

Guiding and Diversifying LLM-Based Story Generation via Answer Set Programming

Phoebe J. Wang¹ and Max Kreminski^{1,2}

¹Santa Clara University, ²Midjourney
wjingsen@gmail.com, mkreminski@midjourney.com

Abstract

Instruction-tuned large language models (LLMs) are capable of generating stories in response to open-ended user requests, but the resulting stories tend to be limited in their diversity. Older, symbolic approaches to story generation (such as planning) can generate substantially more diverse plot outlines, but are limited to producing stories that recombine a fixed set of hand-engineered character action templates. Can we combine the strengths of these approaches while mitigating their weaknesses? We propose to do so by using a higher-level and more abstract symbolic specification of high-level story structure—implemented via answer set programming (ASP)—to guide and diversify LLM-based story generation. Via semantic similarity analysis, we demonstrate that our approach produces more diverse stories than an unguided LLM, and via code excerpts, we demonstrate the improved compactness and flexibility of ASP-based outline generation over full-fledged narrative planning.

1 Introduction

Story generation research today has two major prongs: one that aims to use symbolic or rules-based AI methods (such as planning) to generate plot outlines via hand-crafted bases of story structure knowledge, and one that aims to produce fully realized textual stories via large language models (LLMs). Both approaches have significant strengths and weaknesses. Planners are good at reasoning about long-term story structure and ensuring plot coherence (Riedl and Young, 2010; Cardona-Rivera et al., 2024), but narrative planning domains are notoriously effortful to hand-engineer (Giannatos et al., 2012; Siler, 2022), and their fine-grained encoding of possible character actions and plot elements make them rigid in their adherence to a particular, narrow *type* of story.

LLMs, on the other hand, are effective at open-domain story generation. However, they tend to struggle with long-term story coherence, often failing to produce a satisfying overall story arc or reader experience (Kreminski and Martens, 2022). Additionally, comparing LLM-generated to human-written stories reveals many aesthetic weaknesses (Chakrabarty et al., 2024; Mirowski et al., 2023), in particular a tendency toward unoriginality or *homogeneity* in both the semantic and syntactic aspects of generated stories (Anderson et al., 2024; Begus, 2023).

The clear complementarity of symbolic and neural approaches to story generation has led many researchers to try knitting these approaches together in various ways (Martin et al., 2018; Simon and Muise, 2022; Kelly et al., 2023; Wang et al., 2024). Nevertheless, the rigidity of narrative planning has made it difficult to apply planners to open-domain story generation, even as “plan-and-write” approaches to story generation (Yao et al., 2019) have grown in popularity on the LLM side.

In this paper, we propose a slight variation on the plan-and-write approach—i.e., first generating a story outline and then expanding that outline into a complete story via LLM. Unlike past approaches, we use neither a language model (Fan et al., 2018; Yao et al., 2019; Goldfarb-Tarrant et al., 2020) nor a fine-grained symbolic planning domain (Kelly et al., 2023; Wang et al., 2024) to generate story outlines. Instead, we use a lightweight microtheory of storytelling goal sequencing encoded as an *answer set program*. This enables us to combine the open-domain storytelling capabilities of an LLM with the high-level structural and diversity benefits of a symbolic approach to outline generation. We demonstrate the potential compactness and flexibility of an ASP-based outline generator through code excerpts, and the diversity benefits of our approach through a quantitative analysis of semantic similarities between sets of generated stories.

2 Our Approach

Broadly speaking, we generate stories via a two-step process. First, we **plan** a large and diverse set of potential story outlines by generating the full set of solutions to a small handwritten answer set program that encodes many viable sequences of storytelling goals (Fig. 1). Second, we choose an outline at random from this set and **write** a complete textual story that follows this outline via LLM, conditioning generation on a high-level story premise defined by a short string of text (Fig. 2).

2.1 Plan: Outline Pre-Generation with ASP

In the **plan** phase, we make use of answer set programming (ASP) to generate diverse story outlines. ASP is a form of constraint solving that employs an *answer set solver* to generate all *stable models* (Gelfond and Lifschitz, 1988), or sets of facts, that satisfy a set of (generally hand-written) logical constraints (Brewka et al., 2011).

ASP is widely used in procedural content generation research to generate artifacts ranging from story outlines to game levels (Smith and Mateas, 2011). Generally speaking, ASP-for-PCG approaches break down a generative answer set program into two components; the first part defines the basic *elements* that can be combined to form valid artifacts, and the second part defines *constraints* that restrict how these elements are allowed to be combined. In our case, we model story outlines as consisting of a sequence of *scenes*, each of which describes what happens at a certain time point in the story, and each of which fulfills some *storytelling goal* or *narrative function* related to what the story’s author is trying to convey.

Timeline and Scenes We divide the story timeline into a sequence of scenes, for example, seven scenes, which we found to be long enough to create an interesting narrative. In ASP, this is written as:

```
#const num_scenes = 7.  
scene(1..num_scenes).
```

Our current implementation ultimately translates each scene to a single paragraph of story text, but scenes could also be rendered as shorter or longer spans of text; as sequences of panels in a comic; as clips of video making up a movie; and so on.

Narrative Functions Narrative functions are the atomic units of story outlines, representing what storytelling goal the system is trying to accomplish

in each scene. In our current implementation, each scene performs exactly one narrative function:

```
{ scene_performs_func(S,F)  
  : function(F) } = 1 :- scene(S).
```

So far we have used a relatively minimal set of example narrative functions to test our approach:

```
function(intro_char;intro_rival_char;  
  add_conflict;add_bonding;  
  add_obstacle_towards_major_goal;  
  add_obstacle;add_breakthrough;  
  add_twist;level_up_obstacle).
```

However, we do not intend for this ontology of storytelling actions to be treated as canonical, and we hope that it is revised or replaced with a more rigorously grounded ontology in the future. Future work, for instance, might introduce functions related to the *tone* or *mood* of narration, which could be assigned to scenes in parallel with the plot development functions we have already introduced.

Function Parameters Some narrative functions can take additional *parameters* that help to further specify a high-level storytelling goal (e.g., introducing a character). In our example outline generator, we use both *personality parameters* to add diversity to character introductions and *obstacle parameters* to add diversity to introduced obstacles. In our experience, these parameters are best introduced when the LLM tends to always satisfy a particular narrative function in very similar ways unless given further guidance.

Constraints To prevent the generation of incoherent outlines, we use constraints to rule out certain ways of combining narrative functions. The expressiveness of ASP allows these constraints to be concisely defined. For instance, this constraint prohibits generation of outlines in which two scenes in a row perform the same narrative function:

```
:- scene_performs_func(S1,F),  
   scene_performs_func(S2,F),  
   S2 = S1 + 1.
```

And this constraint ensures that we never attempt to add conflict between characters until at least two characters have been introduced:

```
:- scene_performs_func(S2,add_conflict),  
   0 { scene_performs_func(S1,intro_char)  
     : scene(S1), S1 < S2 } 1.
```

In general, constraints are used to specify invalid *partial orderings* of narrative functions.

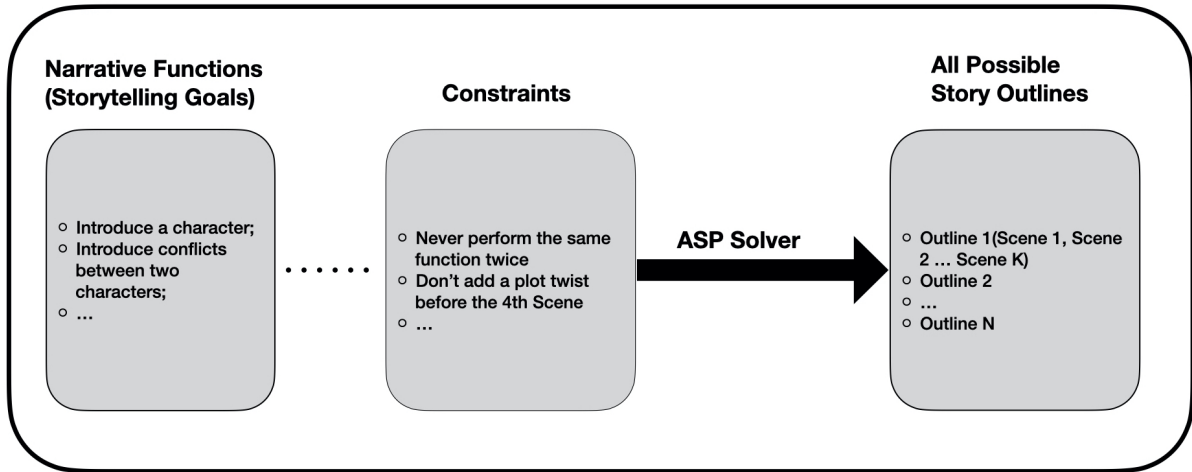


Figure 1: To generate diverse story outlines via ASP, we first specify a set of high-level *narrative functions* or *storytelling goals* that each scene in a story might fulfill. We then define a set of *constraints* that restrict how these narrative functions are allowed to be sequenced and combined. Finally, using an answer set solver, we solve for the set of all valid story outlines that can be produced given the specified narrative functions and constraints.

To maximize the diversity of generated outlines, we intentionally specify only a minimal set of constraints. Our example outline generator makes use of just fifteen total constraints, each consisting of no more than three lines of ASP code. Because story outlines are ultimately translated into complete stories by an LLM, ASP constraints do not necessarily have to prevent every problem that might arise in narrative function sequencing: they only need to prevent problems that the LLM is unable to acceptably “write around”, which can be identified through iterative testing and refinement of the full ASP-to-LLM story generation pipeline.

Though we do not do so here, constraints could also be used to restrict outline generation according to user preferences. For example, a constraint might be used to force generation of outlines that contain at least a certain amount of bonding between characters. An interactive story generation tool based on our approach might even provide a user interface for enabling and disabling various optional constraints.

Implementation We implement ASP functionality using the `clingo` library (Gebser et al., 2011). Having defined an answer set program that describes the design space of valid story outlines, we run the answer set solver to generate all possible outlines that satisfy the given rules. We then save these outlines to a file for later use. Our initial evaluation involved the pregeneration of a set of 401,990 distinct outlines.

2.2 Write: Story Generation with LLM

In the **write** phase, we use an LLM to expand an outline generated via ASP to a complete written story. This phase takes two key inputs:

1. User input: a short textual *premise prompt* that describes the high-level premise of the story to be generated.
2. Outline: a randomly selected story outline from the pre-generated outline pool. In our current implementation, an outline consists of a sequence of seven narrative functions, one function for each scene. Each scene will be written out as one paragraph of story. An example outline looks like this:

```
[‘introduce_character:sunny’,
‘add_obstacle:supernatural’,
‘introduce_character:mysterious’,
‘add_twist’,
‘introduce_character:clumsy’,
‘introduce_rival_character’,
‘level_up_obstacle’]
```

2.2.1 Translate Narrative Functions to Writing Instructions

We use a dictionary to map each narrative function in the chosen outline to a detailed, human-authored *writing instruction* for the LLM to follow.

For example, for the narrative function `introduce_character:sunny`, the resulting instruction is: “*introduce an optimistic and brave*

character who would fit organically. Consider how this character’s presence can be connected to the current plot, setting, or other characters. Craft a scene or situation that allows for a smooth integration of this character, ensuring that their introduction doesn’t feel forced or disconnected from the narrative. Provide context for their sunny and brave nature through their actions, dialogue, or the reactions of other characters, while maintaining the flow and consistency of the story. This character’s introduction should contribute to the development of the plot or the growth of other characters in a meaningful way, while maintaining the flow and consistency of the story”.

2.2.2 Generate Story Using Prompts

We use the gpt-3.5-turbo LLM, via the OpenAI API, to render a selected outline into a full story. A consistent system prompt is used to set the context for the LLM, instructing it to act as a fiction writer.

For the first paragraph of the generated story, the model is prompted with the story premise as well as the appropriate writing instruction for the first narrative function in the outline: “*You’re writing a story about: {premise}. Write the first paragraph of the story. In this paragraph, {instruction}.*” The resulting paragraph is then appended to the conversation history.

For subsequent paragraphs, the same process is repeated, using a slightly modified story continuation prompt: “*Write the next paragraph of the story. Remember the story is about: {premise}. In this paragraph, {instruction}.*” The complete conversation history is fed into subsequent paragraph generation requests as additional context, and the newly generated paragraph is then appended to the conversation history. This process is repeated until every entry in the story outline has been expanded to a full paragraph of text.

2.3 Baseline: Unguided Generation

In addition to an ASP-guided LLM-based story generator, we also implement an unguided LLM-based generator as a baseline for evaluation. This baseline generator follows a nearly identical process to that described above, with one key difference: the paragraph generation prompts issued to the baseline generator do not include the translated instructions from the outline. Instead, they rely solely on the user-supplied premise to guide story generation.

3 Evaluating Story Diversity

3.1 Procedure

We evaluate our approach by generating parallel sets of ten stories each using our (ASP-guided) approach and a naïve (unguided) baseline, then comparing the semantic homogeneity of each story set via a sentence embedding model. Intuitively, we can calculate a *homogeneity score* for a set of texts via the following procedure:

1. Embed each text.
2. Average these embeddings together to calculate a *central* embedding for the set as a whole.
3. Calculate the cosine similarity between this central embedding and the embedding of each text in the set.
4. Average together these similarity scores to determine a summary homogeneity score for the set as a whole.

This is similar to the approach taken in several past studies of the homogenization effects of AI-based creativity support tools (Padmakumar and He, 2024; Anderson et al., 2024).

Homogeneity scores resulting from this approach range from 0 to 1, with higher scores representing more homogenous (i.e., less diverse) sets of texts. A score of 1 corresponds to a set of identical texts, while a score of 0 corresponds to a set of texts that share nothing at all in common.

In practice, an entire generated story is too long to compress into a single embedding using a standard sentence embedding model, so for our evaluation we instead represent each story as a *sequence* of embeddings: one embedding for each paragraph of generated text. We then calculate a separate homogeneity score for each *paragraph index* in a set of stories that each contain the same number of paragraphs. A set of stories that all begin with the same opening paragraph but diverge substantially in the second paragraph, for instance, will have a very high homogeneity score for the first paragraph but a lower homogeneity score for the second.

To validate that our approach improves story diversity across a wide range of different story premises, we repeat our evaluation procedure with six different *premise prompts*: short text strings that a user might input to a story generator to describe the basic premise of a story (e.g., “cat pirate”, “dwarven courtroom drama”).

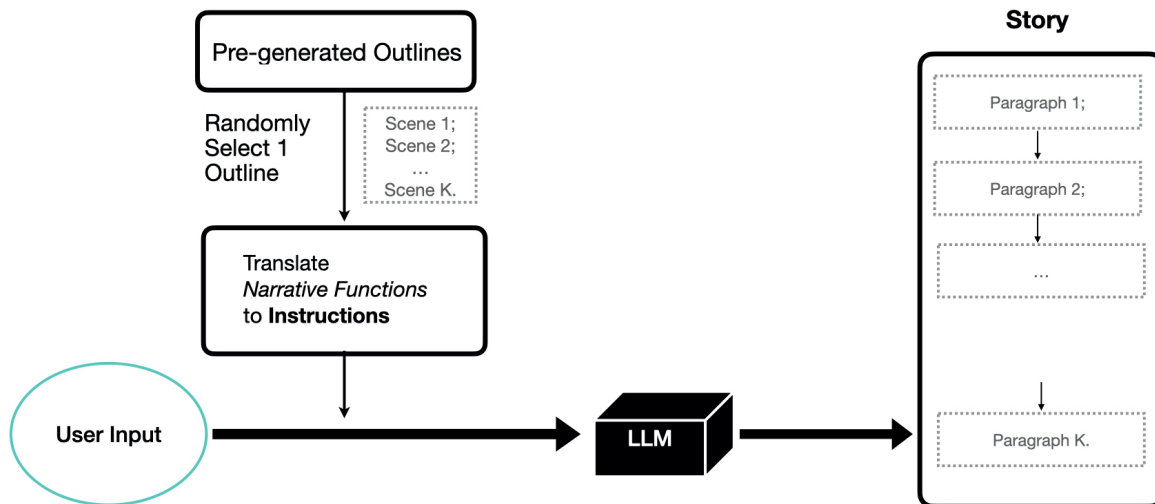


Figure 2: Our overall approach to ASP-guided story generation first takes in a *premise prompt*—a brief textual description of the high-level premise of the story to be generated—and a randomly selected outline from a set of valid outlines generated via ASP. We then translate the sequence of narrative functions defined by the outline to a sequence of *story continuation prompts*. These prompts are fed sequentially into the LLM to generate a sequence of paragraphs, each one taking into account the high-level story premise; the text of the story so far; and the narrative function that this scene (i.e., paragraph) must fulfill. The baseline (unguided) approach is identical, but uses a *neutral* story continuation prompt for each paragraph instead of taking a pre-generated story outline into account.

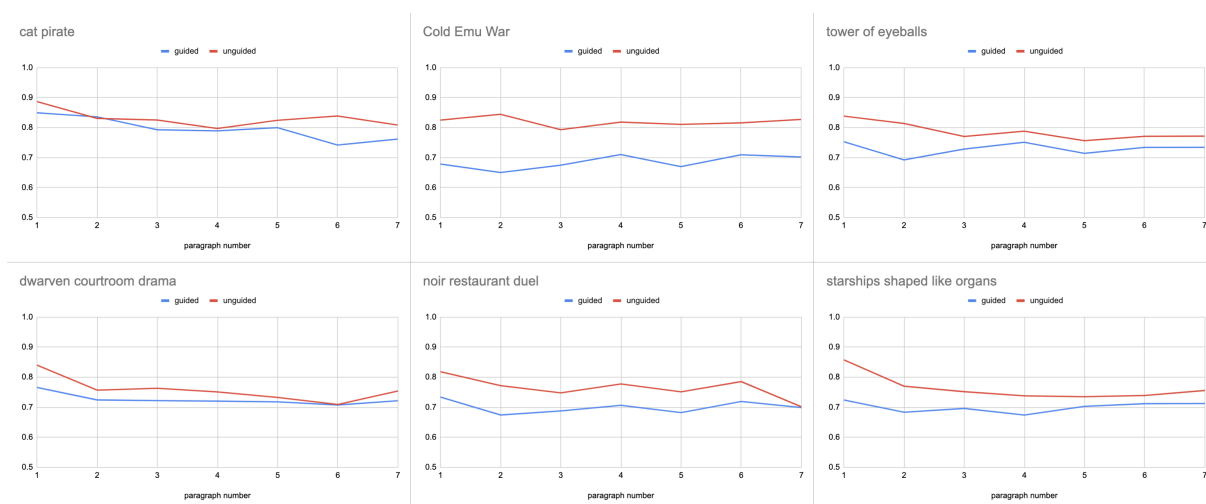


Figure 3: Comparative paragraph-by-paragraph homogeneity scores for stories generated around six different premises, using both an ASP-guided (blue) and an unguided (red) approach. For each premise, we generated a set of ten seven-paragraph stories using each approach, then visualized how semantically homogenous these story sets were at each paragraph. Lower homogeneity scores are better; the maximum possible score is 1 (all stories in the set are identical) and the minimum possible score is 0 (stories share nothing in common whatsoever).

We use the pretrained all-MiniLM-L6-v2 sentence embedding model to conduct semantic similarity calculations. This model has proven to align relatively well with human judgments of semantic similarity in a divergent ideation context (Anderson et al., 2024), and the general class of transformer-based sentence embedding models to which this model belongs has been shown to outperform word embedding models that were previously evaluated as the gold standard for semantic similarity calculation in creativity research (Beaty and Johnson, 2021; Dumas et al., 2021). all-MiniLM-L6-v2 is available as part of the open source SentenceTransformers library (Reimers and Gurevych, 2019), supporting replication of our evaluation approach.

3.2 Results

In general, we found that our ASP-guided approach consistently generated more semantically diverse sets of stories than an unguided approach, regardless of premise and paragraph number (Fig. 3).

Some premises benefited more from ASP guidance than others; for instance, stories generated in response to the “Cold Emu War” premise were much more diverse when ASP guidance was employed. Although the reason for these premise-level differences are not yet clear to us, they seem consistent across multiple different evaluation runs: for instance, story sets produced in response to the “cat pirate” premise consistently seem to benefit the least from ASP guidance in terms of diversity. Perhaps this is because, during pretraining, the LLM was exposed to more closely relevant training data for some of our evaluated story premises than for others; this could give the LLM an especially strong prior for what stories that respond to certain premises “should” look like, resulting in ASP guidance exerting less of an effect on overall story direction for these premises.

We also observed that earlier paragraphs tended to be more homogenous than later paragraphs regardless of whether ASP guidance was used. However, this trend was relatively weak and did not hold across all premises. Insofar as later paragraphs are more diverse than earlier paragraphs, this may be because later paragraphs are generated in the context of more preceding text, so minor differences in earlier paragraphs may “pile up” and be reinforced in later paragraphs as the later paragraphs become progressively more conditioned on earlier story contents.

4 Limitations and Future Work

Evaluation limitations Human readers are the ultimate arbiters of story quality, and although we believe stories generated with ASP guidance are at least no worse than unguided stories, we have not yet conducted any systematic human subjects evaluation of our approach. We use sentence embedding to evaluate the diversity of generated stories; however, embedding models do not align completely with human judgments of semantic similarity (Anderson et al., 2024), and seem to over-prioritize some aspects of sentences (e.g., nouns) compared to others (Nikolaev and Padó, 2023). The replicability of our evaluation is also limited by the use of ChatGPT 3.5 (a closed-source, closed-data, and closed-weights commercial model) as our target LLM, though we would expect to see similar results with any comparable instruction-tuned LLM.

Controllability limitations ASP guidance improves the diversity of generated stories by adding (constrained) *randomness* to the story generation process. This reduces the proportional contribution of language model defaults to the output stories, but doesn’t increase the proportional contribution of *user intent*; in fact, adding randomness without adding more user-controlled input information *reduces* the overall contribution of user intent to the output. This may exacerbate the “dearth of the author” effect, by which creative artifacts produced using AI-based tools can fail to strongly reflect the user’s authorial intent (Kreminski, 2024).

Future work In the future we plan to build an approachable tool that lets users interactively constrain the outline generation (ASP) portion of our generative pipeline, perhaps drawing interaction design inspiration from existing constraint-based tools for mixed-initiative storytelling (Stefnisson and Thue, 2018; Kreminski et al., 2022) and game design (Kreminski et al., 2020). The existence of such an interface would bring the injected randomness back under user control to some extent, helping to reverse the dilution of user intent. We also plan to investigate the use of an LLM to generate new ASP constraints in response to open-ended natural-language statements of storytelling intent from users, in much the same way that Kelly et al. use an LLM to generate narrative planning domains from natural-language descriptions of a high-level story premise. This could enable an even greater degree of user control over story outline generation.

5 Related Work

In the procedural content generation research community, ASP is widely used to specify design spaces of creative artifacts (Smith and Mateas, 2011). We are aware of two prior approaches to generating story outlines with ASP: one that leverages an ASP encoding of event calculus (Chen et al., 2010) and one that encodes a traditional narrative planner as an answer set program (Dabral and Martens, 2020). Both of these approaches differ from ours in that they use ASP to generate sequences of *plot events*; consequently, the resulting answer set programs are relatively complicated due to the need to explicitly encode every action that a character might potentially take, including associated preconditions and postconditions. Our approach, on the other hand, uses ASP to generate sequences of *storytelling goals*; fewer constraints are needed to specify valid outlines in this looser formalism, resulting in a much simpler program.

Luminate (Suh et al., 2024) introduces a new purely LLM-based plan-and-write approach that explicitly aims to generate diverse sets of stories. Given an initial premise, the system first uses an LLM to generate *dimensions* of a design space for stories that address this premise; then uses the LLM again to generate possible *values* for those dimensions; then uses it a third time to generate a structured set of stories that combine these values in diverse ways. This bears some resemblance to how we explicitly construct a design space of story outlines with ASP, and is in fact more flexible than our approach, but does not admit the specification of constraints that *disallow* certain combinations of values (which ASP readily permits).

Outside of plan-and-write (i.e., prompting-focused) approaches, several attempts have also been made to improve the diversity of LLM-generated text via more general, model-focused approaches. These approaches include the use of diverse decoding methods to extract more diverse outputs from a fixed LLM (Ippolito et al., 2019); quality-diversity approaches such as QDAIF (Bradley et al., 2024); and finetuning approaches to force the generation of “diffuse distributions” in contexts where output diversity is desirable (Zhang et al., 2024). All of these approaches are promising in their own right for improving output diversity, though they do not tend to provide new control affordances for story generation in the same way as plan-and-write approaches.

6 Conclusion

We have demonstrated the use of answer set programming (ASP) to guide LLM-based story generation toward greater diversity, using a concise answer set program (more flexible than a full-fledged narrative planning domain) to define a generator of valid story outlines and an LLM to expand these outlines into complete stories. We believe this approach represents a new reference point in the design space of neurosymbolic approaches to story generation, and a potentially compelling way to combine the open-domain story generation capabilities of LLMs with the greater content planning abilities of rules-based methods. We make our code available as a jumping-off point for future research in this direction.¹

Acknowledgments

This work was supported by University Research Grant GR102947, “A Neurosymbolic Approach to Creativity Support for Narrative Writing”, from Santa Clara University.

References

- Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In *C&C '24: Proceedings of the 16th Conference on Creativity and Cognition*.
- Roger E Beaty and Dan R Johnson. 2021. Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2):757–780.
- Nina Begus. 2023. Experimental narratives: A comparison of human crowdsourced storytelling and AI storytelling. *arXiv preprint arXiv:2310.12902*.
- Herbie Bradley, Andrew Dai, Hannah Benita Teufel, Jenny Zhang, Koen Oostermeijer, Marco Bellagente, Jeff Clune, Kenneth Stanley, Gregory Schott, and Joel Lehman. 2024. [Quality-diversity through AI feedback](#). In *The Twelfth International Conference on Learning Representations*.
- Gerhard Brewka, Thomas Eiter, and Mirosław Truszczyński. 2011. Answer set programming at a glance. *Communications of the ACM*, 54(12):92–103.
- Rogelio E Cardona-Rivera, Arnav Jhala, Julie Porteous, and R Michael Young. 2024. The story so far on narrative planning. In *34th International Conference on Automated Planning and Scheduling*.

¹<https://github.com/mkremins/spleenwort>

- Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, Smaranda Muresan, and Chien-Sheng Wu. 2024. Art or artifice? Large language models and the false promise of creativity. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- Sherol Chen, Adam M Smith, Arnav Jhala, Noah Wardrip-Fruin, and Michael Mateas. 2010. Role-Model: towards a formal model of dramatic roles for story generation. In *Proceedings of the Intelligent Narrative Technologies III Workshop*.
- Chinmaya Dabral and Chris Martens. 2020. Generating explorable narrative spaces with answer set programming. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 16, pages 45–51.
- Denis Dumas, Peter Organisciak, and Michael Doherty. 2021. Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of Aesthetics, Creativity, and the Arts*, 15(4):645.
- Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia. Association for Computational Linguistics.
- Martin Gebser, Benjamin Kaufmann, Roland Kaminski, Max Ostrowski, Torsten Schaub, and Marius Schneider. 2011. Potassco: The Potsdam answer set solving collection. *AI Communications*, 24(2):107–124.
- Michael Gelfond and Vladimir Lifschitz. 1988. The stable model semantics for logic programming. In *ICLP/SLP*, volume 88, pages 1070–1080. Cambridge, MA.
- Spyridon Giannatos, Yun-Gyung Cheong, Mark Nelson, and Georgios N Yannakakis. 2012. Generating narrative action schemas for suspense. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 8, pages 8–13.
- Seraphina Goldfarb-Tarrant, Tuhin Chakrabarty, Ralph Weischedel, and Nanyun Peng. 2020. Content planning for neural story generation with Aristotelian rescoring. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4319–4338, Online. Association for Computational Linguistics.
- Daphne Ippolito, Reno Kriz, João Sedoc, Maria Kustikova, and Chris Callison-Burch. 2019. Comparison of diverse decoding methods from conditional language models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3752–3762, Florence, Italy. Association for Computational Linguistics.
- Jack Kelly, Alex Calderwood, Noah Wardrip-Fruin, and Michael Mateas. 2023. There and back again: extracting formal domains for controllable neurosymbolic story authoring. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 19, pages 64–74.
- Max Kreminski. 2024. The dearth of the author in AI-supported writing. *arXiv preprint arXiv:2404.10289*.
- Max Kreminski, Melanie Dickinson, Joseph Osborn, Adam Summerville, Michael Mateas, and Noah Wardrip-Fruin. 2020. Germinate: A mixed-initiative casual creator for rhetorical games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 16, pages 102–108.
- Max Kreminski, Melanie Dickinson, Noah Wardrip-Fruin, and Michael Mateas. 2022. Loose Ends: a mixed-initiative creative interface for playful storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 18, pages 120–128.
- Max Kreminski and Chris Martens. 2022. Unmet creativity support needs in computationally supported creative writing. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, pages 74–82.
- Lara Martin, Prithviraj Ammanabrolu, Xinyu Wang, William Hancock, Shruti Singh, Brent Harrison, and Mark Riedl. 2018. Event representations for automated story generation with deep neural nets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.
- Dmitry Nikolaev and Sebastian Padó. 2023. Representation biases in sentence transformers. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3701–3716, Dubrovnik, Croatia. Association for Computational Linguistics.
- Vishakh Padmakumar and He He. 2024. Does writing with language models reduce content diversity? In *The Twelfth International Conference on Learning Representations*.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Mark O Riedl and Robert Michael Young. 2010. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, 39:217–268.

- Cory Siler. 2022. Open-world narrative generation to answer players' questions. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 18, pages 307–310.
- Nisha Simon and Christian Muise. 2022. Tattletale: storytelling with planning and large language models. In *ICAPS Workshop on Scheduling and Planning Applications*.
- Adam M Smith and Michael Mateas. 2011. Answer set programming for procedural content generation: A design space approach. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3):187–200.
- Ingibergur Stefnisson and David Thue. 2018. Mimi-brunnur: AI-assisted authoring for interactive storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 14, pages 236–242.
- Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminare: Structured generation and exploration of design space with large language models for human-AI co-creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- Yi Wang, Qian Zhou, and David Ledo. 2024. StoryVerse: Towards co-authoring dynamic plot with LLM-based character simulation via narrative planning. *arXiv preprint arXiv:2405.13042*.
- Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019. Plan-and-write: Towards better automatic storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7378–7385.
- Yiming Zhang, Avi Schwarzschild, Nicholas Carlini, Zico Kolter, and Daphne Ippolito. 2024. Forcing diffuse distributions out of language models. *arXiv preprint arXiv:2404.10859*.