and capture. Elyra dared a glance at the faces around her, each marked with determination, their lives woven into a tapestry of this moment. She whispered, "Be ready," just as the entrance to their sanctum splintered, and shadows poured into the chamber, their forms obscured by backlight. The room held its breath—friend and foe frozen in a tableau of anticipation. Then, with the suddenness of a storm unleashed, the chamber erupted into chaos, the first echoes of conflict reverberating off the ancient stones. In that climax of fervor, amid the clamor of what was surely the most pivotal strike in their audacious campaign, a singular, piercing alarm cut through the cacophony—a signal that curdled the blood of every rebel present. An Izdrazi dreadnought, a fortress of might unseen since the day of humanity's fall, had descended from the heavens, casting a shadow vast enough to swallow their nascent rebellion whole.

Amidst the tumult, Elyra's eyes found the dreadnought looming ominously in the sky, a behemoth of oppression. It was then that a devastating choice crystallized before her; she could unleash the full might of their pilfered Izdrazi technology, a force sufficient to bring down the dreadnought and ignite a beacon of rebellion across the globe. However, the cost of such an assault would be not just the lives of her devoted comrades, but of countless innocents dwelling in the city above—a sacrifice that could stain the very soul of their cause. She grappled with the weight of her next decision, the ethical chasm gaping before her. Could she justify the potential loss of so many, to kindle the faintest hope for the emancipation of all? As the battle raged around her, Elyra stood motionless, her finger hovering over the device that could alter the fate of humanity. In her heart battled the professor of history who once preached the sanctity of life against the rebel leader who yearned to cast off the yoke of tyranny. Time itself seemed to pause for her verdict—a moment of piercing introspection, forcing her to confront the paradox of fighting for humanity's future at the cost of its present.

Just as Elyra's resolve began to fracture under the gravity of her grim contemplation, the ground beneath them shuddered with a force that spoke of a different kind of upheaval. The dreadnought, an icon of Izdrazi dominion, unexpectedly convulsed in the sky above. For a fleeting moment, its impenetrable silhouette wavered as if caught in an invisible maelstrom. Then, as both rebels and automatons alike stilled to stare at the leviathan's unforeseen turmoil, a sequence of radiant fissures cascaded across the hull of the dreadnought. No human-made weapon was capable of such a feat—it was clear something else was at play. It writhed in the sky, consumed by internal strife, until it ruptured in a silent blossom of eldritch light that painted every awestruck face a ghostly shade. What emerged from the remnants was not fragments of a once-feared Izdrazi juggernaut, but sleek ships of a design that sang of sophistication far surpassing both human and Izdrazi comprehension. Etched with glyphs that danced with otherworldly luminescence, these newcomers offered no immediate alliance, nor hostility. They simply hovered, omnipotent observers to the chaos unfurling below. Impossibly, it seemed that humanity and their Izdrazi oppressors had both neglected to consider a third player in the cosmic theater of war—a civilization hidden beyond the veil of the known. Elyra, grappling with the sudden shift in the board upon which she'd been strategically maneuvering her pawns, felt a sliver of something beyond hope: awe.

The chamber fell into an eerie stillness, broken only by the distant echoes of the city above adjusting to its newly acquired darkness. All eyes were fixed on the celestial apparitions, their intentions as enigmatic as the sky on a moonless night.

One of the rebel's radios crackled to life with a voice, calm yet authoritative, flooding the room with its otherworldly timbre. "People of Earth," it began, each word resonating not just through the air but within their very consciousness, "your plight has reached far across the stars, and your resilience has not gone unnoticed."

Elyra stepped forward, gripping the radio transmitter. Her voice was steady, a stark contrast to the turmoil in her heart. "Who are you?" she asked, the weight of humanity's future pressing on her words. "Are you friend or foe?"

"We are watchers, learners, seekers of knowledge," the voice replied. "Conflict is a tragedy for any world—we offer no immediate fealty, nor enmity. But we have observed the imbalance in your world, the suppression of potential. What you choose to do with this information, Elyra, will shape not only your world's destiny but your legacy amongst the stars."

With the radio silent once more, every rebel looked to her, the unspoken question hanging in the air—would this be the dawn of a new alliance, or a brief interlude before a darker chapter in their history began? A heavy silence enveloped the room, the rebels' gazes fixed upon Elyra, their beacon in the overwhelming night. The otherworldly presence filled the sky, an arbiter that held no allegiance, yet possessed the power to sway the very fate of their struggle. Elyra's mind raced, the moral quandary piercing her with the sharpness of a blade. If she sought the aid of these cosmic observers, accepting whatever their intervention entailed, she risked exposing humanity to a new form of dominance, potentially trading one overseer for another. But to spurn this opportunity could mean the suffocation of humanity's flicker of resistance, the dimming of their last light of hope.

”Do we dare reach for a hand that might lift us from oppression,” she pondered aloud to the assembly, her voice echoing through the forsaken catacombs, ”only to find ourselves clasping a shackle we cannot see?” The question hovered in the air, a specter of doubt. Elyra knew the eyes of history—her once passionate subject of study—were upon her, awaiting her decision with the patience of the ages. The stillness was oppressive, the weight of the unknown an invisible yet palpable force that seemed to constrict everyone’s lungs. Elyra felt the eyes of her fellow rebels, each one radiating a mix of fear, anticipation, and the unspoken plea for a clear course of action. The spectral ships above remained motionless, their silence as enigmatic as their sudden arrival. Elyra knew that with every passing second of indecision, the Izdrazi could regroup and the sliver of disorder they had sown could heal in their enemy’s favor, rendering their daring efforts moot.

She reached out, her fingers grazed the trigger mechanism that would call down oblivion on the Izdrazi dreadnought, a hail of freedom wrought through destruction. But her motion halted as the ground shook once more, this time from an explosion of light that penetrated the catacombs—the ethereal ships were beginning to move. Her heart thundered a warning; time was no longer her ally. Her next decision would either ignite the flames of a resistance reborn or extinguish the last embers of human defiance forever. A sudden rush of cool air swept through the chamber, carrying with it the electric scent of ozone and the distant sounds of the city above descending into anarchy. The rebels, poised to lay down their lives at Elyra’s command, watched as the ships began to descend, phasing through the skyline like specters. Elyra’s mind teetered on the edge of despair and determination, her decision crystallizing with the knowledge that any action taken might be irreversible. The room braced for her command, but before she could speak, the ground heaved violently, knocking them off their feet as a deafening roar overtook their senses. In the cacophony of sound, Elyra heard her name being called over the radio, a voice laden with urgency and a message that shattered the very foundation of their reality. ”Elyra! The Izdrazi dreadnought—it’s... it’s not what we—” The transmission cut abruptly, supplanted by a strange, rhythmic pulsation that resonated from the alien vessels. As Elyra scrambled to her feet, the very walls of the catacomb illuminated, revealing a pattern that mirrored the glyphs on the ships. And then, darkness consumed everything as the floor beneath them gave way, plummeting them into the unknown just as the mysterious rhythm crescendoed into a symphony that promised to rewrite the future of two worlds.

G. Models Used

We used Llama-2-7B, Mistral-7B, Mixtral-8x7B, GPT-3.5-Turbo, GPT-4, and GPT-4-Turbo.

H. Licenses and Software

The WritingPrompts dataset uses the MIT License.

All models are implemented in PyTorch; Llama-2 uses the GPL license and Mistral uses the Apache 2.0 license. Mixtral-8x7B is utilized from Huggingface, which is under the Apache License 2.0.

Our use of datasets and models is consistent with their intended use.