

# Distributing Creative Responsibility Between a Knowledge-Based Content Determiner and a Neural Text Realizer<sup>\*</sup>

Pablo Gervás<sup>[0000–0003–4906–9837]</sup> and Gonzalo Méndez<sup>[0000–0001–7659–1482]</sup>

Facultad de Informática, Universidad Complutense de Madrid, Madrid, 28040 Spain  
{[pgervas](mailto:pgervas@ucm.es), [gmendez](mailto:gmendez@ucm.es)}@ucm.es  
<http://nil.fdi.ucm.es>

**Abstract.** Knowledge-based solutions for text generation are known to produce outputs that usually sound repetitive and stilted to the human ear. Attention-based neural solutions for generating text have proven to be successful at generating unconstrained prose that is fluent and sounds natural, but they have difficulty in producing texts that comply with a set of restrictions provided as input. The present paper explores combinations of a knowledge-based content generator and a neural text realizer, focusing on how creative responsibility over the final output is distributed over the knowledge-based and the neural modules of the system. A conceptual draft for a story is produced by a knowledge-based solution. The stories are then told using neural generators, with different types of prompt being built as means of requesting specific ways of telling the selected events. The outcomes are evaluated in terms of the percentage of the ideas in the final text that have been contributed by each module.

**Keywords:** storytelling · large language models · creative expansion

## 1 Introduction

Knowledge-based solutions for generating high level conceptual descriptions of contents have over the past decades shown capability to build complex structures using different traditional techniques such planning [12], evolutionary computation [6] or grammars [1]. Yet the rendering as text of contents generated in this way has always been found to lack fluency when compared to equivalent texts generated by humans. Attention-based neural solutions such as transformers have shown an ability to produce correct and fluent text as well as generating instances of basic structures such as simple stories [10, 2]. However, such solutions present difficulties when trying to generate longer stories [8]. They also

---

<sup>\*</sup> This paper has been partially funded by the projects CANTOR: Automated Composition of Personal Narratives as an aid for Occupational Therapy based on Reminiscence, Grant. No. PID2019-108927RB-I00 (Spanish Ministry of Science and Innovation), the ADARVE (Análisis de Datos de Realidad Virtual para Emergencias Radiológicas) Project funded by the Spanish Consejo de Seguridad Nuclear (CSN).

appear to react poorly when asked to generate texts that convey a given set of facts provided as input [5]. The problems observed include omission of requested facts, inclusion of extra facts or significant alteration of the relative order of the facts provided. Although a combination of a knowledge-based solution for generating large-scale structure and a neural solution for generating the final text has a potential to combine the best of both approaches, the possibility of destructive interactions occurring between the creative initiative of each type of modules may set a limit to their overall performance. This issue is particularly relevant when the process is intended not so much to produce a story *ex novo* but to come up with a narrative discourse to convey some subset of a recorded set of facts, intending it to be appreciated as a story.

The present paper explores how creative responsibility over the final output is distributed over the combination of a knowledge-based content generator that selects and structures content from an existing set of facts, and a neural text realizer that accepts a structured selection from the content generator and builds an interesting narrative discourse from it. A conceptual draft for a story is produced by a knowledge-based story sifting solution applied to the log of an agent-based social simulation. The stories are then told using neural generators, with different types of prompt being built as means of requesting specific ways of telling the selected events. The outcomes are evaluated in terms of the percentage of the ideas in the final text that have been contributed by each module.

## 2 Related Work

Two different topics have been reviewed to inform the work presented in this paper: computational approaches to deciding how to best convey a given content, and application of neural large language models to the generation of fluent text.

### 2.1 Computational Approaches to Telling Stories from Facts

An important source of insights on how to select content from an underlying storyworld to create an interesting narrative is recent research on how to curate world simulations to identify valuable stories present in them. Relying on the development of increasingly complex simulations of human behaviour, an approach emerged to produce interesting narratives by mining such simulations for particular subsets of events that exhibit desirable properties. This approach is referred to as *curating storyworlds* and as *story sifting* [9]. The most frequent criteria employed to drive these selection process is the identification of subsets of events that match a particular type of story.

A related approach considers how a subset of events from a given storyworld might better be told. This approach does include a process of selection of a subset of the available events, but it also considers decisions specific to the presentation of the selection, including the possibility of modifying the granularity at which the events are told or the relative order in which they are presented in the discourse for the resulting story. This task is known as *composition of*

*narrative discourse* [4], and it involves the construction of a *narrative discourse* that tells a subset of events extracted from a storyworld to ensure that: (1) the resulting discourse is a linear sequence even if the events involved have taken place over a range of locations and time periods, (2) the events are presented in an order specifically designed to make certain relations between them more easily accessible to the reader.

## 2.2 Large Language Models

A wide variety of Large Language Models (LLMs) have appeared during the last years, specially since OpenAI’s ChatGPT was launched. The ones used for the experiments reported in this paper are Google’s Gemini and Gemma.

Gemini [11] is a family of generative AI models that lets developers generate content handling both text and images as input. These models can generate text, handle multi-turn conversations and also zero, one, and few-shot tasks. These models have been trained with data existing up to the beginning of 2023. The Gemini Pro 1.0 version is available for use through an open web interface or through a commercial web API.

Gemma [3] is a family of lightweight, open models built using the same technology used to create the Gemini models. They are text-to-text, decoder-only LLMs, available in English, with open weights, pre-trained variants, and instruction-tuned variants. Their relatively small size makes it possible to deploy them in environments with limited resources such as a laptop, desktop or the free version of Google Colab. These models were trained on a dataset of text data that includes a wide variety of sources, including web documents, code and mathematical texts, totaling 6 trillion tokens. Their intended use includes text generation, conversational agents or text summarization. The model limitations comprehend open-ended or highly complex tasks and also the use of large contexts as prompts.

## 3 Measuring Relative Creative Responsibility over Different Configurations

This paper addresses the challenge of how to make the most of a subset of events selected from the log of an agent-based simulation to optimise the output of a neural generator set to render that subset as a fluent story. Optimisation may be driven by different criteria, depending on whether faithfulness to the situation taken as reference or creativity oriented to maximising story interest are given priority. The process involves testing a set of different prompts, generated from the conceptual representation of the output of an evolutionary process of selection based on emotional evolution of the characters. The prompts differ in the amount of information they include. The conceptual representation available for the output includes information at different levels of abstraction. The basic level information takes the form of a sequence of events whereby some agent proposes

joint activities to another, and the recipient of the proposal reacts by either accepting it or rejecting it. An additional layer of information registers the changes in the affinities between the agents that results from the interactions conveyed by the events. A final layer of information can be computed from the registered evolution of affinities in terms of the classification of different levels of affinity as different types of relation. These different layers of information contribute to the interest of the final story. In fact, if they are not passed on to the neural generator, it tends to invent additional facts that cover similar details. Therefore, by controlling the amount of information passed on from the conceptual description to the neural generator we can control whether the responsibility of introducing this type of data is assigned to the original simulation or to the neural generator.

### 3.1 Underlying Dynamics of the Agent-based Simulation

The material to be told in the stories generated in this paper has been extracted from the log of an agent-based simulation [7] that models the evolution of affinities between characters as a result of the responses elicited by proposals to carry out joint activities. The simulation is run with 15 agents who do not all know each other. Each agent may or may not have a partner, a small set of friends (between 2 and 4) and may or may not have any enemies (1 or 2 at the start). Interactions are driven by affinities between characters, and also act upon them. Probability of interaction is highest for partners, lower for friends, and lowest for enemies. Acceptance of proposals raises affinity between the characters, rejections and inactivity lower it.

Affinities between two agents A and B are directed, so what A feels for B may differ from what B feels for A. They are represented on a scale between 0 and 100, with 0 representing strong dislike and 100 representing passionate love. The Charade system considers a classification of relations between agents in terms of the affinities between them: *foe* affinity between 0 and 40, *neutral* affinity between 40 and 60, *friend* affinity between 60 and 80 and *mate* affinity between 80 and 100.

The type of relation that holds between two agents determines the subset of activities that they may consider together.

Each agent contributes to the general evolution of the simulation by: (a) proactively proposing interactions to other agents or reacting to proposals received, and (b) by registering changes in affinity towards other agents in response to proposals or reactions. The behaviour of agents is informed by the affinities between them, and it also has the potential to alter the affinities between them.

For a given log, an example of the subset of fact that may be produced by the story sifting process is shown in Table 1.

### 3.2 Template-Based Rendering of Sifted Stories

The original story sifting system included functionality to render the corresponding story as text automatically using basic templates for each of the actions involved. Two progressively more refined options would involve enriching the

	Megan to Tony		Tony to Megan	
	Affinity	Relation	Affinity	Relation
	96	mate	85	mate
Propose mate_dinner_with_candles / Megan / Tony				
Reject mate_dinner_with_candles / Megan / Tony	79	friend		
Propose mate_dinner_with_candles / Tony / Megan				
Reject mate_dinner_with_candles / Tony / Megan			78	friend
Propose friend_serious_talk / Megan / Tony				
Reject friend_serious_talk / Megan / Tony	72	friend		
Propose friend_help / Tony / Megan				
Accept friend_help / Tony / Megan			85	mate
Propose mate_have_coffe / Tony / Megan				
Reject mate_have_coffe / Tony / Megan			78	friend
Propose friend_serious_talk / Megan / Tony				
Accept friend_serious_talk / Megan / Tony	80	mate		
Propose friend_chat / Tony / Megan				
Accept friend_chat / Tony / Megan			85	mate

**Table 1.** Story draft obtained by story sifting from a Charade log. The evolution of affinity and type of relation for each of the characters is shown alongside.

textual rendering with information first on the changes of affinity between the characters taking place at each point, then with information on the changes of how each agent perceives their relation to the other. The selected subset of events shown in Table 1 would be rendered in each case as shown in Table 2.

### 3.3 Customised Prompt Generation from a Conceptual Representation

The ability of neural generators to build a text that renders a given story may be influenced using different strategies to generate the prompt from the available conceptual representation for the story.

The most basic prompt involves passing on to the generator the raw conceptual representation of the story. That would take the following form: `Use the events contained in the following table to write a story: <Table 1>`.

A more refined option would be to pass one of the original textual renderings as part of the prompt instead of the conceptual representation. That would take the following form: `Write a story that narrates the sequence of events described in the following text: <Text rendering option>`, where `<Text rendering option>` would be replaced by one of the options shown in Table 2.

### 3.4 Testing Different Prompting Strategies on Different LLMs

Two different types of model have been used to test the hypotheses of this paper.

The first model considered is the Gemini Pro 1.0 model accessed via its available web interface<sup>1</sup>.

The second model considered is the `gemma-2b-en` model over `keras3.0.5`, `GemmaTokenizer`, the `top_k` sampler to generate the output and a maximum output size of 1000 tokens.

<sup>1</sup> <https://gemini.google.com>

TBT 1	Megan proposes to Tony to dinner with candles as mates. Tony rejects Megan's invitation to dinner with candles as mates. Tony proposes to Megan to dinner with candles as mates. Megan rejects Tony's invitation to dinner with candles as mates. Megan proposes to Tony to serious talk as friends. Tony rejects Megan's invitation invitation to serious talk as friends. Tony proposes to Megan to help as friends. Megan accept Tony's invitation to help as friends. Tony proposes to Megan to have coffee as mates. Megan rejects Tony's invitation to have coffee as mates. Megan proposes to Tony to serious talk as friends. Tony accepts Megan's invitation to serious talk as friends. Tony proposes to Megan to chat as friends. Megan accepts Tony's invitation to chat as friends.
TBT 2	Megan proposes to Tony to dinner with candles as mates. Tony rejects Megan's invitation to dinner with candles as mates. <i>Megan's affinity towards Tony decreases.</i> Tony proposes to Megan to dinner with candles as mates. Megan rejects Tony's invitation to dinner with candles as mates. <i>Tony's affinity towards Megan decreases.</i> (...) Tony proposes to Megan to chat as friends. Megan accepts Tony's invitation to chat as friends. <i>Tony's affinity towards Megan increases.</i>
TBT 3	<b>Megan regards Tony as a mate. Tony regards Megan as a mate.</b> Megan proposes to Tony to dinner with candles as mates. Tony rejects Megan's invitation to dinner with candles as mates. <i>Megan's affinity towards Tony decreases.</i> <b>Megan starts regarding Tony more as a friend.</b> (...) Tony proposes to Megan to chat as friends. Megan accepts Tony's invitation to chat as friends. <i>Tony's affinity towards Megan increases.</i> <b>Tony starts regarding Megan more as a mate.</b>

**Table 2.** Templated-based renderings (*Template-Based Text*, TBT) for the three options: only events, events+affinity changes (*Italic*), and events+affinity changes (*Italic*)+changes in relation (**Bold**)

In both cases, the models are used with no fine-tuning of any kind, and using the default values for temperature. For Gemini, the value used for temperature is 1. For Gemma, the value used for temperature is 0.7.

Table 3 shows examples of the outcomes produced by each of the models when presented with the first type of prompt described in Section 3.3: using as description the content of Table 1.

The outcome produced by the Gemini model expands the given events into convincing sentences. In most cases, it adds information that gives colour to the story, such as inferring specific intentions that the characters may have for their actions (“Megan extended an olive branch to Tony”, “he then shifted gears”) or particular reactions that the events produces in them (“surprisingly, she agreed”, “to his surprise, she accepted”).

It is interesting to note that, although the additional information on variations in affinity or type of relation between the characters is included explicitly in the Table 1 used as input that is used to produce Neural text rendering 1, that information is not included in the resulting text. The model, at least without specific prompting to that effect, is unable to identify that information in the table as relevant to the story.

It is however commendable that, in spite of the fact that the input is presented as a continuous sequence, with no indication of segmentation, the model has identified that proposals of activities should be paired together with the corresponding reactions, and presents each such pair in a separate paragraph. Where there are apparent inconsistencies between the behaviour of the characters and the preceding context, the model introduces idiomatic comments that help to make them plausible (“he then shifted gears”, “surprisingly, she agreed”).

Gemini	<p>Megan and Tony seemed to be stuck in a loop of awkward attempts to connect. First, Megan suggested a romantic dinner with candles, but Tony wasn't interested. Tony then tried the same tactic, but it was Megan's turn to shut him down.</p> <p>Undeterred, they continued their back-and-forth. Megan proposed a serious talk, but Tony wasn't ready for that level of intimacy. Tony then asked for Megan's help, and surprisingly, she agreed. Perhaps this shared experience would bridge the gap between them.</p> <p>(...)</p> <p>Finally, they seemed to find common ground. Tony proposed a casual chat, and Megan agreed. Maybe this would be the start of a new chapter in their relationship, one built on friendship and understanding.</p>
Gemma	<p>Megan and Tony have been friends for a while, but they have not proposed to each other. Megan is interested in Tony, but Tony is not sure if he is interested in Megan. Megan and Tony meet for dinner, and they have a nice conversation. They both like each other, but Megan is not sure if Tony is interested in her. Tony and Megan meet for dinner again, and this time they have a serious conversation. Tony proposes to Megan and they both accept.</p> <p>(...)</p> <p>Megan and Tony still have some challenges in their relationship, but they are always willing to work through them together. Megan and Tony are both very committed to each other, and they are very proud of their relationship.</p>

**Table 3.** Outcomes generated by the tested models in response to prompt that relies on Table 1.

It also often introduces a final sentence to the paragraphs that provide a possible interpretation of the behaviour (“Maybe they could finally clear the air and figure out where they stood.” , “Perhaps they could finally have the honest conversation they both needed”).

Whereas the Gemini model reacts very well to inputs in the form of tables, reproducing the textual content of the tables quite faithfully and ignoring the numerical content, the Gemma model appears to be limited to considering the individual words that appear in the table as inspiration for its story making abilities, and it shows limited understanding of the meaning of the content of the table. The events appear to be used as inspiration for a story, and this leads to two contrasting consequences. First, the resulting story does not faithfully represent the events given as input. Second, the story constructed taken advantage of this freedom has many interesting details and paints a richer picture of the characters than the selection of events.

Table 4 shows examples of the outcomes produced by each of the models when presented with the second type of prompt described in Section 3.3, using as description template-base rendering TBT 1.

For this case, the outcome produced by the Gemini model shows characteristics similar to the outcome produced when prompting with the table.

For the Gemma model, the change from table prompt to text prompt is significant. Whereas outcomes for a table input were very free with the actual content and produced high quality text that conveyed an interesting story, the outcomes for the simple textual input demonstrate two interesting phenomena. Some of the sentences in the output reproduce very faithfully certain sentences in the input. This implies reproduction of the template-based text, which is clumsier. In addition, the model inserts after each such sentence what appears to be a comment phrased in very colloquial style. These comments are written

Gemini	<p>Megan extended an olive branch to Tony, suggesting a casual dinner by candlelight, but Tony politely declined. The tables turned when Tony offered the same invitation, only to be met with rejection from Megan as well.</p> <p>Undeterred, Megan sought a different approach, proposing a heartfelt conversation as friends.</p> <p>(...)</p> <p>Finally, they seemed to find neutral ground. Tony suggested a friendly chat, and Megan readily agreed. This simple act of conversation, devoid of romantic undertones, offered a potential starting point for rebuilding their friendship on a more solid foundation.</p>
Gemma	<p>Megan proposes to Tony to dinner with candles as mates. Tony rejects Megan's invitation to dinner with candles as mates. (Not sure why this is happening. Maybe they don't like each other's plans?). Tony proposes to Megan to dinner with candles as mates. (Same as before, maybe they still don't like each other's plans?).</p> <p>(...)</p> <p>Tony accepts Megan's invitation to serious talk as friends. (Yay, they're talking!). Megan proposes to Tony to chat as friends. (Maybe they just want to catch some waves together?). Tony accepts Megan's invitation to chat as friends. (Yay, they're chatting!)</p>

**Table 4.** Outcomes generated by the tested models in response to prompt that relies on TBT 1.

in brackets. Both of these practices are detrimental to the overall quality of the outcome.

Table 5 shows examples of the outcomes produced by each of the models when presented with the second type of prompt described in Section 3.3, using as description template-base rendering TBT 2.

Gemini	<p>The air between Megan and Tony crackled with unspoken tension. Megan, hoping to salvage a connection, extended an olive branch in the form of a casual candlelit dinner, only to be met with Tony's polite refusal. This initial rejection left a bitter taste in Megan's mouth, and her warmth towards Tony waned.</p> <p>Undeterred, Tony mirrored Megan's invitation, but his attempt was met with the same fate, further diminishing his perception of her. Discouraged, they both retreated, their fondness for each other dwindling with every unsuccessful attempt to connect.</p> <p>(...)</p> <p>Finally, they found a neutral ground: a simple chat as friends. Both readily agreed, their initial awkwardness replaced by a cautious optimism. This simple act of conversation, devoid of past baggage, offered a chance to rebuild their bond on a foundation of mutual respect and understanding. The future remained uncertain, but at least they were facing it together, one friendly chat at a time.</p>
Gemma	<p>Megan proposes to Tony to dinner with candles as mates. This means that Megan is interested in a serious relationship. However, Tony rejects Megan's invitation, which means that he has different interests or priorities, and doesn't want a serious relationship. Megan's affinity towards Tony decreases.</p> <p>(...)</p> <p>Tony proposes to Megan to help as friends. Megan accepts Tony's invitation to help as friends. Tony's affinity towards Megan increases as well. Now we have a relationship where both Megan and Tony want a serious relationship, but also enjoy each other's company as friends. Their affinity towards each other increases.</p>

**Table 5.** Outcomes generated by the tested models in response to prompt that relies on TBT 2. For the Gemma model, only an excerpt of the actual outcome is shown due to its great size, the omitted text follows much the same pattern as the rest)

For the Gemini model, when the information on the evolution of the affinities between characters is included in the textual format for the input in the form of additional sentences (Template-based text rendering 2, used to produce Neural



text rendering 3), one observes that the final sentences that summarise each paragraph are extended to convey these additional details. It is interesting to note that, while the sentences used to include the information on affinities in the input are solidly redundant in their form (determined by the template used for this purpose), the text produced by the model shows a significant effort to use different phrasings for each of the changes in affinity, even when they correspond to similar transitions. This is a very positive feature of these models.

For the Gemma model, the outcome for the prompt that includes the textual rendering of the events and the changes in affinity follows a similar pattern: the template-base sentences that describe the events in the input are rendered verbatim, but the colloquial comments are now replaced with proper sentences based on the information provided on the changes in the affinities between the characters. This slightly improves the quality of the text with respect to the previous version.

Table 6 shows examples of the outcomes produced by each of the models when presented with the second type of prompt described in Section 3.3, using as description template-base rendering TBT 3.

Gemini	<p>Megan and Tony, once comfortable friends, found themselves navigating a confusing dance of mixed signals. Megan initially saw Tony as a potential romantic partner, extending an invitation to a candlelit dinner. Tony, seeing her only as a friend, politely declined, leaving her slightly disappointed. This shifted Megan's perception, pushing her to see Tony solely as a friend.</p> <p>Tony, oblivious to the nuance, tried mirroring the romantic invitation, further confusing Megan and solidifying her view of him as just a friend. This rejection, in turn, dimmed his perception of her.</p> <p>Undeterred, they continued their awkward attempts to connect. Megan proposed a deep conversation, hoping to clear the air, but Tony, hesitant to delve into emotional territory, declined. This further distanced Megan, further solidifying her perception of him as just a friend.</p> <p>(...)</p> <p>Finally, they landed on a simple, neutral ground: a friendly chat. Both readily agreed, a newfound understanding replacing the initial awkwardness. They realized that their initial romantic misinterpretations had created distance, but their honest communication had paved the way for a stronger, more genuine friendship. The future remained uncertain, but they now faced it together, not as potential lovers, but as close friends, ready to navigate their bond with clarity and understanding.</p>
Gemma	<no outcome generated>

**Table 6.** Outcomes generated by the tested models in response to prompt that relies on TBT 3.

The Gemini model also reacts to the introduction of additional information regarding the way in which each character perceives their relationship with the other, and the way that changes through the story. In this case, the model appears to identify this additional information as sufficient to cover the need for a small interpretative sentence at the end of the paragraph explaining what has happened. The changes in the type of relation are included there and now replace altogether the interpretations that had been inferred in the outcomes of the previous experiments.

The Gemma model has proved incapable of generating outcomes for the prompt that provides as input the text rendering that includes the events, the changes in affinities between the characters and the evolution of the type of relation between the characters. We assume the size of the text in questions is beyond the operational range that the model can process.

### 3.5 Discussion of Results

A number of insights can be obtained from the experiments carried out.

First, there are differences in behaviour between the Gemini and the Gemma models to the different formats of input. Gemini reacts well to inputs in the form of tables, reproducing the textual content of the tables faithfully and ignoring the numerical content. Gemma appears to be limited to considering the individual words that appear in the table as inspiration for its story making abilities, and it shows limited understanding of the meaning of the content of the table.

Second, Gemma appears to follow a pattern of adopting a sentence from the input and following with some kind of comment. This shows up in the outcomes for both of the types of prompt that worked. This pattern reminds of the tendency of the Gemini model to end its paragraphs with an interpretative comment. But the Gemma model lacks the ability to identify, cluster and paraphrase the sentences in the input that describe a particular interaction between the characters.

Finally, Gemma shows a limited capacity for processing longer input texts.

The relative responsibility that each intervening module has on the final outcome can be computed by identifying, for each item of information included in the final text, whether it originates from the simulation run to construct the log used as reference or whether it has been introduced by the LLM during the rendering of the text. A number of metrics have been computed to provide indication of this: *covered* is the percentage of the number of information items in the input that have been included in the outcome, *missed* is the percentage of the number of information items in the input that have been omitted in the outcome, *contradicted* is the percentage of the number of information items in the input that have been contradicted in the outcome, *adlibbed* is the percentage of the number of information items in the outcome that have been included beyond those included in the input over the number of information items in the input, *correct* is the percentage of the number of information items in the outcome that have been included beyond those included in the input over the number of information items in the outcome and *incorrect* is the percentage of the number of information items in the outcome that contradict equivalent items included in the input, considered over the number of information items in the outcome.

The metrics for the different prompts and the different models are shown in Table 7. The results for the metrics show that the Gemini model performs significantly better than Gemma on several fronts. First, because Gemma fails to respond in some case. Second, because Gemma cannot interpret correctly the input given in table format. Third, because the texts generated by Gemma in several instances contradict information given in the input.

Prompt	Model	Text size	covered	missed	contradicted	adlibbed	added	correct	incorrect
Table	Gemini	22	48.28	51.72	0.00	24.14	31.82	63.64	0.00
	Gemma	25	0.00	100.00	0.00	86.21	100.00	0.00	0.00
TBT1	Gemini	27	48.28	51.72	0.00	44.83	48.15	51.85	0.00
	Gemma	30	34.48	51.72	13.79	55.17	53.33	33.33	13.33
TBT2	Gemini	43	69.97	31.03	0.00	79.31	53.49	46.51	0.00
	Gemma	38	44.83	44.83	37.93	34.48	26.32	34.21	28.95
TBT3	Gemini	49	72.41	27.59	0.00	96.55	57.14	42.86	0.00
	Gemma	-	-	-	-	-	-	-	-

**Table 7.** Metrics on relative contribution by reference simulation (SIM) and Large Language Model (LLM) for prompt that relies on Table 1.

Both models generate larger texts from the textual input than for the table input that includes the same facts. As the volume of information provided in the verbal prompts increases, the size of the outcome increases. The percentage of incorrect information present in the Gemma outcome also increases.

With respect to creative responsibility, the percentage of the information in the outcome that is introduced unwarranted by the neural generators increases with the volume of information provided in the input: the more information provided, more additional information is added. This is contrary to certain intuitions, where one might expect that more detail is added to information poor inputs, and less to information rich ones. The observed performance suggests that richer information in the input allows for richer elaborations that lead to the addition of more original facts.

## 4 Conclusions

The use of neural models for generating textual rendering of stories sifted from the log of a social simulation shows significant promise in the case of Gemini, much less so for Gemma.

The models respond well in terms of their ability to include all the provided events and present them in the given order. The Gemini model also shows a valuable ability to identify relevant relations between the events, such as contrast, and to react by including linguistic cues to turn these relations into appealing features of the resulting text. This model can also infer plausible motivations for the actions of characters when these are not explicitly provided, leading to text outcomes that are richer in nuance than the corresponding input. In contrast, the Gemma model replicates rather than paraphrase sentences from the input, and when it departs from it produces stylistically poor prose.

Both models do show a certain tendency to fall into recurrent patterns, such as addition of an interpretative sentence at the end of each segment generated from the input. This interpretative sentence summarises the content conveyed by the segment in an attempt to highlight inferred meanings. Texts produced by

the Gemini model exhibit a recurrent structure of the form `<text>`. `Undeterred,`  
`<text> Finally,` `<text>`.

The Gemini model by default avoids repetition of linguistic structures already used, even when the input from which it operates presents redundant structures of the corresponding information. In contrast the Gemma model fails to produce acceptable texts, and for longer texts fails to respond altogether.

From the point of view of creativity, it appears that the Gemini model is better at respecting the creativity of any preceding modules used to generate the input given to it, but both Gemini and Gemma exhibit a substantial amount of creativity in the process of rendering as text inputs for prompts provided.

The quantitative analysis of the generated texts shows that, in a zero-shot scenario, these models do not produce texts that comply well with the required restrictions. Further research with other LLMs and with few-shot scenarios is required in order to figure out whether they perform better or additional fine-tuning of the models is required. Specific automated quantitative metrics to capture whether the desired behaviour is attained would also help.

## References

1. Bringsjord, S., Ferrucci, D.: Artificial intelligence and literary creativity: Inside the mind of brutus, a storytelling machine. Psychology Press (1999)
2. Fang, L., Zeng, T., Liu, C., Bo, L., Dong, W., Chen, C.: Transformer-based conditional variational autoencoder for controllable story generation. arXiv preprint arXiv:2101.00828 (2021)
3. Gemma Team, T.M., Hardin, C., Dadashi, R., Bhupatiraju, S., Sifre, L., Rivière, M., Kale, M.S., Love, J., Tafti, P., Hussenot, L., et al.: Gemma (2024). <https://doi.org/10.34740/KAGGLE/M/3301>
4. Gervás, P.: Composing narrative discourse for stories of many characters: a case study over a chess game. *Literary and Linguistic Computing* **29**(4) (Aug 2014)
5. Gervás, P., Méndez, G.: Evolutionary plot arcs for a series of neurally-generated episodes. In: 15th ICCG. Jönköping, Sweden (jun 2024)
6. Gervás, P., Méndez, G., Concepción, E.: Evolutionary combination of connected event schemas into meaningful plots. *Genetic Programming and Evolvable Machines* **24**(1) (june 2023). <https://doi.org/10.1007/s10710-023-09454-2>
7. Méndez, G., Gervás, P., León, C.: On the Use of Character Affinities for Story Plot Generation, *Advances in Intelligent Systems and Computing*, vol. 416, chap. 15, pp. 211–225. Springer (feb 2016). [https://doi.org/10.1007/978-3-319-27478-2\\_15](https://doi.org/10.1007/978-3-319-27478-2_15)
8. Mikhaylovskiy, N.: Long story generation challenge. In: Mille, S. (ed.) *Proc. of the 16th INLG: Generation Challenges*. pp. 10–16. ACL, Prague, Czechia (Sep 2023)
9. Ryan, J.: *Curating Simulated Storyworlds*. Ph.D. thesis, University of California Santa Cruz, CA, USA (12 2018)
10. See, A., Pappu, A., Saxena, R., Yerukola, A., Manning, C.D.: Do massively pre-trained language models make better storytellers? arXiv:1909.10705 (2019)
11. Team, G.: Gemini: A family of highly capable multimodal models (2023). <https://doi.org/10.48550/arXiv.2312.11805>
12. Young, R.M., Ware, S.G., Cassell, B.A., Robertson, J.: Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* **37**(1-2), 41–64 (2013)