

TaleStream: Supporting Story Ideation with Trope Knowledge

Jean-Peïc Chou
Stanford University

Alexa F. Siu
Adobe Research

Nedim Lipka
Adobe Research

Ryan Rossi
Adobe Research

Franck Dernoncourt
Adobe Research

Maneesh Agrawala
Stanford University
Roblox

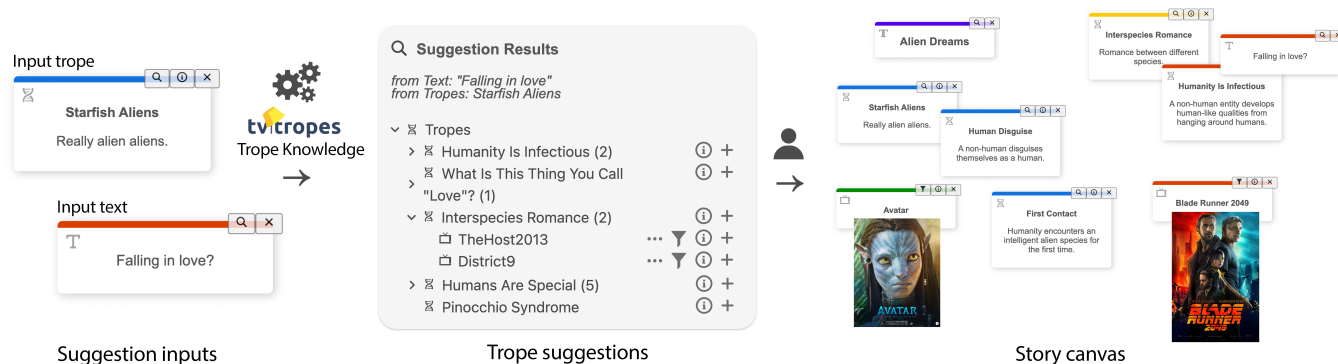


Figure 1: Our system allows users to generate story ideas by providing a set of inputs, which can include tropes and text, e.g., the *Starfish Aliens* trope and "Falling in love?" (left). Leveraging trope knowledge extracted from tvtropes.org, our suggestion algorithm automatically surfaces relevant tropes as story ideas with additional information (middle). Users can add some of these suggestions to their canvas, e.g., *Humanity is Infectious* and *Interspecies Romance* (right). Through iteration with the system, users can ideate and develop their stories.

ABSTRACT

Story ideation is a critical part of the story-writing process. It is challenging to support computationally due to its exploratory and subjective nature. Tropes, which are recurring narrative elements across stories, are essential in stories as they shape the structure of narratives and our understanding of them. In this paper, we propose to use tropes as an intermediate representation of stories to approach story ideation. We present TaleStream, a canvas system that uses tropes as building blocks of stories while providing steerable suggestions of story ideas in the form of tropes. Our trope suggestion methods leverage data from the tvtropes.org wiki. We find that 97% of the time, trope suggestions generated by our methods provide better story ideation materials than random tropes. Our system evaluation suggests that TaleStream can support writers' creative flow and greatly facilitates story development. Tropes, as a rich lexicon of narratives with available examples, play a key role in TaleStream and hold promise for story-creation support systems.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; **Interactive systems and tools**; • **Information systems** → **Users and interactive retrieval**; **Recommender systems**.

KEYWORDS

story ideation, tropes, story grammar, CST, recommender systems

ACM Reference Format:

Jean-Peïc Chou, Alexa F. Siu, Nedim Lipka, Ryan Rossi, Franck Dernoncourt, and Maneesh Agrawala. 2023. TaleStream: Supporting Story Ideation with Trope Knowledge. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*, October 29–November 01, 2023, San Francisco, CA, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3586183.3606807>

1 INTRODUCTION

Finding original and engaging ideas to create compelling stories is a challenging task for authors. Efforts in developing writing support systems have focused on continuing stories by producing text blocks to add and edit [56, 65]. However, beyond sentence generation, previous research has suggested that the main strength and use cases of such tools lie in their suggestive power to help overcome writer's block [6, 17, 24, 75]. In this regard, the focus of our work is to build a system that supports story writing by providing inspiring materials.



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

UIST '23, October 29–November 01, 2023, San Francisco, CA, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0132-0/23/10.

<https://doi.org/10.1145/3586183.3606807>

To ideate stories, drawing inspiration from existing stories is essential as authors consciously or unconsciously borrow story elements from each other. Such shared elements have been theorized by structuralists since the early 20th century [59]. To develop their own stories, authors often rely on well-known narrative structures such as the *Hero's Journey*, on common narrative devices creating or resolving conflicts like *Love Triangles*, or on archetypal characters such as the *Diabolical Mastermind*. As defined by the community from the wiki tvtropes.org, such recurring narrative elements belonging to common knowledge correspond to tropes, i.e. narrative concepts that “the audience will recognize and understand instantly”. Since 2004, through debates on the wiki forum and iterative modifications, thousands of enthusiasts have been establishing an extensive list of more than 24,000 tropes to break down all stories. Following the structuralist view, such patterns are helpful guidelines for structuring stories or sources of inspiration to develop them. Besides, as recognizable and predictable components, tropes shape the audience’s experience of stories and are, therefore, key elements for authors to grasp to build effective stories.

In this paper, we propose to leverage tropes as ideation fuel and story framework to support the design of stories. We introduce TaleStream, a story-creation support system that uses tropes as story-building blocks (Figure 1). To provide inspiring materials, instead of generating sentences, TaleStream suggests ideas in the form of tropes and gives access to related knowledge extracted from the wiki tvtropes.com. As needs, desires, and ideas continuously evolve and vary between authors, TaleStream provides steering controls over the suggestions that can be adapted through the creative process. In this regard, we built suggestion algorithms that were independently evaluated from the system. We found that 97% of the time, trope suggestions generated by our methods provide better story ideation materials than random tropes. In addition, we evaluated TaleStream with experienced story writers, who found the system to be a helpful and flexible creative assistant. Participants shared their enthusiasm for its unique perspective and ability to navigate the space of narratives. The use of tropes was particularly effective in this context, empowering participants to build the backbone of their stories without fear of lacking inspiration while being aware of existing patterns and representations. This work opens up several leads on using tropes as a framework for interacting with intelligent story-creation systems. In summary, this paper makes the following contributions:

- TaleStream, a story creation system that emphasizes story ideation by using tropes as building blocks
- Two approaches to suggest tropes by leveraging online data and methods to steer the suggestion results with various controls
- Results from a summative user study outlining the benefits and limitations of tropes for building stories and story creation tools.

2 RELATED WORK

2.1 Story frameworks

Story frameworks were first theorized in ancient Greece by Aristotle [11], who argued that tragedies should follow a three-act

structure including a beginning, a middle, and an end. In the mid-20th century, structuralist literary theorists regained interest in narrative structures, arguing that all stories shared universal elements and could be essentially reduced to a few narratives. Vladimir Propp’s *Morphology of the Folktale* was a pioneering effort to identify and classify the common narrative elements, or *functions*, of Russian fairy tales [59]. Other works have striven to propose more general story frameworks, breaking down stories into a unique storyline, the *Hero's Journey* [18], seven basic plots [16], or a more exhaustive list of 1,462 plots [25] for instance. The underlying mechanisms aggregating and structuring those elements were theorized in story grammars [52, 66]. Such plot structures are well-known and commonly used by professional story writers. Some of them have been integrated into computational tools to generate stories [30, 36, 38]. For instance, Gervas et al. make use of Propp’s elements to develop a story knowledge named *ProppOnto* [36]. Such approaches correspond to case-based reasoning techniques, which adapt stored stories as frameworks to new contexts. In [71] and [65], sentences are directly fetched from a corpus of stories to propose story continuation. Other methods generate stories by adapting higher-level attributes that can consist of goals and events as story units [62, 72]. Akoury et al. derive a large corpus of story components from STORIUM, an online collaborative game that lets users write stories based on cards as the framework [6]. With TaleStream, we leverage tropes from the wiki tvtrope.org as building blocks of stories. As the result of efforts that have spanned since 2004 by a large community, the more than 24,000 tropes arguably form a comprehensive, organized, and recognizable lexicon for storytelling. Besides, tvtropes.org provide rich information that we propose to make accessible in our system and to reason over for providing suggestions.

2.2 Tropes

Since the creation of tvtropes.org, the wiki’s data has been extracted several times and made available [33, 47] to help people build content generation tools or recommender systems [46]. Many works have proposed global analyses of the website’s rich information [22, 34, 35, 55]. Garcia-Ortega et al. highlighted key statistics about tvtropes.org data [34] and analyzed the trope co-occurrences in movies to determine a classification of tropes [35]. Chou et al. directly examined the website structure to build a trope-based knowledge graph of storytelling that provide semantic relationships [22]. Tropes have been shown to be representative of their works, being indicative of their genre [26, 68, 76] or inducing a character’s persona [12]. Systems such as TropeTwist [8], Story Designer [9], Ghost [39], dairector [29], Dear Leader’s Happy Story Time [42], or StarTropes [32] proposed to conceive stories with tropes as building blocks, plot points, or narrative beats. These systems were partly inspired by artist James Harris’ *Periodic Table of Storytelling*, which draws a comparison between stories, built on tropes, and molecules, composed of atoms [41]. For instance, Alavarez et al. design evolutionary narrative structures as graphs with tropes as nodes [8, 9]. Garcia-Ortega et al. aim at generating lists of tropes by predicting additional tropes that optimize the final story rating [32]. In dairector, improvisers directly interact with the system that proposes tropes as constraints based on prompts and other plot points.

However, these works make use of a limited repertoire of tropes, generally consisting of handpicked ones such as plot tropes or the most recurring ones. This limits users to flesh out stories, whereas we provide a more exhaustive list of tropes ranging from plots to low-level descriptions in TaleStream. In these works, users don't control the design of the story, whereas TaleStream allows users to steer the story idea suggestions. In addition, our work proposes to further the analysis of tropes as structural elements and their use in practice.

2.3 Story assistants

Available story assistant tools are numerous. Commercial tools such as *Dramatica* or *Plottr* help authors structure their plot or narrative in their creative workflow [1, 3]. More recently, there has been a surge of commercial tools leveraging progress with language models [2, 4]. Besides crowdsourcing efforts to help the creation of stories [43, 48], story assistants largely rely on autonomous story generation. Such methods include computational planning, where computers make decisions based on a set of predefined rules and objectives [53, 54, 63], character-based simulation [20], or case-based reasoning [36, 61–63, 72]. Most popular methods now involve language models that infill sentences [10, 44, 45, 73]. Such generative tools, however, necessitate the right level of control [24]. While some have proposed to guide the generation with keywords [31, 45, 70, 74], other works have proposed to use natural language prompts directly [27]. General conversational agents like ChatGPT [5] enable users to refine story generation through an iterative process and are now used by a wider audience [37]. Beyond text, visual elements have also been used as input to control the generation of stories as text [23]. However, the adoption of AI support systems for story creation has raised concerns among professional authors when it comes to translating ideas into words [14]. With TaleStream, we focus on supporting the ideation part of the story creation process rather than the writing. Instead of fully-fleshed linear sentences, our system relies on tropes for suggestions and as building blocks editable from a canvas. As a structured lexicon built on numerous references, tropes enable authors to navigate the space of existing stories to explore and analyze the proposed ideas.

3 TALESTREAM: WORKFLOW

We designed TaleStream by deriving from insights and recommendations of past works studying intelligent story writing support tools [17, 75], as well as from informal interviews that we conducted with five professional story writers. These interviews were aimed at gaining a deeper understanding of their creative processes during story ideation and their experiences with existing intelligent tools. We followed these design guidelines:

- DG1: The system should focus on suggesting ideas and encourage users to adapt the generated results to suit their own story
- DG2: The system should provide suggestions tailored to the needs of the author throughout the creative process
- DG3: The system should provide many suggestions that can be easily replaced and updated.

Our TaleStream interface is designed to help authors build their stories by suggesting adapted story ideas in the form of tropes and by giving them access to resources about these narrative patterns. Tropes, as suggestions, provide specific storytelling devices that can be uniquely fitted into the user's story (DG1). The system was iteratively improved with feedback from the five professional story writers through one pilot study. To demonstrate how TaleStream works in practice, we describe the workflow an author might experience to create a story. Our example shows an author developing a story about an "Alien Child", wishing to incorporate some "Romance" elements, and inspired by the movie *Blade Runner 2049*. As shown in Figure 2, the TaleStream interface consists of three main components: a Canvas (A1), a Control Section (A2, B1, B2), and a Results Section (A3, B3).

3.1 Board

The left pane of the interface shows the author's creation progression as a canvas containing their added story elements (Figure 2 A1). The canvas is a drag-and-drop interface in which users add and edit index card elements of different types (Trope, Text, Movie, Title, Image). As a brainstorming tool, 2D canvas interfaces offer flexibility for exploring, visualizing, and organizing story elements and complex ideas. Canvases are widely used in ideation processes across different fields and can be particularly effective when integrated with intelligent agents organizing, retrieving, and suggesting content [49–51] (DG1). In our example, the canvas was populated by tropes related to a forlorn "Alien Child", a few text cards describing the story in more detail, and a picture of *Blade Runner 2049* as inspiring reference.

3.2 Suggestion controls

During the creation process, users can ask for specific suggestions from TaleStream to develop their stories (DG2). The top-right part of the interface lets authors steer the suggestions based on their creative needs.

3.2.1 User Inputs. Users can specify the inputs of the suggestion results by selecting the canvas tropes and text elements, which can be found in the panel list (Figure 2 A2). We found through our pilot study that users wanted to use both trope and text elements to steer the suggestions. In our example, the author can look for ideas to flesh out the *Starfish Aliens* trope and add the text "falling in love?" as input in the search box. An additional text search box is implemented as a combo box containing all the tropes to facilitate the search for tropes – typing 'astronaut' in the search bar directly gives the tropes that include 'astronaut' in their names (Figure 2 B1).

3.2.2 Filters. We allow users to specify their search with category and movie filters (Figure 2 B2). Categories can notably be used to filter the resulting tropes by narrative function (e.g., Characters, Settings, Beginning Tropes), by theme (e.g., Comedy Tropes, Love Tropes), or by super-trope (e.g., Anti-Hero). With movie filters, users can also get trope suggestions from specific works and get direct inspiration from their content.

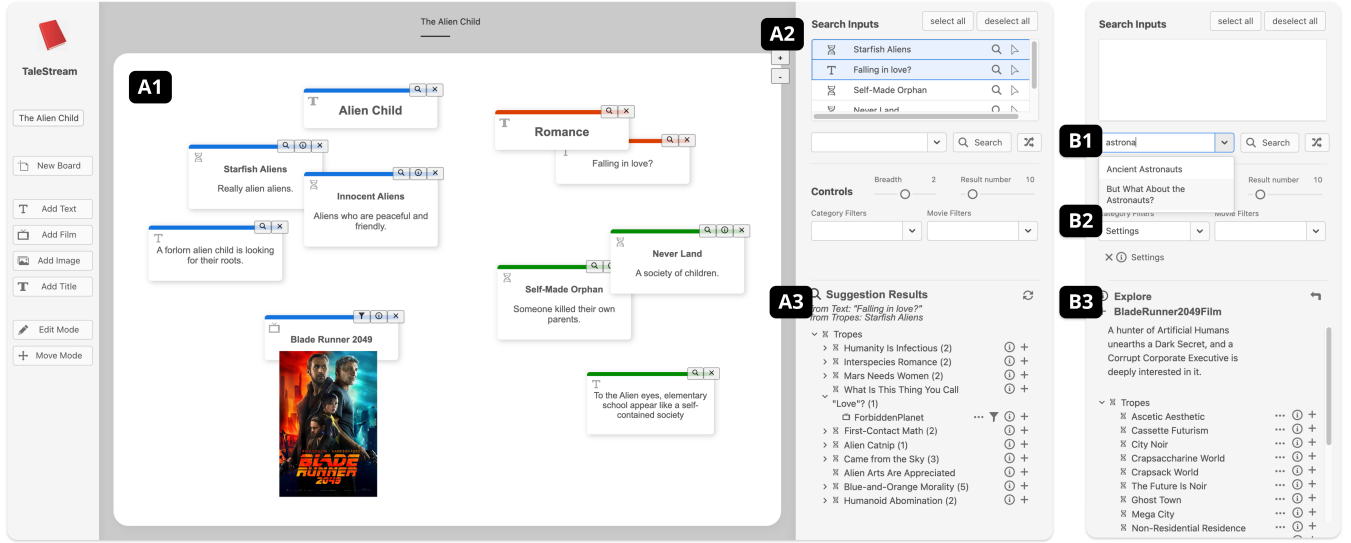


Figure 2: TaleStream interface. The Canvas (A1) contains the story elements that can be organized and edited. The Panel List (A2) enables the users to select the canvas elements for consideration when generating trope suggestions. Trope Suggestion Results (A3) include occurrence examples and additional descriptions. Users can use the Text Search Bar (B1) as an additional text entry or to find specific tropes. Additional Controls (B2) allow them to refine the suggestions by specifying their breadth or filtering trope lists. The Explore mode (B3), accessible from information icons, provides access to information on tropes and movies. For example, users can view a list of the *Settings* tropes used in *Blade Runner 2049*.

3.2.3 Number of suggestions. Participants from our pilot study reported that too many suggestions were cognitively overloading. We, therefore, let authors choose the number of displayed suggestions.

3.3 Results

3.3.1 Suggestions. When asking for suggestions, The Results section of the interface shows a list of trope suggestions (DG3). The author has access to laconic descriptions when hovering over the information icon and can add the corresponding tropes by clicking on the plus button. If a resulting trope co-occurs in movies with at least one of the selected input tropes, up to five of these movies are also displayed, along with a description of how the suggested trope is used in each movie when hovering over the three dots icon. A recap of all the specified inputs is displayed above the results. Responding to the example author’s inputs, the system notably suggests the *Interspecies Romance* (Romance between different species) and *Humanity is Infectious* (A non-human entity develops human-like qualities from hanging around humans) tropes which the author adds to the canvas (Figure 2 A3).

3.3.2 Explore. Authors can also learn more about a specific movie or trope by clicking on the information icons next to them. This switches the Results section to an Explore mode giving additional information about the element. If the element is a trope, authors have access to the categories it belongs to, its sub-tropes, and the movies it appears in along with the description of how it is implemented in them. If it is a movie, the system displays its synopsis and a complete list of its annotated tropes along with their description. These lists of tropes found in the Explore section can be filtered as

well, as shown in the example Figure 2 (B3), where the author examines the *Settings* tropes in *Blade Runner 2049* to break down and imagine a similar dark, futuristic, urban, and ascetic atmosphere.

4 TALESTREAM: TECHNICAL DETAILS

In this section, we focus on the technical details of surfacing relevant story ideas and the control mechanisms. Given some inputs from the user, we generate suggestions in the form of tropes that can be used to develop and revise the story. Our methods are based on data extracted from tvtropes.org (Section 4.1). These inputs can be of different kinds, either a set of tropes (Section 4.2) or some free-form text (Section 4.3), and can be jointly used with additional controls on the generated results (Section 4.4).

4.1 Trope and movie data

Our methods are based on data extracted from tvtropes.org, a wiki-like website on which a community of enthusiast “trophers” defined more than 24,000 tropes with rich information, as shown in Figure 3. On the website, tropes are described by a complete description and a “laconic” one. They are organized and grouped by what the trophers name “indexes”, i.e. categories, which can be tropes themselves. For instance, the *Anti-Hero* trope is an index that encompasses multiple sub-tropes such as the *Byronic Hero*, the *Justified Criminal*, or the *Moral Sociopath* tropes. Tropes can belong to multiple indexes, which typically group tropes according to a narrative function, theme, genre, medium, or by semantic similarity. The smallest indexes only include 1 trope, e.g., the index *Slice of Life*, which only contains the *Slice of Life Webcomics* trope, while the largest, the



Figure 3: Information about the *Vice City* trope extracted from tvtropes.org

index *Comedy Tropes*, includes 1,870 tropes. In addition, the community has annotated trope occurrences in movies and diverse media ranging from literature to advertisement. For each of these occurrences, troopers include a description of the trope implementation in the corresponding work.

Following Chou et al.’s work [22], we collect the wiki’s tropes and their attributes: their laconic definition, the tropes linked in their descriptions (that we will reference as *description* tropes), their indexes, as well as the movies in which they occur with their implementation details. Table 1 gives an overview of the retrieved dataset. We additionally make use of the MovieLens dataset [40] to provide complementary information about movies. This extracted information is used in our suggestion methods described in the next sections and made available in the Explore section of TaleStream.

Table 1: Overview of the retrieved dataset: Number of extracted elements (left) and tropes’ attributes’ mean number (right).

	Number	By trope	Mean number
Tropes	23,665	Description tropes	13.1
Indexes	1,988	Indexes	4.2
Movies	15,304	Occurrences	26.2

4.2 Trope suggestion

Our goal is to assist writers by suggesting story ideas in the form of tropes based on a set of input tropes. While prior studies have explored the use of story generation systems to suggest story ideas, these efforts have mainly focused on evaluating text-based features such as grammar, fluency, or lexical cohesion [60, 64]. The task of generating compelling story ideas is difficult to formulate and evaluate because it is inherently subjective. Our work focuses on ensuring coherence, a key characteristic that has been identified and extensively used in previous literature [6, 7, 19, 67]. The suggested story ideas should fit seamlessly within the user’s narrative, i.e. should be logically consistent with the input tropes.

Our proposed algorithms address the classic trade-off between exploitation and exploration in creativity [15] and recommender

systems [13]. The algorithms are designed to provide suggestions similar to the user’s inputs or introduce options that may be less related to broaden their horizons. By balancing these two approaches, our algorithms aim to offer a more comprehensive and personalized suggestion experience for users.

4.2.1 Index-based method. Our first method leverages the indexation of the tropes from tvtropes.org to provide coherent and closely related suggestions. With this approach, we focus on suggesting tropes that share similarities (e.g., theme, genre, function) with the input tropes, i.e., propose an exploitation method to retrieve suggestions. Similar tropes can encourage authors to imagine how to refine, combine, and develop the inputs. For instance, based on the input trope *Vice City* (an urban town infested with crimes), the trope *Crapsackcharine World* (a dystopian and grim place disguised in a wonderland) would be an output that shares similarities — both describe a place where darkness and terror reign — and that could be used to develop the input directly.

We compute similarities by comparing tropes based on their annotated indexes. For that, we use sklearn TF-IDF [58] by considering tropes as documents and categories as terms. We obtain a corpus of countable indexes for each trope by concatenating the indexes of the trope itself, as well as those of its description tropes, as shown in Figure 4. We use this larger corpus instead of the corpus composed of the trope’s indexes only for two main reasons. First, it allows us to weight indexes based on their frequency of occurrence, rather than treating them all equally. This weighting corresponds to having a variable Term Frequency in TF-IDF. Second, this corpus provides more detailed and nuanced information about a trope that the first-order categories may not capture fully (e.g., *Index Failure*, *Cynism Tropes*, *Horror Tropes*). We compute a similarity score between all tropes and the input trope to determine the ones to suggest. Table 2 shows the tropes with the highest scores when compared to *Vice City* as the input. For multiple trope inputs, we calculate the final score of each trope by multiplying the similarity scores obtained for each input, thus favoring tropes that are relevant in all aspects:

$$s_{ind}(\mathcal{E}_{T_i}, T) = \prod_{T_i \in \mathcal{E}_{T_i}} s_{ind}(T_i, T)$$

where \mathcal{E}_{T_i} is the set of input tropes, T is another trope we compare to, and s_{ind} is the index-similarity scoring function based on sklearn TF-IDF.

4.2.2 Co-occurrence-based method. Our second approach relies on trope occurrences in movies. Tropes that often appear in the same movies together are likely to fit easily into the same story. Unlike the first method, this approach doesn’t necessarily provide tropes that are close in terms of index, i.e. semantic category. The co-occurrence algorithm captures associations between tropes that may not be obvious or direct, resulting in broader and more “exploratory” suggestions. The method is similar to the previous one. We apply TF-IDF, considering tropes as documents, and movies as terms. We consider the list of movies in which a trope appears as its corpus to compute co-occurrence-similarities between tropes.

With this method, the number of directly co-occurring tropes can be limited, restraining the output coverage. This limitation occurs when a trope appears in only a few movies or in movies that don’t have many tropes listed. It poses two problems. Firstly,

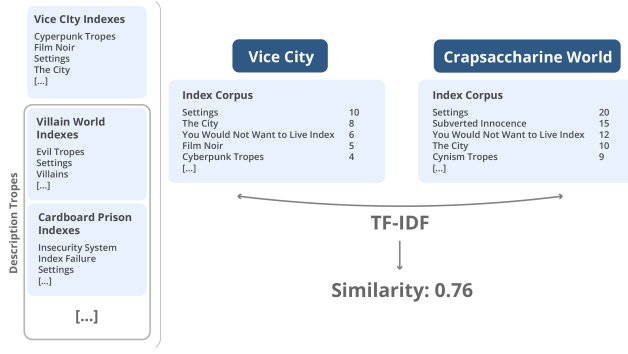


Figure 4: Index-based method for single inputs. Indexes of *Vice City* description tropes are aggregated to obtain a broader and weighted Index Corpus. These corpora are then used to compute similarities between tropes with TF-IDF.

Table 2: Tropes with the highest similarity to *Vice City* based on our methods.

Index	Co-occurrence	Mixed methods
Wretched Hive	False Utopia	City Noir
City Noir	Future Society, Present...	Wretched Hive
The Big Rotten Apple	Terror Hero	City of Adventure
The City	Cataclysm Backstory	Soiled City on a Hill
Crapsaccharine World	Color-Coded Castes	City on a Bottle

the suggested tropes are more likely to come from the same story, leading to less diverse and imaginative suggestions. Secondly, tropes that are only compared to a limited number of others are less likely to be suggested overall.

To address this limitation, we also make use of the description tropes, i.e., tropes mentioned in the description of the input trope. For each description trope, we calculate all TF-IDF co-occurrence scores and multiply them by the description trope index-similarity to the input trope to weigh their contribution. Instead of multiplying the scores, which would favor over-represented tropes that appear in all works (e.g., *Big Bad*, *Shout-Out*, *Oh*, *Crap!*), we keep the maximum score among the computed scores to aggregate the results:

$$s_{co}(T_i, T) = \max_{T_d \in \mathcal{D}_i \cup \{T_i\}} (s_{ind}(T_i, T_d) * \tilde{s}_{co}(T_d, T))$$

where T_i is the input trope, T is another trope we compare to, \mathcal{D}_i is the set of description tropes of the input, s_{co} and s_{cat} are the category-similarity and co-occurrence-similarity functions, and \tilde{s}_{co} is the first-order co-occurrence-similarity function.

Table 2 shows the method results for *Vice City* as input. For multiple input tropes, we calculate the similarity scores for each trope and keep the maximum score among them for the same reason as previously:

$$s_{co}(\mathcal{E}_{T_i}, T) = \max_{T_i \in \mathcal{E}_{T_i}} (s_{co}(T_i, T))$$

4.3 Trope search

We allow users to obtain story ideas from plain text. To implement this feature, we look for the tropes that are the most similar to the input text. We once again use sklearn TF-IDF. Tropes are still the

documents, and we use the examples extracted from the website movie pages to compose their corpus. Each trope corpus is obtained by concatenating its implementation descriptions in movies.

4.4 Suggestion controls

4.4.1 Breadth. We provide authors with a Breadth slider feature that controls the method to use to let them select the desired degree of exploitation versus exploration. Authors can set the Breadth slider to 1 to use the index-based method and to 3 for the co-occurrence method. For a balanced approach, authors can set the slider to 2, which combines both methods by multiplying their scores. A result example of the mixed method is shown in Table 2. To help users understand the connections between the input and output tropes, we provide examples of movies in which both appear together.

4.4.2 Mixing search inputs. To combine suggestions based on both trope and text inputs, output scores from the trope inputs are multiplied by the ones from the text inputs. This ensures that the combined result prioritizes items that are relevant to both the trope and text queries.

$$\tilde{s}(\mathcal{E}_{T_i}, T) = s_{trope}(\mathcal{E}_{T_i}, T) * s_{text}(text)$$

where s_{trope} is the tropes-to-trope similarity function, and s_{text} is the text-to-trope search function.

4.4.3 Temperature. To make the suggestions more diverse and less redundant, we introduce some randomness based on a final score accounting for their similarity score and their ranking to the model. We add a temperature parameter θ that controls the strength of the ranking over the distribution of the final scores, such as:

$$s(T_i, T) = \left(\frac{N_T - \text{rank}_{(\tilde{s}, T_i)}(T)}{N_T} \right)^{\frac{1}{\theta}} * \tilde{s}(T_i, T)$$

where $\text{rank}_{(\tilde{s}, T_i)}(T)$ corresponds to the rank of the trope T among all tropes based on their similarity \tilde{s} to trope T_i , and N_T is the total number of tropes considered.

Outputs are obtained by randomly drawing tropes without replacement. The probability of drawing a trope is proportional to its final score relative to the other final scores.

We empirically set θ to 0.02, deeming that it reasonably randomizes the outputs while providing satisfying suggestions following the score ranking.

5 SUGGESTION EVALUATION

In this evaluation, we show that both of our methods provide valuable suggestions while having the intended characterizations.

5.1 Methodology

We conducted a within-subjects evaluation of our two algorithms, using as a baseline randomly generated tropes. As trope inputs, we randomly selected 36 tropes. For each input, we provided five trope propositions generated by each algorithm and the baseline. Participants were asked to rate six statements on a 7-point Likert scale about the set of propositions in relation to the initial input idea:

- **S1-1:** I am familiar with the Initial Idea.

- **S1-2:** Each Proposition is often used with the Initial Idea.
- **S1-3:** Each Proposition shares similarities with the Initial Idea.
- **S1-4:** Each Proposition can be easily used with the Initial Idea.
- **S1-5:** Each Proposition offers a distinct direction to the story from the others.
- **S1-6:** I would use some of the Propositions to create my own story from the Initial Idea.

The first question allows us to verify that the input is understandable to the participant. Questions S1-2 to S1-5 enable us to characterize the outputs proposed by each algorithm. The last question determines if the algorithms suggestions are actually relevant to create a story.

Along with the first previous input trope, we then provided a randomly selected second input with a new list of five suggested tropes generated by one of the algorithms. Participants were asked to rate two more statements the same way:

- **S2-1:** The Propositions combine the two Initial Ideas.
- **S2-2:** I would use some of the Propositions to create my own story from the two Initial Ideas.

Each participant rated nine distinct sets of inputs. Each algorithm generated suggestions for three of these input sets so that every participant rated each algorithm three times on different inputs. The order of the algorithm appearances was randomized. Each set of inputs handled by one of the algorithms was rated by at least five to 12 participants. Each trope was accompanied by a short description to help them understand the tropes. In total, we recruited 96 users on Prolific to participate in this evaluation. Participants were screened based on self-reporting enjoying and regularly engaging in creating stories.

5.2 Results

The results are displayed in Figure 5. To analyze results, we only consider ratings where participants report being familiar with the initial input (at least 'Somewhat Agree' in S1-1). We removed 90 responses in which candidates declared being not familiar with the Initial Idea given as input for the evaluation. This represents 10% of the total number of responses. We average the participants' ratings for each input and question to obtain a mean rating. With a mean standard deviation of 1.22 for the five to 12 answers, we consider that the participants agreed on the ratings. For each algorithm, we average the mean ratings obtained for each input and employ non-parametric bootstrapping [28] with $R = 1,000$ iterations to derive 95% confidence intervals for all measures.

For single inputs, the Index and Co-occurrence algorithms both provided suggestions that were more likely to be used than the baseline ($\mu = 5.18$, $\mu = 4.65$, and $\mu = 4.08$ respectively, for S1-6). For this question, the mean ratings were higher than the baseline for 97% of the Index suggestions (35 out of 36 inputs) and for 75% of the Co-occurrence suggestions (27 out of 36 inputs). Although the Co-occurrence method relies on trope appearance with one another frequency, the Index method tropes were reported to be more frequently used with the input (Figure 5). Overall, the Index algorithm was largely found to provide suggestions that were considered the most often used with (S1-2), similar to (S1-3), and easily

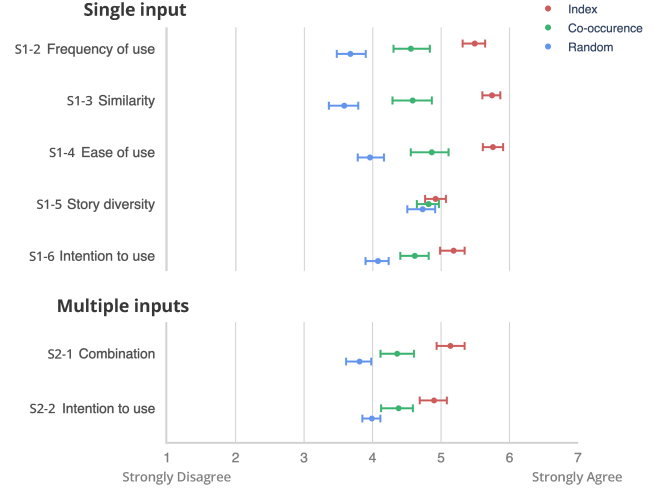


Figure 5: Suggestion evaluation results for our proposed Index and Co-occurrence-based methods and a baseline generating random tropes.

usable with (S1-4) the input trope. Besides, we checked that the suggestions generated by the Index and Co-occurrence algorithms were almost fully distinct: on average, only 0.28 suggestion out of the five proposed were the same. As a result, we conclude that our algorithms provide different suggestions that would be rather useful for developing the input trope and that the Index algorithm provides suggestions that are most closely related to the input.

In addition, each of the three compared methods was similarly rated regarding to diversity ($\mu = 4.92$, $\mu = 4.82$, and $\mu = 4.73$ for Q1-5). In other words, our methods suggestions were deemed to provide story directions as distinct as random trope suggestions. Finally, we note that the "Frequency of use with," "Similarity," and "Ease of use" properties are strongly correlated when examining each participant's answer to each input ($p(S1-1, S1-2) = 0.80$, $p(S1-1, S1-3) = 0.68$, and $p(S1-2, S1-3) = 0.75$ for the random suggestions ratings), which may reflect some semantic overlap.

We observed similar results for multiple inputs in terms of intentions to use, with overall positive feedback for both our algorithms. However, the virtual adoption of the suggestions is this time lower, with middling Co-occurrence suggestions. The Index algorithm demonstrates better semantic combination of the two inputs ($\mu = 5.13$), compared to the Co-occurrence algorithm ($\mu = 4.36$) and the baseline ($\mu = 3.81$), showing the algorithms' difference in characterizations again.

6 TALESTREAM EVALUATION

We conducted a user study to gain insights and feedback on the potential, limitations, and future opportunities of TaleStream and trope-based human-AI story co-creation. We focused on learning how our system and the use of tropes facilitate the exploration of story ideas and story co-creation. As this study seeks to understand

the users' engagement with tropes and the system, we did not conduct a formal comparison to existing human-AI story co-creation tools but relied on participants' experiences.

6.1 Participants

We recruited 10 participants from our institutions by word-of-mouth and through mailing lists. Participants completed a screening survey about their writing background before the study. We selected 10 participants who reported writing stories at least a few times a week, including five hobbyists and five experts. We considered participants who reported writing stories for professional purposes as experts and those who wrote for personal enjoyment as hobbyists. Our participants came from diverse backgrounds, including journalism, theater, literature, improvisation, cinema, and role-playing games. Participants presented a wide variety of experiences providing a range of perspectives on the use of tropes in storytelling. Four experts had more than ten years of professional experience (U1, U2, U8, U10) in their field, and the fifth one had five (U9). We did not collect hobbyists' years of experience. None of the participants had extensive experience with AI tools (the experiments were conducted before ChatGPT). Participants were compensated \$25 for the one-hour experiment.

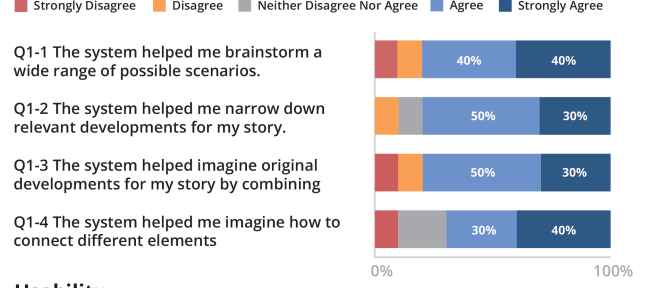
Table 3: Information about the participants of the system evaluation.

Participant	Experience	Fields
U1	Expert	Animation, Film-making
U2	Expert	Advertisement, Film-making
U3	Hobbyist	Literature
U4	Hobbyist	Film-making
U5	Hobbyist	Improvisation, Theater
U6	Hobbyist	Literature
U7	Hobbyist	Role-play
U8	Expert	Film-making
U9	Expert	Theater
U10	Expert	Film-making

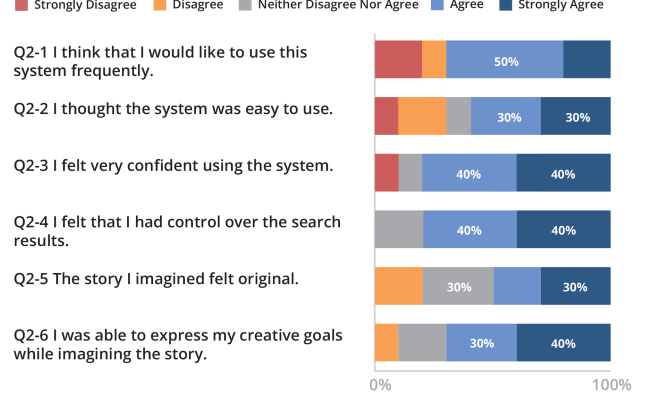
6.2 Procedure

We conducted remote studies on Zoom that were recorded. Participants were first given an overview of the study (5 min) before proceeding to a tutorial on using the tool (15 min). Participants used Chrome Remote Desktop to access the system on the interviewer's computer. Since our goal was to understand how experienced story writers interacted with our system and tropes, we asked participants to use our system to imagine a new story by filling the canvas with elements that they wanted to use in their story. We specifically encouraged them to conceive a story that they had not thought about before. Participants were not expected to complete a full story and we did not restrict how or when they should use the suggestions. We asked participants to think aloud during the experiment to learn about their rationales and reactions in using the tool. Afterward, we conducted semi-structured interviews (20 min) about their experience with the system, asking them to reflect on their own practices to draw comparisons.

Creativity Support



Usability



Controls

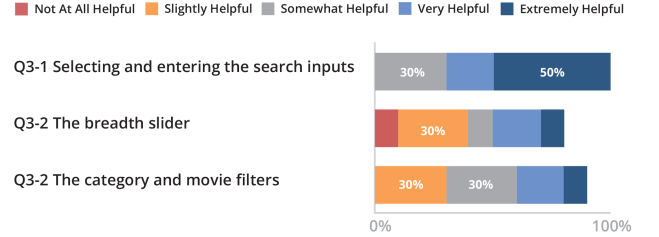


Figure 6: Results of TaleStream evaluation survey on creativity support, usability, and controls.

Lastly, participants were asked to fill out a questionnaire divided into three sections. Our tool was designed as an aid to authors, we, therefore, focused on obtaining participants' self-assessment of their results and their experience with the tool rather than relying on a third party's assessment of their creations. In the first section, participants rated their level of agreement with questions related to the system's support for divergent and convergent thinking using a 5-point Likert scale. The second section assessed the system's usability, asking participants to rate their impressions on various aspects of the system's design and functionality. Finally, the third section focused on the helpfulness of the features controlling the suggestion results. Participants were asked to rate each feature on a scale from "Not at all helpful" to "Extremely helpful," with an additional option for "Did not use." The full list of questions can be found in Figure 6.

7 RESULTS

7.1 Creating with TaleStream

7.1.1 Creativity support. In general, all participants expressed their enthusiasm for TaleStream: “It’s terrific. It was awesome. I love it.” (U4). 7 out of the 10 said they would use the system frequently (Q2-1 in Figure 6). Reflecting on their creative processes and the tools they were familiar with, participants found TaleStream unique and complementary to their workflow.

All 10 participants deemed the system very useful and efficient for writer’s block, helping them to ideate and develop compelling stories in a few minutes, which they would normally struggle to do in days in regular workflows. TaleStream supported convergent thinking. 8 participants agreed that the system helped them narrow down the possibilities and focus on some ideas (Q1-2). U10 liked that some ideas were “far more specific”, which “helped [them] specify and decide on which tropes [they] actually wanted to pursue”. Besides, the system helped connect different ideas (Q1-4) for 7 participants. U1 saw the system as “a GPS that takes [them] from point A to point B without replacing A or B”, helping them figure out how to fill out some holes.

The system also supported divergent thinking. The system expanded the participants’ range of story possibilities according to 8 participants (Q1-1). For U4, “inputting a term or word [branched] out almost like a spider web and [gave] different options”. Finally, the system’s suggestions combining ideas helped 8 participants to broaden their insights (Q1-3). For instance, U10 was surprised to be able to find tropes combining concepts related to Western and Sci-Fi and was inspired by the “mashup of different types of tropes” they obtained.

7.1.2 Collaborating with TaleStream. Most participants (8/10) felt they had adequate controllability over the suggestion results (Figure 6 Q2-4). Participants found selecting the inputs and entering text particularly helpful in guiding the search (Q3-1). Participants showed diverse opinions about the breadth slider and the filters (Q3-2 and Q3-3). In open-ended feedback, some participants described forgetting about or leaving aside the breadth feature due to the lack of time. The filters were used by half of the participants specifically when looking for inspiration for specific narrative holes, such as characters and settings (U6, U7), or from particular movies (U7, U10) for instance. Overall, participants found the controls understandable and easily usable to steer the suggestion results.

The system’s effectiveness, freedom (Q2-6), and ease of use (Q2-2) helped build confidence (Q2-3) and trust with the participants who felt they were collaborating with the system, referring to it as “a storyteller assistant” (U1), an “effective brainstorming partner” (U6), or “a writing partner” (U10).

7.2 Using tropes

7.2.1 A common story lexicon. While 6 out of the 10 participants had not previously heard about the concept of trope, all participants were familiar with most of the encountered tropes. Some even reckoned using these story mechanisms all the time, mostly unknowingly (U1, U5, U7, U9). This natural familiarity makes tropes particularly evocative, allowing participants to get a quick picture of possible stories (U1, U3, U5, U7, U10): “It gave me a visual of

what could be happening much faster than having to write it down” (U3). This common language was also deemed to be excellent for quickly communicating ideas (U8, U9) and for “getting everyone on the same page” (U8), which are key goals when making canvases.

This common language is built on a history of occurrences that participants enjoyed having access to easily. It helped them better understand some tropes (U7, U8, U10) and explore implementations of tropes (U4, U5, U7, U8, U9). Exploring the movies *The Groundhog Day* and *Looper*’s tropes, U7 stated: “The most important thing for me is understanding what this particular character or structure is, and then being able to translate that to a different set”. However, U2 and U6 felt a bit overwhelmed by the number of examples they were unfamiliar with, wishing they could filter some. Besides, U6 was “afraid of looking at the examples” because they did not want to over-rely on them.

7.2.2 Reinventing tropes. Originality is indeed often a concern when using tropes. Many participants mentioned overreliance on tropes as a trap (U2, U5, U6, U7, U8). All participants reported that TaleStream pushed them to be more conscious of these misuses, making U1 “more confident in building the story” or helping U4 “not plagiarize because the idea is out there”. Once aware of these pitfalls, tropes are great assets for story creation. All participants pointed out the diversity of uses of a single trope and the opportunity to reinvent each of them. With TaleStream, tropes were never seen as rigid building blocks, but as loose ideas that can be freely interpreted, interrogating more than imposing. As such, and despite relying on tropes to build their stories, participants generally considered their story created in less than 20 minutes original (Q3-4).

7.2.3 Building blocks. Within TaleStream, tropes were used as building blocks of stories. On average, more than half of the final canvas elements were tropes. Most of these tropes were suggested by TaleStream based on the participants’ inputs. Additionally, half of the text elements were directly used in conjunction with a specific trope to develop or detail it. U8 compared tropes to a “pre-built Lego set”, while U2 believes that “stories are made of other stories”. Overall, creating with tropes helped participants envision their stories at a high level, making them more aware of the underlying structure and increasing their confidence in the creation process. Participants used tropes to start building the “scaffolding” (U7) of their stories (U1, U5, U7, U10). Tropes then served to flesh out the backbone (U1, U2, U3, U5, U8, U9, U10) with “smaller and smaller details, almost like brushstrokes” (U3). Finally, participants described how they implemented the tropes in their stories in their own words, which felt like “coloring a story” (U5).

7.3 Summary

Overall, our study suggests that TaleStream provides an original and effective way to build stories with controls that support writers in their creative flow. Tropes, as story building blocks, proved to be a shared language among the participants, which they naturally adopted to explore existing works as well as imagine and conceive their stories. Working with tropes helped participants be more confident in the creation process by making them aware of the

underlying structure and patterns that constitute the core of their stories.

8 LIMITATIONS AND FUTURE WORK

8.1 Experiments limitations

Both of our experiments present limitations to consider. The mixed algorithm slider was introduced to the system only after conducting our suggestion evaluation of the Index and Co-occurrence algorithms for trope suggestions which revealed that both algorithms produced useful but distinct trope suggestions. Although we did not directly evaluate the mixed algorithm, participants in our system study did use the mixed slider to adjust the suggestions according to their needs and reported their satisfaction in both the questionnaire and the interview.

In the evaluation of TaleStream, our objective was to analyze the interactions between story writers and the system. However, the experiments did not entirely represent how a writer would use our tool. The study was limited to a 20-minute story ideation task, which may not have allowed participants to fully grasp the system. Participants were also asked to generate original ideas from scratch, which occasionally felt artificial (U6, U9). All participants expressed curiosity in trying TaleStream throughout the entire story creation process, not just at the beginning. It is also important to note that the insights derived from our results may not capture the characteristics of all workflows and that the features and design of our system may not be optimal for all cases. Our participants represented a relatively diverse range of backgrounds, each requiring different modes of thinking and creation support. Additionally, storytelling expertise could also influence the user experience. Although we did not observe significant differences in the interactions with the system between experts and hobbyists, our experiment was not designed to point them out. Our only observation was that expert story writers were more familiar with tropes and reported commonly using tropes (U1, U2) or having learned about them during their curriculum (U10). To obtain more realistic results, a randomized controlled and longitudinal field trial involving story writers over an extended period would be beneficial in exploring how creators perceive and integrate TaleStream into their existing creative workflows in different contexts.

8.2 Developing the trope framework

In this work, tropes were employed as a framework to approach narrative design, serving as building blocks. Our evaluation results highlighted several promising directions for further development. Overall, tropes prove to be effective elements for navigating the story space and obtaining tailored suggestions. We could explore additional mechanisms to support the search for tropes. For instance, incorporating category filter suggestions could assist users in refining their searches. We could also leverage insights from systems that incorporate users' direct feedback or analyze their activity [49] and develop analogous suggestions adaptation mechanisms for tropes.

Aside from tropes, canvases typically rely on visual elements. Many participants expressed their desire for visual aids to complement the tropes, which would be beneficial for grounding ideas (U2, U8, U9, U10). To illustrate specific tropes, images could either

be directly extracted from scenes in which they appear or generated based on wiki data. Other participants desired more detailed suggestions, such as "random details" about what characters have for breakfast within the Morning Routine trope (U7). This level of specificity could be provided through fine-tuned text generation using trope examples from tvtropes.org. More broadly, tropes could be used as high-level controls for generating text and guiding the flow of the story. Some participants envisioned a side text editor linked to the canvas and its elements. Within this framework, it could be easier and more natural for humans to express their creative intentions to computers while enabling computers to respond according to the story structure and stakes.

8.3 Inclusiveness and awareness

Tropes can perpetuate a narrow perspective, harmful representations, and stereotypes that can negatively impact individuals and communities. While tropes serve as convenient shortcuts for conveying familiar ideas, they can also hinder the exploration of more nuanced narratives by constraining people's imagination and inadvertently promoting laziness. Other biases can strengthen this danger. Our dataset was extracted from the English version of tvtropes.org, which strongly focuses on popular Western culture: U9 could not find a specific Kenyan film in the filters. It is essential for creators and consumers to be critical of these tropes, actively challenging and dismantling them to foster a more inclusive and equitable cultural landscape. One important challenge for future work is to address the limitations and biases of the dataset that will almost inexorably lead to biased results [57]. By aiming for greater diversity, inclusivity, and comprehensiveness in media representation, we can foster more nuanced storytelling systems. This involves seeking annotations from a broader array of sources, including non-Western references, to ensure a richer and more culturally varied narrative landscape. This could be accomplished by incorporating data from tvtropes.org in different languages or by developing automated methods for trope detection, an active area of research [21, 69]. Several other potential approaches can be considered to address biases. Firstly, implementing mechanisms to flag or censor tropes that are deemed problematic can help mitigate biases. Additionally, providing contextual guidelines that encourage critical thinking or offering examples of how to subvert each suggested trope¹ to propose alternative viewpoints are strategies to explore.

9 CONCLUSION

In this paper, we introduce TaleStream, a canvas system that suggests story ideas in the form of tropes. The trope and text elements on the canvas can be selected to generate trope suggestions which can be explored with movie examples and steered with additional controls on the story space. Our technical evaluation of the suggestion algorithms shows that our methods provide valuable results with different characterizations. The system evaluation revealed that TaleStream supports creative abilities for story ideation, provides reliable controllability, and is perceived as a pleasant partner accompanying the creative flow. The use of tropes in TaleStream was found to be particularly effective for quickly visualizing ideas

¹Following the *Playing with a Trope* article from tvtropes.org

through references, being aware of common pitfalls, and structuring stories, making users more confident while creating. This work opens up new ways to leverage tropes to support story creation as an intermediate comprehensive lexicon of storytelling.

ACKNOWLEDGMENTS

We thank Charly Lehuédé, Egan Tizzoni, Loïc Matos, Marc Löning, Ian-Christopher Tanoh, and Saehui Hwang for valuable conversations, our user study participants for their insights, and the reviewers for their feedback. The first author was partially supported by the Brown Institute for Media and Innovation.

REFERENCES

- [1] 1994. Dramatica. <https://dramatica.com/>
- [2] 2019. AI Dungeon. <https://play.aidungeon.io/main/home>
- [3] 2020. Plotr. <https://plottr.com/>
- [4] 2021. Sudowrite. <https://www.sudowrite.com/>
- [5] 2022. Introducing ChatGPT. <https://openai.com/blog/chatgpt>
- [6] Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyyer. 2020. STORIUM: A Dataset and evaluation platform for machine-in-the-loop story generation. *arXiv preprint arXiv:2010.01717* (2020).
- [7] Arwa I. Alhussain and Aqil M. Azmi. 2021. Automatic Story Generation: A Survey of Approaches. *Comput. Surveys* 54, 5 (May 2021), 103:1–103:38. <https://doi.org/10.1145/3453156>
- [8] Alberto Alvarez and Jose Font. 2022. TropeTwist: Trope-Based Narrative Structure Generation. In *Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3555858.3563271> event-place: Athens, Greece.
- [9] Alberto Alvarez, Jose Font, and Julian Togelius. 2022. Story Designer: Towards a Mixed-Initiative Tool to Create Narrative Structures. In *Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3555858.3555929> event-place: Athens, Greece.
- [10] Prithviraj Ammanabrolu, Wesley Cheung, William Broniec, and Mark O. Riedl. 2020. Automated Storytelling via Causal, Commonsense Plot Ordering. <https://doi.org/10.48550/arXiv.2009.00829> arXiv:2009.00829 [cs].
- [11] Aristotle. 2006. *Poetics*. ReadHowYouWant.com. Google-Books-ID: sywJT24pBb8C.
- [12] David Bamman, Brendan T. O'Connor, and Noah A. Smith. 2013. Learning Latent Personas of Film Characters. In *Annual Meeting of the Association for Computational Linguistics*.
- [13] Andrea Barraza-Urbina. 2017. The Exploration-Exploitation Trade-off in Interactive Recommender Systems. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 431–435. <https://doi.org/10.1145/3109859.3109866>
- [14] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (Virtual Event, Australia) (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1209–1227. <https://doi.org/10.1145/3532106.3533506>
- [15] Margaret A. Boden and Research Professor of Cognitive Science Margaret A. Boden. 2004. *The Creative Mind: Myths and Mechanisms*. Psychology Press. Google-Books-ID: 6Zkm4dz32Y4C.
- [16] Christopher Booker. 2004. *The Seven Basic Plots: Why We Tell Stories*. A&C Black. Google-Books-ID: XEUamcjBo9IC.
- [17] Alex Calderwood, Vivian Qiu, K. Gero, and Lydia B. Chilton. 2020. How Novelists Use Generative Language Models: An Exploratory User Study. <https://www.semanticscholar.org/paper/How-Novelist-Use-Generative-Language-Models%3A-An-Calderwood-Qiu/8cf1fc0b87dfda2a11bfaaa3a0bf9f9e069bb0f>
- [18] Joseph Campbell. 2008. *The Hero with a Thousand Faces*. New World Library. Google-Books-ID: 1ItuFuXlvFgMC.
- [19] Louis Castricato, Spencer Frazier, Jonathan Balloch, and Mark Riedl. 2021. Fabula Entropy Indexing: Objective Measures of Story Coherence. In *Proceedings of the Third Workshop on Narrative Understanding*. Association for Computational Linguistics, Virtual, 84–94. <https://doi.org/10.18653/v1/2021.nuse-1.9>
- [20] Marc Cavazza, Fred Charles, and Steven J. Mead. 2001. Characters in Search of an Author: AI-Based Virtual Storytelling. In *Virtual Storytelling Using Virtual Reality Technologies for Storytelling (Lecture Notes in Computer Science)*, Olivier Balet, Gérard Subsol, and Patrice Torguet (Eds.). Springer, Berlin, Heidelberg, 145–154. https://doi.org/10.1007/3-540-45420-9_16
- [21] Chen-Hsi Chang, Hung-Ting Su, Jui-Heng Hsu, Yu-Siang Wang, Yu-Cheng Chang, Zhe Yu Liu, Ya-Liang Chang, Wen-Feng Cheng, Ke-Jyun Wang, and Winston H. Hsu. 2021. Situation and Behavior Understanding by Trope Detection on Films. In *Proceedings of the Web Conference 2021 (WWW '21)*. Association for Computing Machinery, New York, NY, USA, 3188–3198. <https://doi.org/10.1145/3442381.3449806>
- [22] Jean-Peic Chou and Marc Christie. 2021. Structures in Tropes Networks: Toward a Formal Story Grammar. In *Proceedings of the Twelfth International Conference on Computational Creativity*. Mexico, Mexico. <https://hal.inria.fr/hal-03777738>
- [23] John Joon Young Chung, Woosok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3491102.3501819> event-place: New Orleans, LA, USA.
- [24] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, New York, NY, USA, 329–340. <https://doi.org/10.1145/3172944.3172983>
- [25] William Cook. 2011. *PLOTTO: the master book of all plots*. Tin House Books.
- [26] Anupam Datta, Sophia Kovaleva, Piotr Mardziel, and Shayak Sen. 2017. Latent factor interpretations for collaborative filtering. *arXiv preprint arXiv:1711.10816* (2017).
- [27] Alexandre Duval, Thomas Lamson, Gaël de Léséleuc de Kérouara, and Matthias Gallé. 2021. Breaking Writer's Block: Low-cost Fine-tuning of Natural Language Generation Models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, Online, 278–287. <https://doi.org/10.18653/v1/2021.eacl-demos.33>
- [28] B. Efron. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7, 1 (1979), 1–26. <https://www.jstor.org/stable/2958830> Publisher: Institute of Mathematical Statistics.
- [29] Markus Eger and Kory W. Mathewson. 2018. dAIrector: Automatic Story Beat Generation through Knowledge Synthesis. *CoRR abs/1811.03423* (2018). <http://arxiv.org/abs/1811.03423> arXiv: 1811.03423.
- [30] C. Fairclough and P. Cunningham. 2003. A Multiplayer Case Based Story Engine. <https://www.semanticscholar.org/paper/A-Multiplayer-Case-Based-Story-Engine-Fairclough-Cunningham/aea44aaafca25c3d4c1919a57258528d7dfdb798>
- [31] Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical Neural Story Generation. <https://doi.org/10.48550/arXiv.1805.04833> arXiv:1805.04833 [cs].
- [32] Rubén Héctor García-Ortega, Pablo García-Sánchez, and Juan Julián Merelo-Guervós. 2020. StarTropes, a film trope rating optimizer using machine learning and evolutionary algorithms. *Expert Systems* 37, 6 (2020), e12525. Publisher: Wiley Online Library.
- [33] Rubén H García-Ortega, Juan J Merelo-Guervós, Pablo García Sánchez, and Gad Pitaru. 2018. Overview of PicTropes, a film trope dataset. *arXiv preprint arXiv:1809.10959* (2018).
- [34] Rubén Héctor García-Ortega, Pablo García Sánchez, and Juan J Merelo-Guervós. 2020. Tropes in films: an initial analysis. *arXiv preprint arXiv:2006.05380* (2020).
- [35] Pablo García-Sánchez, Antonio Velez-Estevéz, Juan Julián Merelo, and Manuel Jesús Cobo. 2021. The Simpsons did it: Exploring the film trope space and its large scale structure. *Plos one* 16, 3 (2021), e0248881. Publisher: Public Library of Science.
- [36] Pablo Gervás, Belén Díaz-Agudo, Federico Peinado, and Raquel Hervás. 2005. Story plot generation based on CBR. *Knowledge-Based Systems* 18, 4 (Aug. 2005), 235–242. <https://doi.org/10.1016/j.knosys.2004.10.011>
- [37] Robert A. Gonsalves. 2023. Using ChatGPT as a Creative Writing Partner – Part 1: Prose. <https://towardsdatascience.com/using-chatgpt-as-a-creative-writing-partner-part-1-prose-dc9a9994d41f>
- [38] D. Grasbon and N. Braun. 2001. A Morphological Approach to Interactive Storytelling. (2001). <https://publica.fraunhofer.de/handle/publica/338099>
- [39] Andrea Guarneri, Laura Ripamonti, Francesco Tissoni, Marco Trubian, Dario Maggiorini, and Davide Gadia. 2017. GHOST: a GHOst STory-writer. 1–9. <https://doi.org/10.1145/3125571.3125580>
- [40] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems* 5, 4 (Dec. 2015), 19:1–19:19. <https://doi.org/10.1145/2827872>
- [41] James Harris. 2017. The Periodic Table of Storytelling. <https://jamesharris.design/periodic/>
- [42] Ian Horswill. 2021. Dear Leader's Happy Story Time: A Party Game Based on Automated Story Generation. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 12, 2 (June 2021), 39–45. <https://doi.org/10.1609/aiide.v12i2.12902>
- [43] Chieh-Yang Huang, Shih-Hong Huang, and Ting-Hao Kenneth Huang. 2020. Heteroglossia: In-Situ Story Ideation with the Crowd. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376715>
- [44] Yichen Huang, Yizhe Zhang, Oussama Elachqar, and Yu Cheng. 2020. INSET: Sentence Infilling with Inter-Sentential Transformer. In *Proceedings of the 58th*

- Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 2502–2515. <https://doi.org/10.18653/v1/2020.acl-main.226>
- [45] Daphne Ippolito, David Grangier, Chris Callison-Burch, and Douglas Eck. 2019. Unsupervised Hierarchical Story Infilling. In *Proceedings of the First Workshop on Narrative Understanding*. Association for Computational Linguistics, Minneapolis, Minnesota, 37–43. <https://doi.org/10.18653/v1/W19-2405>
- [46] Alister Johnson. 2018. Scaling Collaborative Filtering with PETSc. In *2018 IEEE International Conference on Big Data (Big Data)*. 4237–4244. <https://doi.org/10.1109/BigData.2018.8622202>
- [47] Malte Kiesel and Gunnar Aastrand Grimnes. 2010. DBTropes-a Linked Data Wrapper Approach Incorporating Community Feedback.. In *EKAU (Posters and Demos)*. Citeseer.
- [48] Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S. Bernstein. 2017. Mechanical Novel: Crowdsourcing Complex Work through Reflection and Revision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 233–245. <https://doi.org/10.1145/2998181.2998196>
- [49] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design Ideation with Cooperative Contextual Bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300863>
- [50] Janin Koch, Nicolas Taffin, Michel Beaudouin-Lafon, Markku Laine, Andrés Lucero, and Wendy E. Mackay. 2020. ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 45 (may 2020), 27 pages. <https://doi.org/10.1145/3392850>
- [51] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) (DIS '20). Association for Computing Machinery, New York, NY, USA, 407–418. <https://doi.org/10.1145/3357236.3395494>
- [52] George Lakoff. 1972. Structural Complexity in Fairy Tales. (1972). <https://escholarship.org/uc/item/6h38w8jc>
- [53] Michael Lebowitz. 1984. Creating characters in a story-telling universe. *Poetics* 13, 3 (June 1984), 171–194. [https://doi.org/10.1016/0304-422X\(84\)90001-9](https://doi.org/10.1016/0304-422X(84)90001-9)
- [54] James R. Meehan. 1977. TALE-SPIN, an interactive program that writes stories. In *Proceedings of the 5th international joint conference on Artificial intelligence - Volume 1 (IJCAI'77)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 91–98.
- [55] Clayton Mellina and Stacey Svetlichnaya. 2011. Trope propagation in the cultural space.
- [56] 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* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. <https://doi.org/10.1145/3544548.3581225>
- [57] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (FAT* '19). Association for Computing Machinery, New York, NY, USA, 220–229. <https://doi.org/10.1145/3287560.3287596>
- [58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [59] Vladimir Propp. 1928. *Morphology of the Folktale*. Vol. 9. University of Texas Press.
- [60] Christopher Purdy, Xinyu Wang, Larry He, and Mark Riedl. 2018. Predicting generated story quality with quantitative measures. In *Proceedings of the Fourteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE'18)*. AAAI Press, Edmonton, Alberta, Canada, 95–101.
- [61] Rafael Pérez Y Pérez and Mike Sharples. 2001. MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence* 13, 2 (April 2001), 119–139. <https://doi.org/10.1080/09528130010029820> Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/09528130010029820>
- [62] Mark Riedl. 2009. Vignette-Based Story Planning: Creativity Through Exploration and Retrieval. *Proceedings of the International Joint Workshop on Computational Creativity 2008* (Jan. 2009).
- [63] Mark O. Riedl and R. Michael Young. 2010. Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research* 39, 1 (Sept. 2010), 217–268.
- [64] Melissa Roemmele, A. Gordon, and R. Swanson. 2017. Evaluating Story Generation Systems Using Automated Linguistic Analyses. <https://www.semanticscholar.org/paper/Evaluating-Story-Generation-Systems-Using-Automated-Roemmele-Gordon/cf222683dd06990d76da48612c4dbe72d62f968d>
- [65] Melissa Roemmele and Andrew S. Gordon. 2015. Creative Help: A Story Writing Assistant. In *Interactive Storytelling (Lecture Notes in Computer Science)*, Henrik Schoenau-Fog, Luis Emilio Bruni, Sandy Louchart, and Sarune Baceviciute (Eds.). Springer International Publishing, Cham, 81–92. https://doi.org/10.1007/978-3-319-27036-4_8
- [66] David E. Rumelhart. 1975. NOTES ON A SCHEMA FOR STORIES. In *Representation and Understanding*, DANIEL G. Bobrow and ALLAN Collins (Eds.). Morgan Kaufmann, San Diego, 211–236. <https://doi.org/10.1016/B978-0-12-108550-6.50013-6>
- [67] Manasvi Sagarkar, John Wieting, Lifu Tu, and Kevin Gimpel. 2018. Quality Signals in Generated Stories. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, New Orleans, Louisiana, 192–202. <https://doi.org/10.18653/v1/S18-2024>
- [68] John R Smith, Dhiraj Joshi, Benoit Huet, Winston Hsu, and Jozef Cota. 2017. Harnessing ai for augmenting creativity: Application to movie trailer creation. In *Proceedings of the 25th ACM international conference on Multimedia*. 1799–1808.
- [69] Hung-Ting Su, Po-Wei Shen, Bing-Chen Tsai, Wen-Feng Cheng, Ke-Jyun Wang, and Winston H. Hsu. 2021. TrUMAN: Trope Understanding in Movies and Animations. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 4594–4603. <https://doi.org/10.1145/3459637.3482018>
- [70] Simeng Sun, Wenlong Zhao, Varun Manjunatha, Rajiv Jain, Vlad Morariu, Franck Dernoncourt, Balaji Vasan Srinivasan, and Mohit Iyyer. 2021. IGA: An Intent-Guided Authoring Assistant. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 5972–5985. <https://doi.org/10.18653/v1/2021.emnlp-main.483>
- [71] Reid Swanson and Andrew S. Gordon. 2012. Say Anything: Using Textual Case-Based Reasoning to Enable Open-Domain Interactive Storytelling. *ACM Transactions on Interactive Intelligent Systems* 2, 3 (Sept. 2012), 16:1–16:35. <https://doi.org/10.1145/2362394.2362398>
- [72] Scott R. Turner. [n.d.]. MINSTREL: A computer model of creativity and storytelling. Ph.D. University of California, Los Angeles, United States – California. <https://www.proquest.com/docview/304049508/abstract/7A5295B0C69E46D3PQ/1> ISBN: 9798209134299.
- [73] Su Wang, Greg Durrett, and Katrin Erk. 2020. Narrative Interpolation for Generating and Understanding Stories. <https://doi.org/10.48550/arXiv.2008.07466> arXiv:2008.07466 [cs].
- [74] Peng Xu, Mostofa Patwary, Mohammad Shoyebi, Raul Puri, Pascale Fung, Anima Anandkumar, and Bryan Catanzaro. 2020. MEGATRON-CNTRL: Controllable Story Generation with External Knowledge Using Large-Scale Language Models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 2831–2845. <https://doi.org/10.18653/v1/2020.emnlp-main.226>
- [75] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. <https://doi.org/10.1145/3490099.3511105>
- [76] Cecilia Åijälä and others. 2020. Using Film-Trope Connections for Clustering Similar Movies. (2020). Publisher: Helsingin yliopisto.