The Generative Fairy Tale of Scary Little Red Riding Hood

Lasse Harde
lharde18@student.aau.dk
Aalborg University
Aalborg, Denmark

Lasse Jensen
lonj18@student.aau.dk
Aalborg University
Aalborg, Denmark

Johan Krogh
jkrogh18@student.aau.dk
Aalborg University
Aalborg, Denmark

Adrian Plesner
aplesn18@student.aau.dk
Aalborg University
Aalborg, Denmark

Oliver Sørensen
ossa18@student.aau.dk
Aalborg University
Aalborg, Denmark

Henning Pohl
henning@cs.aau.dk
Aalborg University
Aalborg, Denmark

Figure 1: We used generative AI to create a version of Little Red Riding Hood that can adapt to the viewer's fear profile and emotional state. Through changes in scariness level and personalized content additions the story is optimized to the viewer.

ABSTRACT

Advances in generative text-to-image models are enabling new forms of personalized and adaptive media. We investigate the potential of such techniques through a generative adaptation of the fairy tale of Little Red Riding Hood. Specifically, we test two kinds of adaptations: (1) continuously adapting the visuals based on a face-to-emotion model, and (2) eliciting viewers' fears and adapting the story accordingly. In either case, the adaptive versions are designed to make the story more scary and thus enhance the viewing experience in this dimension. We compare both variants against a baseline condition in a between-subjects study with 97 participants. Our results show that these adaptations significantly alter the viewing experience, modulated by viewers' genre preferences.

CCS CONCEPTS

• **Information systems → Multimedia content creation**; • **Human-centered computing** → *Empirical studies in HCI*; • **Applied computing** → *Media arts.*

IMX '24, June 12–14, 2024, Stockholm, Sweden

ACM ISBN 979-8-4007-0503-8/24/06
https://doi.org/10.1145/3639701.3656303

KEYWORDS

generative media, adaptive media, interactive video

ACM Reference Format:
Lasse Harde, Lasse Jensen, Johan Krogh, Adrian Plesner, Oliver Sørensen, and Henning Pohl. 2024. The Generative Fairy Tale of Scary Little Red Riding Hood. In *ACM International Conference on Interactive Media Experiences (IMX '24), June 12–14, 2024, Stockholm, Sweden.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3639701.3656303

1 INTRODUCTION

There has been a continuous shift away from shared media experiences to more personal ones. In the early days of television, for example, only few channels were available and the family gathered around the single receiver in the home [7]. But with time, the number of media offerings has grown, and the audience has split alongside [39]. Today, few media events (e.g., world cup finals) draw in large numbers of synchronous viewers, and media consumption is increasingly personal and on-demand [42]. This trend is amplified by newer channels like Youtube, TikTok, and Twitch, which allow an ever growing number of creators to reach their audiences and viewers to watch recommended sequences of clips tailored towards their interests. As described by Evens et al., preference for streaming video is also related to a desire for community building [17] and thus more personal media use should not be understood as a trend towards individualization. For the case of TikTok, Lee et al. also point out that personalization can support self-reflection and connection to others [32]. Interactive video content, such as Netflix's Bandersnatch [46], offers further personalized experiences depending on each viewer's choices. The experiences here are reminiscent of "choose your own adventure" books and "visual novel" video games. A form of implicit interactive media are Varisco and Interlandi's "unconscious and enactive" interactive movie experiences, where the viewer's emotional state and eye gaze influence the content [53]. More personalized media experiences and consumption promise to cater more closely to audiences [20] (e.g., by addressing smaller niches or tailoring content to the identities and communities of viewers), thus improving the overall viewing experience (as also described by Evens et al. [18]).

A recent trend for videos, which further pushes the personalization of media experiences, is (semi-)automatic generation of content. For example, Balanzategui described the rise of algorithmically-generated "disturbing" Youtube videos aimed at children [3]. Another example are generated videos on platforms like Facebook (e.g., for memorial videos [31]) that automatically splice together photos and videos from users' histories. In addition to algorithmic video creation, an emerging class of AI systems also allows to create video content from textual descriptions. For example, Meta's *Make-A-Video* [48] can generate short clips in different kinds of styles from text, an image, an image pair, or another clip as input. With ongoing improvements around generative AI, video content can potentially be adapted to every viewer's preferences and context.

With the technology advancing rapidly, a wide range of video personalization and adaptation is possible. This could be changing the style of a video from realistic to painterly, changing the hair color of an actor or replacing them altogether, or adding ones own pets to a scene. Yet, our understanding of the impact of such changes is lagging behind the technological progress. Would it actually make for a better movie watching experience if, for example, a viewer's preferred music and settings are worked into the film? Inversely, could movies enhance an eerie mood by incorporating themes or set pieces the viewer more strongly reacts to? We focus to better understand the effects of adaptations and personalizations that alter the overall tone of a narrative, specifically, how *scary* that story is.

We investigate this question for a concrete piece of media: a short adaptation of the *Little Red Riding Hood* fairy tale. Its narrative and setting provide a fitting framework to experiment with adaptations that shift the tone towards being scarier. Using generative models, we transform a script of that story into a short movie with animation and narration. In addition to the baseline (i.e., a neutral rendition of the story), we explore two approaches: (1) a dynamic viewing experience that continuously adapts the shown content based on the viewers' emotions, and (2) a personalized viewing experience that modifies the story according to a viewer's profile. In both cases, the adaptations revolve around increased *scariness* of the fairy tale, meaning that: (1) the dynamic approach offers two levels of increasingly scary versions of the video, and that (2) the personalized approach elicits viewers' fears and then incorporates these into the story. While this approach conceptually allows for realtime adaptation of a narrative, we pre-generate the video clips as the used generative models are currently not fast enough. In summary, we investigate two research questions:

RQ1 How suitable are generative models as design material for adaptive filmic experiences?

RQ2 Does adapting content for scariness with generative models result in a better experience?

We compared the three viewing experiences in a between-subjects study with 97 participants and found effects for our adaptations, particularly on the perceived level of suspense. This is modulated by viewers' genre preferences, with participants who like horror and thrillers particularly enjoying the experience. Participants also found the noticeable AI-style of the generated videos themselves interesting and novel. Our results show the potential for adapting content to viewers through generative models. With most of the adaptation driven by text prompts, this kind of generative AI approach can be highly individualized and flexible.

2 RELATED WORK

Our work builds on advances in generative images and videos, but also connects to the area of interactive media experiences.

2.1 Generative Images

In their literature review, Ko et al. [30] pointed out the many tasks and roles tackled with the help of generative models in HCI research. They followed this up with artist interviews which further shed light on the complex roles of such models in their creative process. The target audience for *Opal* [36] instead are news editors, who are enabled to generate news illustrations of improved quality and range. Just as we use generative models to tell a fairy tale, the *StoryGAN* [49] architecture was designed for "story visualization", generating image sequences that fit the overall themes of a story as it evolves. StoryGAN maps sentences to images, but does not yet connect them together for a continuous story video.

As generative models most commonly make use of text prompts for input, that aspect of their use has received increased attention. For example, Liu and Chilton [34] have conducted a series of experiments to better understand the effects of prompts and to consequently derive design guidelines, such as to focus on the subject and style parts of the prompt. An automatic method for improving prompts is *RePrompt* [57], which was designed to optimize for emotional expressivity of the resulting images. Results showed that images created from prompts improved through *RePrompt* better aligned with the intended emotion as well as the input text. We drew on these results in our own prompt engineering.

Two other methods for controlling the results of image-generating models are *Make-A-Scene* [19] and *Initial Images* [43]. The former introduces a "scene layout" (i.e., an input image showing the desired segmentation of the output) which yielded results favored by human raters and also increases the controllability of the image generation (demonstrated through a generated childrens's story). *Initial Images* also investigates complementary image prompts for generative models. The results showed that such images help to better align the output with intentions and support a range of aesthetics. In our work we apply the above methods (image prompts in particular) in the creation of an interactive video experience.

2.2 Generative Video

While we focus on direct AI-generated video in this paper, there are also other text-to-video approaches. For example, Wang et al.'s *Write-A-Video* [55] pulls in and composes clips from video repositories based on an input text. Similarly, Yu et al.'s *Text2Video* [58] augments text scripts with video game footage and images, to enable novices to create narrative videos. Another approach by Chi et al. [8] uses synthesized talking heads as part of a system that converts documents to video prototypes. Chi et al. [9] also showed how tutorial scripts can automatically be turned into instructional videos. With *VScript* [27], Ji et al. take this a step further and generate full scripts from a genre and some starting words, which are then illustrated with suitable clips retrieved from a video database. Finally, *SmartShots* [50] takes data tables as input and then combines that with other media to generate data visualization videos.

Where the above systems are not fully generative, Meta's *Make-A-Video* [48] can create short video clips directly from input text, images, and videos. Similarly, Google's *Imagen Video* [25] and *Phenaki* [54] have similar capabilities. Other recent examples in that space are *Text2Video-Zero* [28] and *VideoFusion* [38]. Focusing on dance videos, Wang et al. [56] show that a reference photo and a set of target poses are enough to generate such content. On the commercial side, runway Research's *Gen-2* and tools by Fulljourney AI[1] offer video generation as well. This next generation of text-to-video models hints that the process is set to become more powerful and approachable, further demonstrating the potential for personalized interactive media as we describe it in this paper.

Closest to the approach taken in our work is Liu et al.'s *Generative Disco* [35] system. Using that system, users can annotate moments during a song with textual descriptions and the system then interpolates between these to generate music videos. Users can freely segment a song into intervals, which they can also describe using text. The system is able to suggest prompts based on the user descriptions. During subsequent interpolation between user-defined keyframes, the system also makes use of musical features (i.e., its "energy"). Our generated videos are also based on keyframe interpolation, albeit with guidance to ensure story alignment (i.e., consistent narrative and action) instead of fit to a music track.

2.3 Interactive Experiences

There is a long history of interactive experiences, which adaptive media is a part of. For example, de Sena Caires' [16] interactive film, "Transparency", demonstrated how viewer choices can result in different recombinations of clips while maintaining a filmic narrative. There have not been many interactive video experiences that have reached a wider audience and we can thus look at for understanding their reception. An exception is Netflix's *Bandersnatch*, for which Roth and Koenitz have investigated audience reactions [46]. While their participants found that experience usable and overall enjoyable, they also indicated a lack of agency and a feeling that their choices did not matter much. This also manifested in limited exploration of the different paths and endings. In contrast, we do not require viewers to make active choices and instead *implicitly* adapt the experience based on their preferences or reactions.

[1]https://research.runwayml.com/gen2 and https://fulljourney.ai/ respectively.

As Barber and Kudenko described [4], interactive narratives can also be fully generated based on users' preferences and actions. Similarly, de Lima et al. [14] recombine narrative variations based on personality profiles. Another example is work by Nakasone et al., who developed a *interactive storytelling model using rhetorical structure theory* [40]. Here, which story events take place is determined by structural criteria, but also to optimize for user interests. Guerrini et al.'s [23] presented a form of interactive movietelling, where AI is used to recombine video based on user specifications (e.g., which characters should be included).

Approaches to interactive media that require active viewer choices and manually designed branching are limited in complexity, with effort commonly scaling with the number of possible combinations. Furthermore, visual adaptation has similar limitations in such approaches, particularly if continuous change is desired. Moving toward automated adaptations, such as via the generative models we explore, has the potential to alleviate these limitations.

2.4 Emotion-Driven Experiences

"Emotion driven user experiences", as described by Bisogni et al. [6], adapt content to how a user feels at a given moment. Building on previous work [1], they described two VR applications that incorporate emotion detection via webcam: a horror experience and an air-hockey game. In both cases, the adaptations are small, such as the ambient lighting in the room adapting to users' expressions. After a jumpscare occurs in the horror experience, the game waits to detect fear and then transports the user to a serene and relaxing scene. Emotion detection was also used by Damiano et al. [13], albeit for a theater experience. Here, a camera films the audience and emotion labels are determined for visible faces. This data is mapped to a two-dimensional model of polarity (positive or negative) and intensity. Depending on the detected audience emotions, different video clips were played back behind the performer, who also engaged with the audience around these clips. As described by Zhao et al. [60], viewers' emotional reactions to imagery varies, which makes it hard to predict how any one viewer would react to generated imagery. However, they also showed that additional (e.g., social) data enables better prediction of the viewer's emotional reactions, demonstrating the suitability of personalized adaptations.

Emotional state can also be inferred from bio-signals, such as electromyography (EMG) or galvanic skin response (GSR). For example, Gilroy et al. [21] monitor such signals and adapt the narrative accordingly. Similarly, Liu et al. [37] used electroencephalography (EEG) signals to personalize advertising, such as by adjusting colors to counter low arousal. Cohendet et al. [12] used EEG readings to record the viewer's emotional responses and describe how this then can be used to create "emotional interactive movies". As Kirke et al. [29] demonstrated, adaptation can also work for groups. They recorded several bio-signals from four different viewers and then derived a group measure of emotional arousal. Based on this arousal reading the shown movie then adapts live and switches between clips to adapt the narrative to the audience.

Just as with interactive experiences, manually designed emotion adaptations is limited in possible complexity. Generative emotional adaptations again have the potential to reduce required effort and open up how much adaptation is feasible.

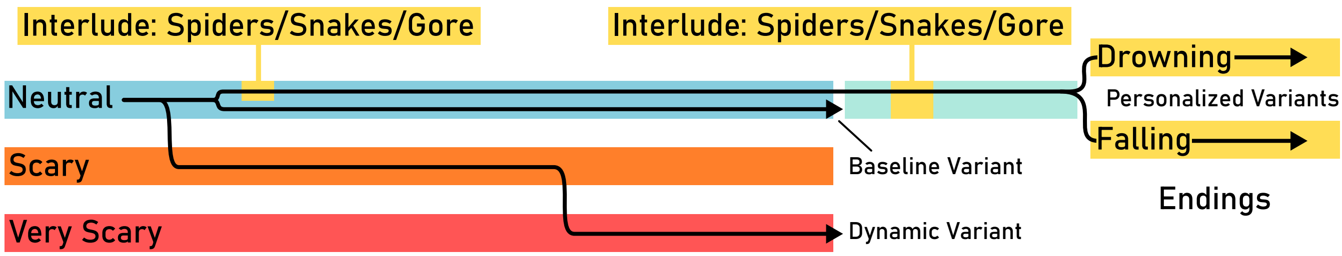

Figure 2: We created three variants of the experience: (1) a baseline one in a neutral tone, (2) a dynamic one that can get scarier in two levels, depending on the detected viewer emotions, and (3) a personalized one that adds additional interludes and endings based on a user's fears. Potential paths through the video material are shown here as arrows crossing the available scenes.

3 GENERATIVE LITTLE RED RIDING HOOD

We build upon the fairy tale of *"Little Red Riding Hood"* to create a generative interactive media experience. In that tale, a girl is on her way to visit her grandmother and meets a wolf that convinces her to take a detour. The wolf uses that time to rush to and eat the girl's grandmother, then disguising itself as said grandmother and waiting for the girl. When the girl arrives at the grandmother's house, she is puzzled by the wolf in disguise and questions it. At that point the wolf also eats the girl, with some versions of the tale continuing to have a woodcutter come by to rescue the girl and grandmother from inside the wolf. As de Lima et al. described, several variants of the story exist [15] which can also be recombined to create additional versions. In fact, in subsequent work they presented an interactive storytelling system that adapts the narrative according to personality profiles [14]. The story than varies, for example, in which path is taken to the grandmother's house.

We adapt the first part of the tale, up to where Little Red Riding Hood and the wolf part ways. Hence, our story takes place solely in the forest and only includes the characters of Little Red Riding Hood and the wolf. Both characters are visually distinct and easily identifiable, which is beneficial for generative media. Furthermore, the forest setting allows for a good amount of visual variation and thus for us to experiment with different modifications of the scene. Our script is based on a play[2], which has the advantage of being more concise and simple than the original fairy tale. We further split the script into scenes along narrative and acting character boundaries. Each scene then translates into a *clip* to be generated.

We create three different versions of our Little Red Riding Hood adaptation: (1) a *baseline* version with just one visual style and neutral tone, (2) a *dynamic* version that supplements the baseline with a "scary" and a "very scary" variant, and (3) a *personalized* version that, based on a choice of fears (snakes, spiders, gore, drowning, heights), adds additional interludes and endings to the story. The used script and scene information is available in Appendix A, see Figure 2 for an overview of the story variants. Depending on the viewed version, a sequence of the generated clips are then presented to the viewer. For the baseline and personalized versions, this is a static sequence, while the dynamic version changes which clips are shown during the viewing. Because the generative models used for video generation are too slow for realtime use, all clips are pregenerated. However, as video generation improves the process we describe here would also work for realtime adaptation of narratives.

[2]https://www.allansonstreetprimary.co.uk/wp-content/uploads/2020/12/Little-Red-Riding-Hood-Play-script.pdf

3.1 Generative Narration

We generated the narration audio using the ElevenLabs *Prime Voice AI*[3] text-to-speech software. Furthermore, we added ambient and effect sounds to the clips to supplement the narration. This includes, for example, growling noises for the wolf, rustling bushes, streaming rivers, and rock slides. We mixed together the sound effects with the narrator audio, placing them manually to fit the narration.

3.2 Generative Video

Text-to-image and text-to-video generation are fast moving areas with new models released regularly. We picked Stable Diffusion [44] for our project, as an open-source and freely available option. Specifically, we used Stable Diffusion v1.5[4] as a base model. Stable Diffusion has been trained on a wide range of images and styles, but has not been fine-tuned to any specific image category. For our fairy tale setting, we desired a semi-realistic style and it was particularly important to us that the visuals for Little Red Riding Hood and the wolf were of high quality. We hence turned to DreamShaper v4[5] as a model that has been fine-tuned for this very purpose. Furthermore, selecting a model with a specific style also aids visual consistency of the generated material.

We used the AUTOMATIC1111[6] web-based user interface (version 1.2) for controlling Stable Diffusion. We used several techniques within that environment to generate images, but primarily the *text-to-image* and *image-to-image* features. The former uses a pair of text prompts (positive and negative) for input, while the latter supplements this with an image the output should be based on as well. To generate videos, we used the Deforum[7] extension for Stable Diffusion (integrated into the web interface[8]). In particular, we made use of a feature for interpolating between two images, by repeatedly running the image-to-image process. This allowed us to first generate keyframes for each clip and then fill in intermediate frames. We also used ControlNet [59], which enables more

[3]Available at https://beta.elevenlabs.io/
[4]See https://huggingface.co/runwayml/stable-diffusion-v1-5
[5]See https://civitai.com/models/4384/dreamshaper
[6]See https://github.com/AUTOMATIC1111/stable-diffusion-webui
[7]See https://deforum.github.io/
[8]Using https://github.com/deforum-art/sd-webui-deforum

fine-grained control over where things should be in a sequence of frames. This, in combination with iterative prompt design, allowed us to achieve acceptable levels of visual consistency. With advances in this area (e.g., Midjourney's upcoming "consistent characters" feature[9]) such tight control will likely be less needed in the future.

3.2.1 Overall Prompt Design. While there are some design guidelines for image-generation prompts (e.g., [34]), this still is a process that requires exploration and experimentation. Hence, in our own process we primarily tried out a wide range of prompt variations to see what works well for our story. One aspect of the design is the selection of terms for the positive and negative prompts that steer the text-to-image model towards higher quality output. For example, our positive prompts commonly include terms like 'cinematic', 'high resolution', 'high detail', '8k', and 'intricate sharp details'. On the other hand, we list artifacts to avoid in the negative prompt, such as 'long neck', 'bad anatomy', 'extra arms', 'missing legs', 'poorly drawn face', and 'disfigured'. Furthermore, this also includes abstract terms like 'jpeg artifacts', 'text', 'signature', 'username', 'watermark', and 'low res', which we include to avoid some issues related to the quality of the training data.

3.2.2 Artists. A powerful way to steer the artistic style of generated images is to refer to the art of others. For example, instead of requesting Cubist art in general, asking for images in the style of Picasso further focuses the process. We refer to two artists in particular: Thomas Kinkade and Tomasz Alen Kopera. Thomas Kinkade is known for his idyllic and pastoral landscape paintings and we hence refer to him in scenes that primarily depict scenerie. On the other hand, Tomasz Alen Kopera's work commonly has a more surreal, dark, and mysterious tone and often features creatures and characters. These are desireable features for clips like those showing the wolf, but also closeups of Little Red Riding Hood.

3.2.3 Text-to-Image Parameters. The classifier free guidance scale parameter controls how much the given prompts influence the image generation process. It ranges from 1–30, where low values allow the model to deviate more and high values exert more control. At the extremes this leads to images not aligning with the prompt and artifacts, so commonly values in the range of 5–15 are recommended. The default value is set to 7, which is also what we use throughout this project.

We picked the *Euler Ancestral* sampling method for the denoising process. As an ancestral sampler, it does not always converge on the same result, but on the other hand creates a wider, "more creative" set of results. With this sampler, the number of sampler steps also has a larger in influence on the results. We use 20 sampler steps for all clips we generated.

3.2.4 Image-to-Image Parameters. In addition to the text-to-image prompts, we also employed initialization images for most scenes. This allowed us to better specify how we would roughly like the scene to be arranged (see Figure 3 for an example). We use both generated and existing images for these initialization images. In case of the former, we combined text-to-image prompts with a follow-up image-to-image optimization. For the latter, we pulled in online images that fit the desired scene composition and style. We made manual additions to some of these online images and also further modified them with image-to-image prompts.

Figure 3: For many scenes we used initialization images (left), which we then augment with additional prompts to generate the final frames (right) for our clips.

3.2.5 Animation Parameters. With Deforum, our video clips are generated through a process of repeated image-to-image executions. We used a frame rate of 12 fps for our videos, which is common in animation. This aligns well with our visual style and also accentuates how the clips are generated. The noise and variation of the generative models shows as visible morphing and blending between frames. Instead of jerky frame-by-frame animation, the resulting imagery is more akin to a smooth interpolation between related images. We combine this with the slight zoom and panning effect of Deforum to make the clips more animated.

An important parameter for this process is the strength schedule, which determines how much a previous frame influences the next one. We use values between 0.55–0.75 for the strength schedule, depending on the scene and content within. For example, setting it to lower values creates sequences that are more eery, with visual elements shifting around more. The cadence parameter on the other hand determines how many frames are generated through diffusion and how many are interpolated. We use a cadence of two which means that every other frame receives diffusion. This creates more vivid video, but comes at the cost of more flickering and higher computational cost.

3.2.6 ControlNet. For some scenes we required more fine-grained control over the movement within them. For example, when Little Red Riding Hood walks down a path in the forest, we found that a text prompt on its own was not sufficient to generate this with the tools at the time. Here we make use of ControlNet where the generated frames are further constrained to a secondary input. This can be skeletal animations, depth images, facial expressions, or scribbles, with us using the latter. With scribbles, outlines and edges define where image features in the final output should be located. We use a manual process with hand-drawn and animated scribbles that provide a coarse reference sequence. Figure 4 shows an example of how this only requires little detail but can generate the desired sequence. This level of control was not necessary for most clips, given they did not contain movement of characters, and we thus only used ControlNet in about a quarter of the scenes.

[9]See, for example, https://www.theverge.com/2024/3/12/24098282/midjourney-is-testing-a-highly-requested-consistent-characters-feature

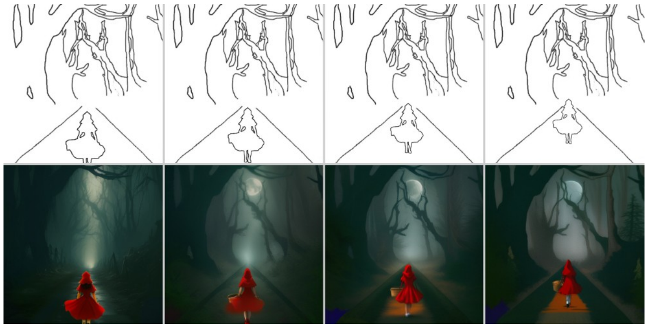

Figure 4: We used ControlNet for more fine-grained steering of the generated animations. Shown here is the input scribbles (top) and resulting frames (bottom) for a sequence of Little Red Riding Hood walking through the forest (left to right: frames 1, 30, 60, and 90).

Figure 5: In addition to the neutral baseline (left), the dynamic version of our story also offers a scary (middle) and very scary (right) variant.

3.3 Dynamic Condition

The dynamic condition adjusts the scariness of the story as it is being watched. This aims to provide a suitably scary experience for different viewers, optimizing for this aspect of the story. We vary scariness in three levels: neutral (i.e., the baseline experience), scary, and very scary. The two scarier versions of clips are generated through prompt changes from the baseline, maintaining the overall narrative while adapting the visuals.

To make clips scarier, we add terms such as 'aggressive', 'evil', and 'nightmare fuel' to the positive prompt and terms like 'colorful', 'life', 'green' to the negative prompt. For the very scary version we accentuate this further, such as by asking for 'very aggressive' within the positive prompt. We show an example of these prompt differences for one clip in Appendix B. Figure 5 also shows a comparison of the three levels of scariness within one scene.

We switch between the different levels of scariness on a clip-by-clip basis. Thus, at the end of a clip, before the next one is played, we determine which level is appropriate. If the viewer is too scared, we tone it down and if they are not scared we increase scariness. In lieu of a continuous measure of how scared the viewer is at a given moment, we chose a scheme that only relies on discrete emotion labels. As shown in Figure 6, we base our adaptation just on whether the emotion of 'fear' was detected in the viewer or not. Emotion detection always runs for the whole current clip and represents the prevalent emotion during that time.

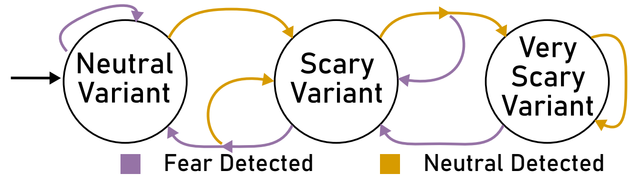

Figure 6: The dynamic viewing experience switches between three levels of scariness, depending on which emotional state is detected for the viewer. Switches happen based on (one arrowhead) the detected emotion from the current clip, or (two arrowheads) the detected emotions from the current and previous clip.

When viewers did not feel fear during a clip, we increase the scariness level in the next one and conversely decrease it if they did feel fear. Furthermore, to move from 'scary' to 'very scary' and from 'scary' to 'neutral', we require that the corresponding emotion was not just prevalent during the current clip, but also the one before that. This 'sticky' logic was chosen to prevent too much oscillation between the three scariness levels. As the dynamic condition (as well as the baseline condition) contains 17 scenes altogether, there are 16 potential instances for a shift in scariness level.

3.4 Personalized Condition

In the personalized condition the story is modified to cater to individual users. Given that the clips are created with a generative model, a wide range of adaptations can be performed. In line with the overall approach of our project, our personalizations also focus on scariness and how to make the story more scary for individual viewers. In contrast to the dynamic condition, the personalizations are driven by viewer profiles entered before the story is watched.

Given the goal of scariness, we decided to personalize the story based on viewers' individual fears. Specifically, we incorporate a set of five fears: (1) arachnophobia, the fear of spiders, (2) ophidiophobia, the fear of snakes, (3) hemophobia, the fear of blood and injuries, (4) acrophobia, the fear of heights, and (5) thalassophobia, the fear of deep water. We selected these from among common fears [2], but also for their suitable integration with the given story. For example, fears of crowds or public speaking are impossible to integrate with the story of Little Red Riding Hood.

As shown in Figure 2, we insert personalized content in three places of the story (also see Figure 7). First, as Little Red Riding Hood walks through the forest, she stops to pick up a flower. As she grabs it, she is startled by either spiders, snakes, or a gory display, depending on the picked fear. Later on, Little Red Riding Hood is walking on the wrong path the wolf sent her on. Here, she again encounters a representation of her fears (either spiders, snakes, or gore), which terrifies her and makes her run away. Finally, we end the story with one of two personalized endings, with Little Red Riding Hood either drowning, or falling to her death. The choice here depends on whether viewers indicated a stronger fear of heights or deep water.

4 EVALUATION

We compared our two interactive video experiences against the baseline condition in a between-subjects study.

Figure 7: The personalized version of our story adds two interludes (left and middle column) and endings (right column) based on the viewer's reported fears.

4.1 Participants

We recruited 97 participants (age 19–59, M=24.6, SD=5.9; 17 women, 79 men, and 1 who preferred not to disclose) around our institution. Participants provided informed consent and were not reimbursed for their participation. We randomly assigned participants to one of the three story versions, resulting in 31 of them watching the baseline condition, 31 the personalized condition, and 35 participants watching the dynamic condition.

4.2 Apparatus

The study was run on desktop computers with a *Logitech c930e* webcam attached and a headset connected. Participants sat down in front of this computer to watch the videos and listened to the audio through the headset. We ran this study in quiet and low-traffic shared spaces on campus, where participants were not disrupted, but also not sitting on their own. In the baseline and personalized conditions the video to watch is fixed, but for the dynamic condition many different clip sequences are needed. Hence, we use a video streaming setup where a server generates an MPEG-DASH stream that is then shown to participants in a web application. That server at viewing time then splices together the clips as needed for a seamless viewing experience on the client side.

In the dynamic condition, our viewing application also continuously streams webcam images to a server for the emotion detection. There, we first find the face in the image (using OpenCV's HAAR cascade classifier) and then crop and rescale that area. This data is then processed by a custom model we trained for emotion recognition. Building on the *Xception* [10] architecture, we further train that model with the *FER2013* dataset [22]. This dataset contains 35887 images across seven emotion dimensions, including "happy" (8989 images), "neutral" (6198 images), and—most relevant for our purposes—"fear" (5121 images). We only used images for these three emotions (20308 images overall), reserving 20% for testing and using the rest for training. We trained for 25 epochs with a batch size of 32 and made use of data augmentation[10] to increase the amount of training data. Cross-validation with the resulting model showed an overall accuracy of 75% on the test set (62% for fear, 94% for happy, and 59% for neutral). The emotions detected by this model are streamed back to the viewing application.

On the client side, we cache emotion labels during a clip and then compute overall emotion when it is time to request the next one. We label a viewing of a clip as "fearful" if 10% or more of the webcam frames were labeled as fear and as "neutral" otherwise. We do this because scared reactions can be short and we deem some presence of fear to be sufficient for a scary viewing experience overall. This approach also has the benefit of alleviating low emotion labeling accuracy. As some time is needed to request the next clip and queue it up in the stream, we send this request four seconds before a clip ends.

Table 1: Participants answered 11 questions across three different factors: enjoyment, curiosity, and suspense.

Question	Dimension
The experience was enjoyable	Enjoyment
The experience was exciting	Enjoyment
The experience was pleasant	Enjoyment
During the experience, I felt curious	Curiosity
During the experience, I felt interested	Curiosity
During the experience, I felt stimulated	Curiosity
During the experience, I felt bored (R)	Curiosity
I found myself staring at the screen in anticipation	Suspense
I found myself wishing for a particular story outcome	Suspense
The story did not affect me (R)	Suspense
Some moments were rather suspenseful	Suspense

Note: (R) = reverse scored

4.3 Measures

Our adaptation of Little Red Riding Hood constitutes an interactive storytelling experience. For such experiences, Roth has described [45] a range of concepts and measures from which we draw for our own evaluation. Namely, we adapt items on *enjoyment* [45, p. 45], *curiosity* [45, p. 47], and *suspense* [45, p. 51]. As not all items were applicable for our form of interactive storytelling, we did not include the full scales, but selected the items appropriate for our situation (see Table 1). Participants rated these items on 5-point Likert scales.

4.4 Procedure

All participants first provided informed consent where we also let them know of the study purpose and that they could stop participating at any time. With our experience designed for scariness, this

[10]Using Tensorflow's ImageDataGenerator.

was important so participants would not feel stuck in an unpleasant situation or out of control. This also addresses the main ethical concerns around uncomfortable interactions, previously raised by Benford et al. [5]. Participants then answered some questions on their video viewing preferences. We asked how often they watch short (<30 minutes), medium (30–60 minutes) and long (>60 minutes) videos, and whether they enjoy (yes/no answers) the genres of horror, thriller, and fairy tale. Participants in the personalized condition were also asked to rate a selection of fears on a five-point scale (Low–High). For each fear (spiders, snakes, gore, heights, deep water) we showed four representative images and asked "How uncomfortable do these images make you?" This data then determines which kind of personalizations the participants see, with the clips to use based on their highest ranked fears among the set used for the interludes (i.e., spiders, snakes, or gore) and the ones used for the endings (i.e., heights or deep water). The distribution between the six personalized variants was not uniform with 13 participants most afraid of spiders and heights, followed by spiders and drowning (7), gore and drowning (4), gore and heights (3), snakes and drowning (3), and snakes and heights (1). After watching the video for their assigned conditions, participants then answered the 11 questions on the viewing experience. Participants could also leave comments on the form as well as during the post-experience conversations we had with them.

5 RESULTS

First, we checked for reliability of the used scale by computing Cronbach's alpha values. The resulting alpha values were acceptable for enjoyment (0.56), curiosity (0.75), and suspense (0.53). We thus compute and retain all three measures for subsequent analyses, also referring to Schrepp's [47] recommendations for handling of alpha values in studies of user experience. The individual responses (see Figure 8) show some variation within each of the three measures. For example, within the enjoyable measure (first three questions) the personalized videos were seen as more exciting, but as less pleasant than the baseline, possibly due to our personalizations focusing on making content scarier.

For each measure we then fit a linear model with video condition and the three viewing preferences as predictors. We then used stepwise model selection to determine the best fitting model. All three models include `condition` and `horror` as predictors as well as their interaction, `condition:horror`. For enjoyment and curiosity, `fairytale`, `thriller` as well as the interaction of `horror:thriller` are included. Furthermore, the curiosity model also contains `fairytale:thriller` and `condition:fairytale`. Table 3 in the appendix provides full details on these models.

Table 2 shows the results of ANOVAs for the three models. This shows that which factors significantly influenced a measure varies across them. The video condition on its own was a significant predictor for curiosity and suspense. Figure 9 also shows these main effects, with differences most pronounced for suspense ratings. Post-hoc testing with pairwise t-tests and holm-bonferroni correction showed a significant difference between the baseline and the dynamic condition for suspense ($p < 0.01$).

Our analysis showed several main effects for genre preferences. As shown in Figure 10, liking any of the three genres overall lead to an increase in enjoyment and curiosity. Post-hoc t-tests with holm-bonferroni correction showed significant differences for viewers liking horror on enjoyment ($p < 0.01$) and curiosity ($p = 0.02$) as well as for viewers liking thrillers on curiosity ($p = 0.02$).

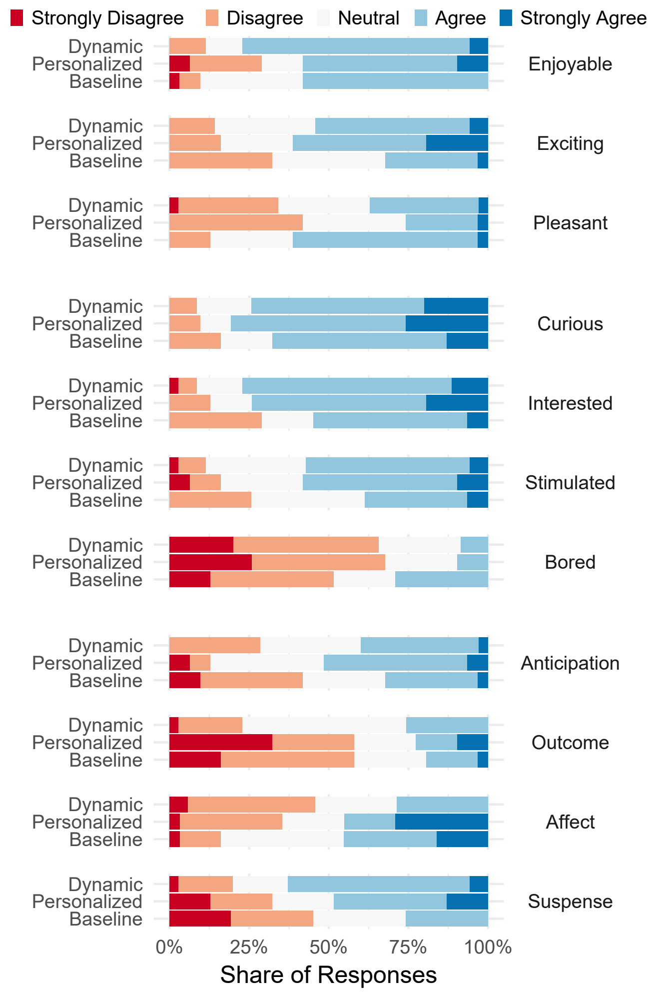

Figure 8: Distribution of participant responses for the three video watching conditions. Questions per Table 1.

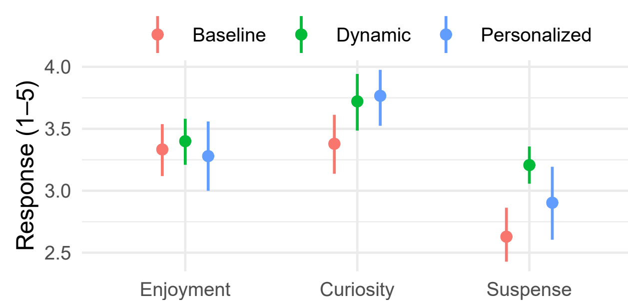

Figure 9: Differences between the three video watching conditions on the three experience measures.

Table 2: Results from separate ANOVAs for the three measures of video watching experience.

Measure	Factor	SS	df	F	p-value	
Enjoyment	Condition	1.02	2	1.62	0.20	
Enjoyment	Fairytale	1.10	1	3.50	0.065	
Enjoyment	Horror	4.64	1	14.78	< 0.001	***
Enjoyment	Thriller	1.85	1	5.90	0.017	*
Enjoyment	Horror:Thriller	5.24	1	16.67	< 0.001	***
Enjoyment	Condition:Horror	1.50	2	2.38	0.099	
Curiosity	Condition	2.44	2	3.18	0.047	*
Curiosity	Fairytale	0.60	1	1.58	0.21	
Curiosity	Horror	1.72	1	4.49	0.037	*
Curiosity	Thriller	2.56	1	6.69	0.011	*
Curiosity	Horror:Thriller	1.55	1	4.04	0.048	*
Curiosity	Condition:Horror	2.39	2	3.12	0.049	*
Curiosity	Fairytale:Thriller	1.27	1	3.32	0.072	
Curiosity	Condition:Fairytale	1.73	2	2.25	0.11	
Suspense	Condition	5.39	2	6.37	0.003	**
Suspense	Horror	1.12	1	2.64	0.11	
Suspense	Condition:Horror	2.02	2	2.39	0.097	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

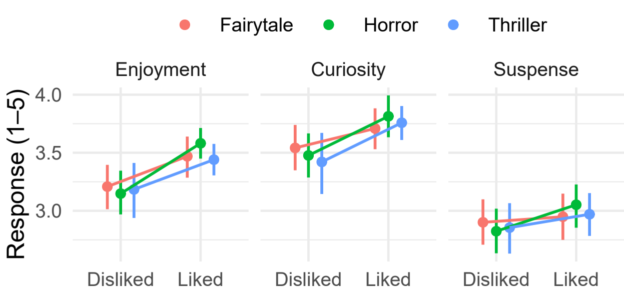

Figure 10: Participants' genre preferences had an effect on their level of enjoyment and curiosity.

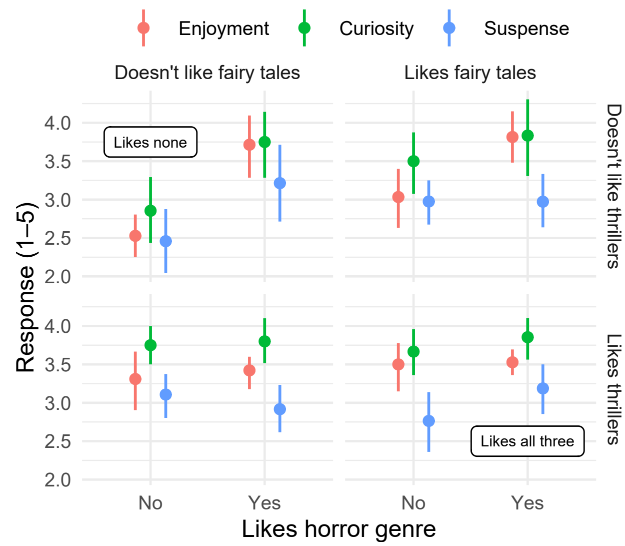

Figure 11: The three genre preferences overlap with participants liking none, some, or all of them. This results in differences in the viewer experience measures.

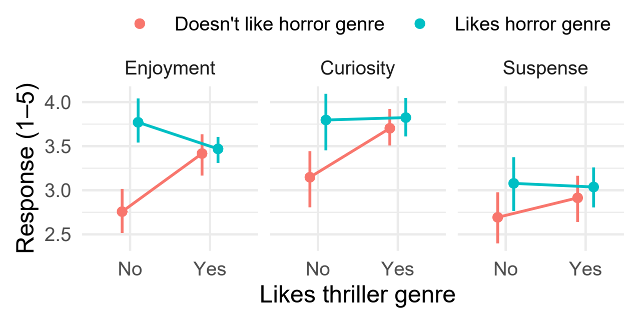

Figure 12: There were significant interactions between preferences for the thriller and horror genres.

In addition to effects of individual genre preferences, there are also interactions between them. In particular, our ANOVA results (see Table 2), showed a significant interaction effect for `horror:thriller`. Post-hoc t-tests with holm-bonferroni correction showed several significant differences. For enjoyment, people disliking all three genres differed from those liking just thrillers ($p < 0.001$), just horror ($p < 0.001$) as well as both horror and thrillers ($p < 0.001$). The same holds for curiosity with $p = 0.02$, $p = 0.02$, and $p < 0.01$ respectively. As Figure 11 shows, people disliking all three genres in general stand out from all other combinations of genre preference. The effect of horror and thriller preferences was most pronounced, as also shown in Figure 12. Noteworthy here is that viewers who like just horror enjoyed the experience more than those who also liked thrillers, albeit at a non-significant level ($p = 0.06$).

Condition effects are also modulated by genre preferences, in particular with respect to curiosity. The only significant predictor here was `condition:horror`, which we further investigated with post-hoc t-tests with holm-bonferroni correction. This showed a significant difference in curiosity ratings between viewers not liking horror watching the baseline condition and viewers liking horror watching the personalized condition ($p = 0.04$). There also was a significant difference in suspense between the same baseline viewers and viewers of the dynamic condition who did not like horror ($p = 0.03$). In the former case, personalized video increased curiosity by 0.7 points and in the latter case the dynamic video increased suspense by 0.66 points. If we focus only on those participants liking horror and compare conditions, pairwise t-tests with holm-bonferroni correction show a significant difference between the baseline and the personalized variants for curiosity ($p < 0.01$). For participants not liking horror, the same test shows significant differences between the dynamic variant and the baseline ($p = 0.01$) and personalized ($p0.02$) variants for suspense. If people liked horror the personalized variant provided the most curiosity inducing experience and if they did not like horror then the dynamic condition created the most suspense.

Finally, we took a closer look at the shown sequences in the dynamic condition. In general, the actually watched clips tended towards scarier versions as few participants exhibited signs of fear when watching the baseline version. We then checked whether the mean scariness level and the variance in scariness level had an influence on the responses using linear models. Neither the former ($p = 0.50$ for enjoyment, $p = 0.24$ for curiosity, and $p = 0.18$ for suspense) or the latter ($p = 0.14$ for enjoyment, $p = 0.07$ for curiosity, and $p = 0.22$ for suspense) showed a significant effect. This confirms that participants actually had a fairly uniform experience in the dynamic condition, as also indicated by the average clip scariness of 1.6 on a 0 (neutral) to 2 (very scary) scale.

5.1 Participant Comments

Only a few participants used the opportunity to provide written comments at the end of the study. These were 16 participants, similarly spread across the conditions (4 baseline, 5 dynamic, and 7 personalized). We list them here, annotated by the participant age, gender, and which video condition they were a part of.

Many of the participants' comments were on the visual style of the video: "The way the images were merged/changing was a bit uneasy on the eyes" (F24, baseline), "The visuals of the video was different in a good way" (M22, baseline), "The only reason I found the video interesting,was because of the shifting images. It a well-known story, so the pictures was the only new thing for me" (M23, baseline), "The art style of the video was something completely unique, in a good way" (M26, dynamic), and "I found the images the most interesting part of the experience" (M26, personalized).

On the other hand, the video style also was distracting and disorienting for others: "It was somewhat disorienting" (F24, dynamic), "I felt it was a bit uncomfortable with switches between images that didn't quite match with the previous image. A bit like stuttering" (M24, dynamic), "I honestly got very distracted by the (I assume AI generated) picture of a wolf which suddenly had two lower jaws. I felt more like a disjoint and surreal video than an engaging story. I was honestly mostly curious about what would be generated next — I cannot recall many of the details of the story." (M29, personalized), and "It was somewhat disorienting" (F24, dynamic).

Two participants also commented on the audio, noting that "The voice didn't quite match the expression in the video" (M24, dynamic), and desiring "some kind of underlying music, creating a bit more suspense or something?" (M21, personalized).

Lastly, a few participants commented on story elements and the overall experience: "The last sea monster came very suddenly without any build up, which was more funny than scary. And the wolf was more goofy than scary" (M23, personalized), "The wolf got a cuter expression at one point, which ruined the story a little, because you had to perceive it as dangerous" (M24, dynamic), "I did not like big spider :(" (M27, personalized), "Poor Little Red Riding Hood :(" (F27, personalized), and "Exciting" (M22, personalized).

6 DISCUSSION

How much participants enjoyed and got engaged by our story depended on the version they saw, but also on their genre preferences in general. Given the subject matter as well as our manipulation around fears and scariness, this influence of genre preference is not surprising. This is, for example, in line with results by Thompson et al. [51], who found that genre preference affected enjoyment of as well as attention and connection to a movie. We also found an effect of genre preference on curiosity, with people liking horror experiencing more curiosity. Work by Clasen et al. [11] provides clues to the potential reasons behind this link, as they found that people liking the horror genre in general are more sensation seeking. This is also mentioned by Hoffner and Levine [26] in their meta-analysis on the enjoyment of fright. Furthermore, they point to aggressiveness and empathy as predictors, with people who are more aggressive and have less concern for others enjoying horror more. Relevant to our results, given our predominantly male sample, is that Hoffner and Levine also report sex and age to be predictors of this enjoyment. Males and adolescents enjoy horror the most, with this enjoyment decreasing for younger as well as older viewers. Thus, our sample likely tended to enjoy horror—and thus also our manipulations—more than the overall population would.

In our study we saw overall effects of condition, but only one significant post-hoc test for the difference between the baseline and the dynamic variant for suspense. Given that the dynamic condition tended towards scarier clips, this is then not necessarily due to the dynamic adaptation, but could just be due to that overall increase in scariness. The differences are more pronounced once we focus on participants with specific genre preferences, horror in particular. Between the conditions, those not liking horror differed in suspense, while those liking horror differed in curiosity. One finding here is that those liking horror rated the personalized variant higher for curiosity. This hints that personalization got people already interested in the genre more engaged. Such an effect is in line with previous findings, such as by Peng et al. [41], who found that personalized animated movies got viewers more intrigued and invested compared to a non-personalized control. Moreover, with their gaze-based video personalization system, Heck et al. [24] saw positive effects of personalization on attention and involvement.

Similarly, we saw that for participants not liking horror the dynamic variant created a feeling of more suspense. A reason for this could be that this condition is less predictable than the other ones, with potential shifts in scariness. As Lehne and Koelsch [33] note, uncertainty is a key component of suspense experiences. Given that the dynamic experience currently only varies scariness, this effect of suspense could be enhanced by introducing additional dynamic components, such as adding eery music depending on the viewers' emotional state.

Instead of personalization aiming at increasing the horror aspects of the story, a more fitting approach for viewers not liking horror might be to personalize other aspects, such as the setting or characters. As systems like EmoWare [52] have demonstrated, personalization can also be performed across videos, where emotional reactions to one determine which one to show next. Combining within-video adaptations like ours with across-video personalization could thus further enhance the viewing experience.

6.1 Limitations

The extent of our story adaptations was limited by the capabilities of the used generative models as well as the amount of manual labor that was still required in the workflow. Due to this, the control over movement in the scenes was limited and some aspects of animation, such as lip-syncing the characters with the audio were not feasible. Furthermore, the generated clips all exhibit a noticeable "AI style" with morphing features and scene objects blending in and out. These kinds of technical limitations of the underlying models are an active area of improvement, so future work will likely be able to create better clips with more control. This in turn will enable a larger number of adaptations, where our cost per variant was too high to implement and test more than the five personalizations and three scariness levels we cover. The focus on one specific story also means that it remains unclear whether the results generalize to other narratives.

Another limitation is our participant sample, which was mostly male and with an average age of 25. This was suitable for our chosen approach to adapt a story for scariness and fears, but it is unclear whether the found effects translate to more general audiences and other stories and themes. As the story itself and the adaptations are intimately connected, the results are also specific to this story. For example, while we incorporate common fears, they are instantiated in specific story moments with the character of Little Red Riding Hood. It remains an open question whether the same addition of, for example, snakes yields similar results in other stories and settings.

6.2 Implications

Reflecting on our initial research questions, we find that generative-AI indeed allowed for complex yet semi-automated content adaptations (**RQ1**) and that such adaptations indeed have an influence on the viewer experience (**RQ2**). Where manual content creation limits the amount of adaptivity in interactive media, use of generative models opens up a much wider range of cost-effective adaptations. Where we already saw an effect of adaptation, we posit that more extensive and more fine-grained adaptations will enhance this further. But this process also raises new issues for content creators, such as how to best define the boundaries for generative models. In our process, we did this with extensive prompt design and common keywords and descriptions for scenes. However, while this is more flexible than traditional video creation, it is not clear this is sufficient for more dramatic adaptations. For example, if we allow viewers to adapt the geographic setting of a narrative (e.g., having it take place in their home town) then different kind of constraints might be needed for different locales.

7 CONCLUSION

Generative models enable new and more flexible ways to create media, which opens up a new space for content adaptation and customization. We have presented a generative version of the fairy tale of Little Red Riding Hood, where the story can (1) dynamically adapt along three levels of scariness, or (2) be extended with sequences personalized to a viewer's fears. The former incorporates camera-based emotion recognition for view-time adaptation, while the latter builds on viewer profiles elicited beforehand. We tested the two variants and a neutral baseline in a user study with 97 participants. Our results show that the viewed version as well as the preferences of the viewers had an effect on the viewer experience. Furthermore, preferences and story variants interact, and people liking the horror genre, for example, differ in their reception from those that do not. These findings show the potential for AI-generated media to cater more closely to viewers, extending and altering stories to make them align with individuals' preferences. As our results are for one specific narrative, future work needs to investigate whether this translates to other settings, but also to other groups of viewers. With the rapid and ongoing improvements of generative models, further customization and adaptation are being opened up. Furthermore, where our process still required substantial manual control and oversight, models such as Runway Research's *Gen-2*[11] are enabling direct generation of video clips with much more fine-grained control over the output. Our work shows a concrete application of this technology and the potential of the resulting media experiences.

REFERENCES

[1] Andrea F. Abate, Aniello Castiglione, Michele Nappi, and Ignazio Passero. 2020. DELEX: A DEep Learning Emotive EXperience: Investigating Empathic HCI. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 14, 8 pages. https://doi.org/10.1145/3399715.3399820

[2] Stewart Agras, David Sylvester, and Donald Oliveau. 1969. The epidemiology of common fears and phobia. *Comprehensive Psychiatry* 10, 2 (1969), 151–156. https://doi.org/10.1016/0010-440X(69)90022-4

[3] Jessica Balanzategui. 2021. 'Disturbing' children's YouTube genres and the algorithmic uncanny. *New Media & Society* (2021), 14614448211049264. https://doi.org/10.1177/14614448211049264

[4] Heather Barber and Daniel Kudenko. 2009. Generation of Adaptive Dilemma-Based Interactive Narratives. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 4 (2009), 309–326. https://doi.org/10.1109/TCIAIG.2009.2037925

[5] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2005–2014. https://doi.org/10.1145/2207676.2208347

[6] Carmen Bisogni, Lucia Cascone, Aniello Castiglione, and Ignazio Passero. 2021. Deep learning for emotion driven user experiences. *Pattern Recognition Letters* 152 (2021), 115–121. https://doi.org/10.1016/j.patrec.2021.09.004

[7] Barry Brown and Louise Barkhuus. 2011. *Changing Practices of Family Television Watching*. Springer London, London, 93–110. https://doi.org/10.1007/978-0-85729-476-0_6

[8] Peggy Chi, Tao Dong, Christian Frueh, Brian Colonna, Vivek Kwatra, and Irfan Essa. 2022. Synthesis-Assisted Video Prototyping From a Document. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 16, 10 pages. https://doi.org/10.1145/3526113.3545676

[9] Peggy Chi, Nathan Frey, Katrina Panovich, and Irfan Essa. 2021. Automatic Instructional Video Creation from a Markdown-Formatted Tutorial. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 677–690. https://doi.org/10.1145/3472749.3474778

[10] Francois Chollet. 2017. Xception: Deep Learning With Depthwise Separable Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[11] Mathias Clasen, Jens Kjeldgaard-Christiansen, and John A. Johnson. 2020. Horror, personality, and threat simulation: A survey on the psychology of scary media. *Evolutionary Behavioral Sciences* 14, 3 (2020), 213–230. https://doi.org/10.1037/ebs0000152

[12] Romain Cohendet, Matthieu Perreira Da Silva, Anne-Laure Gilet, and Patrick Le Callet. 2017. Emotional interactive movie: adjusting the scenario according to the emotional response of the viewer. *EAI Endorsed Transactions on Creative Technologies* 4, 10 (1 2017). https://doi.org/10.4108/eai.4-9-2017.153053

[13] Rossana Damiano, Vincenzo Lombardo, Giulia Monticone, and Antonio Pizzo. 2019. All about face. An experiment in face emotion recognition in interactive dramatic performance. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. 1–7. https://doi.org/10.1109/ACIIW.2019.8925032

[14] Edirlei Soares de Lima, Bruno Feijó, and Antonio L. Furtado. 2020. Adaptive storytelling based on personality and preference modeling. *Entertainment Computing* 34 (2020), 100342. https://doi.org/10.1016/j.entcom.2020.100342

[11] See https://research.runwayml.com/gen2

[15] Edirlei Soares de Lima, Antonio L. Furtado, and Bruno Feijó. 2015. Storytelling Variants: The Case of Little Red Riding Hood. In *Entertainment Computing - ICEC 2015*, Konstantinos Chorianopoulos, Monica Divitini, Jannicke Baalsrud Hauge, Letizia Jaccheri, and Rainer Malaka (Eds.). Springer International Publishing, Cham, 286–300.

[16] Carlos Duarte de Sena Caires. 2007. Towards the interactive filmic narrative—"Transparency": An experimental approach. *Computers & Graphics* 31, 6 (2007), 800–808. https://doi.org/10.1016/j.cag.2007.08.003

[17] Tom Evens, Amandine Henderickx, and Peter Conradie. 2023. Technological affordances of video streaming platforms: Why people prefer video streaming platforms over television. *European Journal of Communication* (2023), 02673231231155731. https://doi.org/10.1177/02673231231155731

[18] Tom Evens, Amandine Henderickx, and Lieven De Marez. 2021. Generation Stream: The Audiovisual Repertoire of Teenagers. *Journal of Broadcasting & Electronic Media* 65, 2 (2021), 185–204. https://doi.org/10.1080/08838151.2021.1944146

[19] Oran Gafni, Adam Polyak, Oron Ashual, Shelly Sheynin, Devi Parikh, and Yaniv Taigman. 2022. Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 89–106.

[20] David Geerts, Evert van Beek, and Fernanda Chocron Miranda. 2019. Viewers' Visions of the Future. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford (Manchester), United Kingdom) *(TVX '19)*. Association for Computing Machinery, New York, NY, USA, 59–69. https://doi.org/10.1145/3317697.3323356

[21] Stephen Gilroy, Julie Porteous, Fred Charles, and Marc Cavazza. 2012. Exploring Passive User Interaction for Adaptive Narratives. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (Lisbon, Portugal) *(IUI '12)*. Association for Computing Machinery, New York, NY, USA, 119–128. https://doi.org/10.1145/2166966.2166990

[22] Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. 2013. Challenges in Representation Learning: A Report on Three Machine Learning Contests. In *Neural Information Processing*, Minho Lee, Akira Hirose, Zeng-Guang Hou, and Rhee Man Kil (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 117–124.

[23] Fabrizio Guerrini, Nicola Adami, Sergio Benini, Alberto Piacenza, Julie Porteous, Marc Cavazza, and Riccardo Leonardi. 2017. Interactive Film Recombination. *ACM Trans. Multimedia Comput. Commun. Appl.* 13, 4, Article 52 (aug 2017), 22 pages. https://doi.org/10.1145/3103241

[24] Melanie Heck, Janick Edinger, Jonathan Bünemann, and Christian Becker. 2021. The Subconscious Director: Dynamically Personalizing Videos Using Gaze Data. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 98–108. https://doi.org/10.1145/3397481.3450679

[25] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. 2022. Imagen Video: High Definition Video Generation with Diffusion Models. arXiv:2210.02303 [cs.CV]

[26] Cynthia A. Hoffner and Kenneth J. Levine. 2005. Enjoyment of Mediated Fright and Violence: A Meta-Analysis. *Media Psychology* 7, 2 (2005), 207–237. https://doi.org/10.1207/S1532785XMEP0702_5

[27] Ziwei Ji, Yan Xu, I-Tsun Cheng, Samuel Cahyawijaya, Rita Frieske, Etsuko Ishii, Min Zeng, Andrea Madotto, and Pascale Fung. 2022. VScript: Controllable Script Generation with Visual Presentation. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Taipei, Taiwan, 1–8. https://aclanthology.org/2022.aacl-demo.1

[28] Levon Khachatryan, Andranik Movsisyan, Vahram Tadevosyan, Roberto Henschel, Zhangyang Wang, Shant Navasardyan, and Humphrey Shi. 2023. Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators. arXiv:2303.13439 [cs.CV]

[29] Alexis Kirke, Duncan Williams, Eduardo Miranda, Amanda Bluglass, Craig Whyte, Rishi Pruthi, and Andrew Eccleston. 2018. Unconsciously interactive Films in a cinema environment—a demonstrative case study. *Digital Creativity* 29, 2-3 (2018), 165–181. https://doi.org/10.1080/14626268.2017.1407344

[30] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-Scale Text-to-Image Generation Models for Visual Artists' Creative Works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 919–933. https://doi.org/10.1145/3581641.3584078

[31] Alex Lambert, Bjorn Nansen, and Michael Arnold. 2018. Algorithmic memorial videos: Contextualising automated curation. *Memory Studies* 11, 2 (2018), 156–171. https://doi.org/10.1177/1750698016679221

[32] Angela Y. Lee, Hannah Mieczkowski, Nicole B. Ellison, and Jeffrey T. Hancock. 2022. The Algorithmic Crystal: Conceptualizing the Self through Algorithmic Personalization on TikTok. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 543 (nov 2022), 22 pages. https://doi.org/10.1145/3555601

[33] Moritz Lehne and Stefan Koelsch. 2015. Toward a general psychological model of tension and suspense. *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.00079

[34] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. https://doi.org/10.1145/3491102.3501825

[35] Vivian Liu, Tao Long, Nathan Raw, and Lydia Chilton. 2023. Generative Disco: Text-to-Video Generation for Music Visualization. arXiv:2304.08551 [cs.HC]

[36] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. https://doi.org/10.1145/3526113.3545621

[37] Yisi Liu, Olga Sourina, and Mohammad Rizqi Hafiyyandi. 2013. EEG-Based Emotion-Adaptive Advertising. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 843–848. https://doi.org/10.1109/ACII.2013.158

[38] Zhengxiong Luo, Dayou Chen, Yingya Zhang, Yan Huang, Liang Wang, Yujun Shen, Deli Zhao, Jingren Zhou, and Tieniu Tan. 2023. VideoFusion: Decomposed Diffusion Models for High-Quality Video Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10209–10218.

[39] Sidneyeve Matrix. 2014. The Netflix Effect: Teens, Binge Watching, and On-Demand Digital Media Trends. *Jeunesse: Young People, Texts, Cultures* 6, 1 (2014), 119–138. https://doi.org/10.3138/jeunesse.6.1.119

[40] Arturo Nakasone, Helmut Prendinger, and Mitsuru Ishizuka. 2009. ISRST: GENERATING INTERESTING MULTIMEDIA STORIES ON THE WEB. *Applied Artificial Intelligence* 23, 7 (2009), 633–679. https://doi.org/10.1080/08839510903205423

[41] Fengjiao Peng, Veronica Crista LaBelle, Emily Christen Yue, and Rosalind W. Picard. 2018. A Trip to the Moon: Personalized Animated Movies for Self-Reflection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173827

[42] Marc Ferrer Prada and Josep Valor. 2022. Linear Broadcasting vs Streaming: the jury is still out. https://blog.iese.edu/the-media-industry/2022/03/16/linear-broadcasting-vs-streaming-the-jury-is-still-out/

[43] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art. In *Creativity and Cognition* (Venice, Italy) *(C&C '22)*. Association for Computing Machinery, New York, NY, USA, 15–28. https://doi.org/10.1145/3527927.3532792

[44] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV]

[45] Christian Roth. 2016. *Experiencing Interactive Storytelling*. Ph. D. Dissertation. Vrije Universiteit Amsterdam.

[46] Christian Roth and Hartmut Koenitz. 2019. Bandersnatch, Yea or Nay? Reception and User Experience of an Interactive Digital Narrative Video. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford (Manchester), United Kingdom) *(TVX '19)*. Association for Computing Machinery, New York, NY, USA, 247–254. https://doi.org/10.1145/3317697.3325124

[47] Martin Schrepp. 2020. On the Usage of Cronbach's Alpha to Measure Reliability of UX Scales. *Journal of Usability Studies* 15, 4 (2020), 247–248.

[48] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, and Yaniv Taigman. 2022. Make-A-Video: Text-to-Video Generation without Text-Video Data. arXiv:2209.14792 [cs.CV]

[49] Gábor Szűcs and Modafar Al-Shouha. 2022. Modular StoryGAN with Background and Theme Awareness for Story Visualization. In *Pattern Recognition and Artificial Intelligence*, Mounîm El Yacoubi, Eric Granger, Pong Chi Yuen, Umapada Pal, and Nicole Vincent (Eds.). Springer International Publishing, Cham, 275–286.

[50] Tan Tang, Junxiu Tang, Jiewen Lai, Lu Ying, Yingcai Wu, Lingyun Yu, and Peiran Ren. 2022. SmartShots: An Optimization Approach for Generating Videos with Data Visualizations Embedded. *ACM Trans. Interact. Intell. Syst.* 12, 1, Article 4 (mar 2022), 21 pages. https://doi.org/10.1145/3484506

[51] Jacqueline M. Thompson, Ben Teasdale, Sophie Duncan, Evert van Emde Boas, Felix Budelmann, Laurie Maguire, and Robin I. M. Dunbar. 2021. Differential effects of film genre on viewers' absorption, identification, and enjoyment. *Psychology of Aesthetics, Creativity, and the Arts* 15, 4 (2021), 697–709. https://doi.org/10.1037/aca0000353

[52] Abhishek Tripathi, T. S. Ashwin, and Ram Mohana Reddy Guddeti. 2019. EmoWare: A Context-Aware Framework for Personalized Video Recommendation Using Affective Video Sequences. *IEEE Access* 7 (2019), 51185–51200. https://doi.org/10.1109/ACCESS.2019.2911235

[53] Laura Varisco and Giulio Interlandi. 2020. Designing Unconscious and Enactive Interaction for Interactive Movie Experience. In *Distributed, Ambient and Pervasive Interactions*, Norbert Streitz and Shin'ichi Konomi (Eds.). Springer International Publishing, Cham, 362–375.

[54] Ruben Villegas, Mohammad Babaeizadeh, Pieter-Jan Kindermans, Hernan Moraldo, Han Zhang, Mohammad Taghi Saffar, Santiago Castro, Julius Kunze, and Dumitru Erhan. 2022. Phenaki: Variable Length Video Generation From Open Domain Textual Description. arXiv:2210.02399 [cs.CV]

[55] Miao Wang, Guo-Wei Yang, Shi-Min Hu, Shing-Tung Yau, and Ariel Shamir. 2019. Write-a-Video: Computational Video Montage from Themed Text. *ACM Trans. Graph.* 38, 6, Article 177 (nov 2019), 13 pages. https://doi.org/10.1145/3355089.3356520

[56] Tan Wang, Linjie Li, Kevin Lin, Yuanhao Zhai, Chung-Ching Lin, Zhengyuan Yang, Hanwang Zhang, Zicheng Liu, and Lijuan Wang. 2024. DisCo: Disentangled Control for Realistic Human Dance Generation. arXiv:2307.00040 [cs.CV]

[57] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. 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 22, 29 pages. https://doi.org/10.1145/3544548.3581402

[58] Yipeng Yu, Zirui Tu, Longyu Lu, Xiao Chen, Hui Zhan, and Zixun Sun. 2021. Text2Video: Automatic Video Generation Based on Text Scripts. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM '21)*. Association for Computing Machinery, New York, NY, USA, 2753–2755. https://doi.org/10.1145/3474085.3478548

[59] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543 [cs.CV]

[60] Sicheng Zhao, Hongxun Yao, Yue Gao, Rongrong Ji, Wenlong Xie, Xiaolei Jiang, and Tat-Seng Chua. 2016. Predicting Personalized Emotion Perceptions of Social Images. In *Proceedings of the 24th ACM International Conference on Multimedia* (Amsterdam, The Netherlands) *(MM '16)*. Association for Computing Machinery, New York, NY, USA, 1385–1394. https://doi.org/10.1145/2964284.2964289

A SCRIPT

The three conditions include different kinds of clips. The *baseline* condition only includes the "neutral" variant clips. The *dynamic* condition includes clips with "neutral", "scary", or "very scary" variants and the specific video is picked dependent on the user state. In the personalized condition, the "neutral" variant clips are supplemented with personalized clips. These are two interlude clips that show either "snakes", "spiders", or "gore", a transition clip, and one of two personalized endings: "drowning" or "falling". Also see Figure 2 for an illustration of the different variants.

A.1 Main Story

Clip 1 (13 seconds).

Narrator: "Little Red Riding Hood went through the forest with the little basket in her hand. She wanted to pick some flowers for her grandmother, so she went off the road to find some. (Pause 1sec) She looks down on the forest floor."
Visuals: Little red riding hood that walk through a forest path with a basket in her hand.
Variants: neutral, scary, very scary

Clip 2 (6 seconds).

Narrator: <silent>
Visuals: The forest floor.
Variants: neutral, scary, very scary

<u>Personalized Interlude</u> Clip 1 (21–26 seconds).

Narrator: "Unbeknownst to the girl. She wasn't the only one interested in the flower."
Visuals: She bends down to grab it, but is startled by what she fears the most.
Variants: spiders, snakes, gore

Clip 3 (10 seconds).

Narrator: "With some flowers in her basket, she went back to the road, leading to her grandmother's house. "
Visuals: Little red riding hood walking from flower spot, back to the path.
Variants: neutral, scary, very scary

Clip 4 (7 seconds).

Narrator: <silent>
Visuals: Little red riding hood walking on a forest path, the forest is now more dense than before.
Variants: neutral, scary, very scary

Clip 5 (4 seconds).

Narrator: "Little Red Riding Hood looks at the forest scenery as she goes further and further into the forest."
Visuals: Panning over the treeline, consisting of some bushes.
Variants: neutral, scary, very scary

Clip 6 (3 seconds).

Narrator: " On her way through the forest she hears something rustling in the bushes."
Visuals: Image of a bush, with glowing eyes.
Sounds: Branch snapping.
Variants: neutral, scary, very scary

Clip 7 (5 seconds).

Narrator: "The big bad wolf emerges from the bushes in front of little red riding hood. She lets out a little yelp."
Visuals: Image of a wolf.
Sounds: Wolf growl.
Variants: neutral, scary, very scary

Clip 8 (4 seconds).

Narrator: "Hello little girl, where are you headed? The wolf asked in a friendly voice."
Visuals: Image of a wolf.
Variants: neutral, scary, very scary

Clip 9 (8 seconds).

Narrator: "Little Red Riding Hood gulped and composed herself. I'm on my way to visit my sick Grandmother. My mother brought me some bread and honey to give to her."
Visuals: Images of little red riding hood.
Variants: neutral, scary, very scary

Clip 10 (8 seconds).

Narrator: "The wolf nods compassionately, and says: Your Grandmother is a lucky woman, little girl. Where does she live, exactly?"
Visuals: Image of the wolf.
Variants: neutral, scary, very scary

Clip 11 (5 seconds).

Narrator: "She lives just across the forest in the first little house in the village — the one with the white door."
Visuals: Image of a forest tree lodge, with a white door.
Variants: neutral, scary, very scary

Clip 12 (9 seconds).

Narrator: "Hmm, interesting, the wolf uttered under his breath. I will leave you to your plans, and wish your Grandmother a speedy recovery from me."
Visuals: Image of the wolf smiling.
Variants: neutral, scary, very scary

Clip 13 (9 seconds).

Narrator: "Certainly, Mr. Wolf. You are very friendly, but I will get going, I bid you farewell."
Visuals: Image of little red riding hood.
Variants: neutral, scary, very scary

Clip 14 (10 seconds).

Narrator: "Goodbye little girl. Wait, look, do you see that road? The wolf points and says: Just follow that road and you will get to your grandmother's house."
Visuals: Image of a forest road.
Variants: neutral, scary, very scary

Clip 15 (5 seconds).

Narrator: "Yes, I do."
Visuals: Image of little red riding hood.
Variants: neutral, scary, very scary

Clip 16 (5 seconds).

Narrator: "Little red riding hood smiled and said Thank you mr. wolf as she disappeared from the wolf's sight."
Visuals: Image of little red riding hood walking down a forest path.
Variants: neutral, scary, very scary

Clip 17 (6 seconds).

Narrator: "The wolf smirked as she naively strolled on the longer road."
Visuals: Image of wolf being suspicious.
Variants: neutral, scary, very scary

A.2 Story Extension in Personalized Condition

Personalized Transition Clip 1 (10 seconds).

Narrator: "After walking for a while, Little Red Riding Hood didn't feel like she was headed for her grandmas. I just hope I won't end at the part of the forest my mother warned me about, the girl thought to herself"
Visuals: —
Variants: neutral

Personalized Transition Clip 2 (3 seconds).

Narrator: "What was that?"
Visuals: —
Variants: neutral

Personalized Transition Clip 3 (5 seconds).

Narrator: <silent>
Visuals: Zooming in on some dense bushes*
Variants: neutral

Personalized Interlude Clip 2 (29-33 seconds).

Narrator: <silent>
Visuals: Little Red Riding Hood watch through the bushes in horror, witnessing her deepest fears
Variants: spiders, snakes, gore

Personalized Transition Clip 4 (4 seconds).

Narrator: "She runs…terrified of what she saw creeping, behind the trees"
Visuals: —
Variants: neutral

A.3 Added Endings in Personalized Condition

Personalized Ending Clip A-1 (15 seconds).

Narrator: "A strong river stream, surges below Little Red Riding Hood, as She decides to cross a river. The tree log she walks on feels unstable, she can feel how old and rotten it is. She slips!"
Visuals: —
Variants: drowning

Personalized Ending Clip A-2 (7 seconds).

Narrator: "She screams and cries as she is now at the mercy of the strong river stream."
Visuals: —
Variants: drowning

Personalized Ending Clip A-3 (12 seconds).

Narrator: "It comes to a stop. She is now in a huge lake, tired from struggling against the current. With little energy left, she can barely stay above water…"
Visuals: —
Variants: drowning

Personalized Ending Clip A-4 (8 seconds).

Narrator: "Slowly fading underwater in the vast lake…her lungs feel heavyldots she thinks of her mother. one last time."
Visuals: —
Variants: drowning

Personalized Ending Clip A-5 (4 seconds).

Narrator: <silent>
Visuals: Sea monster watches Little Red Riding Hood
Variants: drowning

Personalized Ending Clip B-1 (15 seconds).

Narrator: "A cliffside appeared as the forest ended. The girl, franticly threw herself down a slope, instead of falling straight to her demise."
Visuals: —
Variants: falling

Personalized Ending Clip B-2 (7 seconds).

Narrator: "Unfortunately... The hillside had wide chasms, and the girl caught one. Holding on for her dear life... but without the strength, she lets go."
Visuals: —
Variants: falling

Personalized Ending Clip B-3 (4 seconds).

Narrator: <silent>
Visuals: Falling into the hole
Variants: falling

Personalized Ending Clip B-4 (12 seconds).

Narrator: "The wide hole, becomes narrow and rugged with stones…Her blood is spilled as she collides.. but she keeps falling."
Visuals: —
Variants: falling

Personalized Ending Clip B-5 (8 seconds).

Narrator: "An opening…finally appears. And she can rest"
Visuals: —
Variants: falling

B PROMPT EXAMPLE

Neutral (baseline), scary, and very scary versions of the prompt used for clip 10, where the wolf talks to Little Red Riding Hood.

Neutral, Positive Prompt. nervous, tense, scared, looking nervous, looking scared Close up of cunning wolf looking sly with a forest in the background in the style of thomas kinkade

Neutral, Negative Prompt. bad anatomy, bad proportions, blurry, cloned face, cropped, deformed, dehydrated, disfigured, duplicate, error, extra arms, extra fingers, extra legs, extra limbs, fused fingers, gross proportions, jpeg artifacts, long neck, low quality, lowres, malformed limbs, missing arms, missing legs, morbid, mutated hands, mutation, mutilated, out of frame, poorly drawn face, poorly drawn hands, signature, text, too many fingers, ugly, username, watermark, worst quality, snow, smiling, big eyes, furry wolf, furry

Scary, Positive Prompt. cinematic, colorful background, concept art, dramatic lighting, high detail, highly detailed, intricate, intricate sharp details, octane render, smooth, studio lighting, trending on artstation, high resolution, best resolution, 8k, crazy, aggressive, eye scar, scar, red eyes, evil, nightmare fuel, crazy eyes, slightly open mouth, deforestation, dead trees, small fangs Closeup of grey creepy wolf with red eyes, slightly open mouth

Scary, Negative Prompt. bad anatomy, bad proportions, blurry, cloned face, cropped, deformed, dehydrated, disfigured, duplicate, error, extra arms, extra fingers, extra legs, extra limbs, fused fingers, gross proportions, jpeg artifacts, long neck, low quality, lowres, malformed limbs, missing arms, missing legs, morbid, mutated hands, mutation, mutilated, out of frame, poorly drawn face, poorly drawn hands, signature, text, too many fingers, house, buildings, taverns ugly, username, orange eyes, watermark, worst quality, lantern, asphalt, asphalt road, lantern, stairs, colorful, green, leaves, reflection, castles, fog, decorations, symbols, paint, jewelry, indian objects, spiritual objects, spiritual painting, yellow eyes, grass, life

Very Scary, Positive Prompt. red eyes, fangs, big fangs, biting, evil, nightmare fuel, closeup of a very aggressive and scary wolf with red eyes and big fangs

Very Scary, Negative Prompt. green forest, green leaves, green bushes, tongue

Table 3: Regression table for models fitted to predict enjoyment, curiosity, and suspense ratings across the story variants.

	Dependent variable:		
	Enjoyment	Curiosity	Suspense
	(1)	(2)	(3)
Constant	2.599^{***}	3.045^{***}	2.579^{***}
	(2.269, 2.929)	(2.607, 3.483)	(2.287, 2.871)
ConditionDynamic	0.250	−0.003	0.659^{**}
	(−0.103, 0.603)	(−0.511, 0.506)	(0.250, 1.067)
ConditionPersonalized	−0.294	−0.355	0.004
	(−0.677, 0.090)	(−0.888, 0.179)	(−0.436, 0.445)
Fairytale	0.216	0.071	
	(−0.010, 0.441)	(−0.459, 0.602)	
Horror	1.086^{***}	0.251	0.129
	(0.597, 1.574)	(−0.291, 0.793)	(−0.341, 0.599)
Thriller	0.719^{***}	0.813^{***}	
	(0.408, 1.030)	(0.388, 1.238)	
HorrorTRUE:Thriller	-0.983^{***}	-0.535^{*}	
	(−1.455, −0.511)	(−1.056, −0.013)	
ConditionDynamic:Horror	−0.345	0.256	−0.200
	(−0.901, 0.211)	(−0.358, 0.871)	(−0.841, 0.440)
ConditionPersonalized:Horror	0.267	0.783^{*}	0.490
	(−0.301, 0.835)	(0.155, 1.411)	(−0.166, 1.147)
FairytaleTRUE:Thriller		−0.475	
		(−0.987, 0.036)	
ConditionDynamic:Fairytale		0.470	
		(−0.133, 1.072)	
ConditionPersonalized:Fairytale		0.651^{*}	
		(0.029, 1.273)	
Observations	97	97	97
R^2	0.348	0.311	0.184
Adjusted R^2	0.289	0.222	0.139
Residual Std. Error	0.561 (df = 88)	0.619 (df = 85)	0.650 (df = 91)
F Statistic	5.880^{***} (df = 8; 88)	3.486^{***} (df = 11; 85)	4.093^{**} (df = 5; 91)
Note:			*p<0.05; **p<0.01; ***p<0.001