Story-Driven: Real-time Context-Synchronized Storytelling in Mobile Environments

Jan Henry Belz
jan_henry.belz@porsche.de
Porsche AG
Stuttgart, Germany
University of Ulm
Ulm, Germany

Lina Madlin Weilke
lina.weilke1@porsche.de
Porsche AG
Stuttgart, Germany

Anton Winter
anton.winter@porsche.de
Porsche AG
Stuttgart, Germany

Philipp Hallgarten
philipp.hallgarten1@porsche.de
Porsche AG
Stuttgart, Germany

Enrico Rukzio
enrico.rukzio@uni-ulm.de
University of Ulm
Ulm, Germany

Tobias Grosse-Puppendahl
tobias.grosse-puppendahl@porsche.de
Porsche AG
Stuttgart, Germany

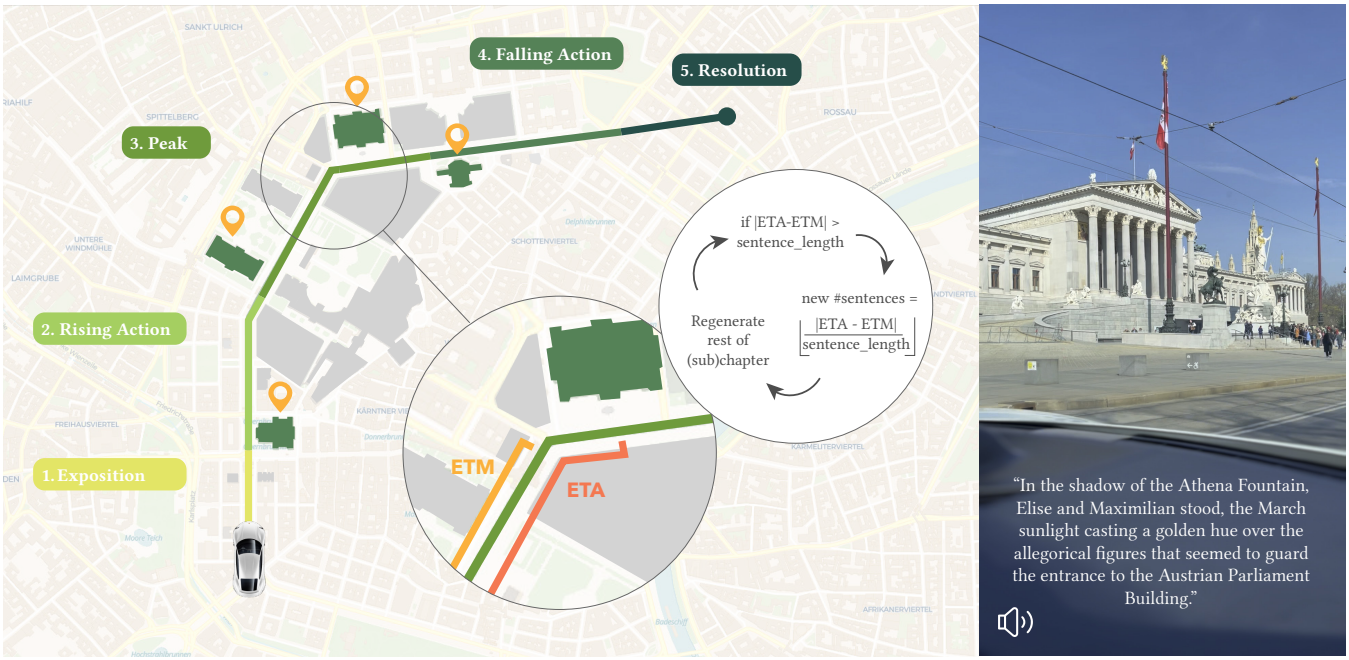


Figure 1: *Story-Driven* is a mobile system that creates a context-synchronized storytelling experience for any given journey. A customized audiobook is narrated as the user travels, weaving contextual information into the story. By integrating real-world elements into the story world, the line between the physical and virtual worlds is blurred. This is achieved by calculating the difference between the user's physical Estimated Time of Arrival (ETA) at a location and the Estimated Time to Mention (ETM) of that location in the story. With AI-generated storytelling, the narrative is aligned with the real world in real-time, creating a unique storytelling experience for every journey.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676372

ABSTRACT

Stories have long captivated the human imagination with narratives that enrich our lives. Traditional storytelling methods are often static and not designed to adapt to the listener's environment, which is full of dynamic changes. For instance, people often listen to stories in the form of podcasts or audiobooks while traveling in a car. Yet, conventional in-car storytelling systems do not embrace the adaptive potential of this space. The advent of generative AI

is the key to creating content that is not just personalized but also responsive to the changing parameters of the environment. We introduce a novel system for interactive, real-time story narration that leverages environment and user context in correspondence with estimated arrival times to adjust the generated story continuously. Through two comprehensive real-world studies with a total of 30 participants in a vehicle, we assess the user experience, level of immersion, and perception of the environment provided by the prototype. Participants' feedback shows a significant improvement over traditional storytelling and highlights the importance of context information for generative storytelling systems.

CCS CONCEPTS

• **Human-centered computing → Sound-based input / output**; **Personal digital assistants**.

KEYWORDS

Storytelling, Large-Language-Models, Driving Experience, Auditive Interfaces

ACM Reference Format:
Jan Henry Belz, Lina Madlin Weilke, Anton Winter, Philipp Hallgarten, Enrico Rukzio, and Tobias Grosse-Puppendahl. 2024. Story-Driven: Real-time Context-Synchronized Storytelling in Mobile Environments. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3654777.3676372

1 INTRODUCTION

Storytelling is said to be an ancient human universal, bringing together generations of people since time immemorial [50]. The element of narrative entertainment has evolved since then: the writing and printing of books has given millions of people access to storytelling, and the development of digital audio technology has taken storytelling to the next level by combining the characteristics of professional voice actors with the narrative talent of book authors. The resulting audiobooks and podcasts are the modern manifestation of traditional storytelling and have become a common medium for hands-free entertainment while performing multitasking tasks such as housework or traveling.

Traveling and commuting are often time-consuming processes, with little possibility of using this time for other tasks. Particularly, commuting by car demands the driver's visual and hands-on attention at all times, making longer travel a dull task. Here, audiobooks have emerged as a popular activity for both passengers and drivers [34]. However, traditional auditive means of entertainment do not make use of the many potential advantages that come with the mobility of traveling by car. The interior of a vehicle presents a highly unique design space: being a confined space, it is partially isolated from external influences. Moreover, it is highly mobile, with the environment changing frequently. This results in exceptionally variable context factors along the ride, which can be both a blessing and a curse for context-aware in-vehicle systems.

The emergence of Generative AI applications in recent years has made it possible to create personalized content in any modality: image (e.g., DALL-E[1]), video (Sora[2]), audio (MusicGen[3]), and text generation (ChatGPT[4]) have become much more human-like through the impact of Generative AI. Especially the introduction of the text-based tool *ChatGPT* resulted in an enormous public interest by providing access to Generative AI for a non-technical audience. Current large language models (LLMs) show great performance as conversational interfaces, excelling in their ability to interpret natural language. Besides mere informational question-answering, LLMs are also able to provide great entertaining value by generating content that is tailored exactly to the user's needs: Video games can provide more immersive player interaction [14, 27], quiz applications adjust to the user's preferences in real-time, and conversational assistants offer a much more human-like experience.

Especially in mobile settings, such as traveling in a car, auditory interfaces benefit from text-based LLMs due to their hands-free and eyes-free interaction. The qualities of LLMs make them very suitable for automated storytelling [11]. They provide an acceptable narrative quality while allowing many possibilities for customization and adaptation.

To make use of this potential, we propose a novel system for context-synchronized storytelling in mobile settings named Story-Driven. By combining the changing environment during a drive with the adaptive qualities of an LLM, the system is able to create an immersive storytelling experience tailored to any given route. Using contextual and geographic information about the vehicle's surroundings, interesting locations along the route are woven into an AI-generated story. GPS and traffic information are used to synchronize the storytelling with the real world, blurring the line between fiction and reality. We propose that this context-synchronized storytelling can be utilized for creating much more believable environments [20] and a more engaging listening experience [51].

Contributions

We enable reproduction of our research by providing datasets and source code to the community[5]. In summary, we provide the following core contributions:

- We present an experienceable end-to-end system for generative storytelling in mobile environments. Designing this system poses the challenge of merging mobile, real-time context into a context-synchronized, immersive experience.
- We introduce a novel method that writes and adopts stories based on contextual information and timing, which we validate in both a simulator setting and in several real-world scenarios.
- We evaluate the system in two real-world user studies (n=5 and n=25) to assess the user experience in comparison to a baseline system.

[1] https://openai.com/dall-e-3, last accessed on 2024/02/19
[2] https://openai.com/sora, last accessed on 2024/02/19
[3] https://musicgen.com/, last accessed on 2024/02/19
[4] https://chat.openai.com/, last accessed on 2024/02/19
[5] See supplementary materials, will be made accessible on GitHub

2 RELATED WORK

2.1 Context in Mobile HCI

Context-aware systems (CAS) have become an important paradigm in the field of mobile Human-Computer Interaction (HCI) [9]. Abowd et al. [1] have pioneered the research on context-aware computing, framing context as *implicit situational information*. In many scenarios, this describes a vast amount of information that becomes relevant. As Baldauf et al. [3] stated, "It is a challenging task to define the word context". Numerous definitions of the term exist and include several factors, such as the user's location, environment, identity, emotional state, and attentional focus [15, 39, 40]. Notably, many of these factors change within a mobile context: Where location is the most frequently used attribute of context [3], it is also the most variable for mobile applications.

Contextual information is collected by so-called sensors, which can be divided into three categories: physical, virtual, and logical sensors [3]. In current mobile applications, physical sensors are the most frequently deployed sensors, e.g., in location-based services (LBS) [37]. With the increase of mobile devices (such as smartphones), LBS have become more popular over the years [3]. Wireless cellular networks provide access to information via the internet to any location in the world, with 5G technology enabling latencies and bandwidths for computation-expensive real-time applications such as large-scale vehicular communication networks [48]. Car-based context-aware applications present the pinnacle of the variability of context information. With vehicles traveling at higher speeds than most mobility solutions, the environment context changes with even higher frequency. This creates both challenges and opportunities. Kari et al. [24] make use of fast-changing environment variables by adjusting in-vehicle music to external affordances (e.g., tunnel entrances). Bethge et al. [6] involve the user's emotional state for navigation purposes, combining internal and external contextual information.

2.2 Narrative Storytelling

Narrative storytelling is an ancient human trait that has been studied across multiple disciplines, such as literature [4], psychology [2], and media studies [38]. How a story is written and told impacts the audience's engagement and attention. For instance, the Dual-Coding Theory by Paivio [32] states that the human brain encodes information in both visual and verbal modalities. If both modalities are utilized, the information is easier to retrieve. This is further supported by the Multimedia Learning Theory [30]. Therefore, a visual-auditive storytelling experience is easier to perceive and remember [23].

Visual storytelling has the power to enhance verbal storytelling through images: "A picture is worth a thousand words." [7]. This is reflected in recent approaches to automate visual storytelling. Huang et al. [22] argue that mere descriptive image-to-text generation is not suitable for natural HCI, as they propose a more narrative approach to achieve conversational human-AI communication. Following this narrative focus, other approaches in the domain have explored the generation of image-based questions as a means to deepen engagement and interactivity. Suwono et al. [43] leverage geographical and contextual information to integrate visual storytelling with question generation. The introduction of multiple modalities to automated storytelling and question generation has shown to create more meaningful and appealing experiences [56].

2.3 Automated Storytelling

Usage of LLMs is widespread among the general public [16] and with their constant technological advancements they can fully automate the process of crafting a narrative or telling a story. Collaborative systems can provide human writers with the assistance of AI-generated content [13, 28, 52, 57] or pave the way for users to automatically generate content to their liking. Researchers have implemented different tools to generate stories with LLMs that cater to users' input, which can take the form of a selection of story elements [33], a number of topics [29], or rough sketches of a plot [35]. Chung et al. [12] proposed an interactive sketching system that allows users to sketch narratives by sketching rough graphs, while an LLM generates the corresponding narrative.
LLMs are not only applicable for user-controlled story generation, but they are also capable of generating high-quality narratives on their own. Though public skepticism towards AI persists, Chu and Liu [11] found that LLMs such as ChatGPT hold the potential to outperform human writers. They can offer progressive narratives and provide innovative plot twists, even though they lack imagination for diverse scenarios and rhetoric [5].

To get the best results from an LLM's story-telling skills, it can be beneficial to structure and plan narratives before generating whole stories. This approach helps provide coherence and impede repetition [42, 55]. Researchers have experimented with differing methods of providing a structured plot, including modeling storylines from image sequences [21]. Yang et al. [54] have publicized an extensive line of work aiming to generate coherent long-form story content with LLMs. They proposed Re3[54], a framework that generates long stories with recursive reprompting and revision, and later on implemented a detailed outline control [53] and a concrete outline control [47] to improve scalability, story coherence, and pacing in automated story-telling.

Automated storytelling is applicable to many different areas. Researchers have experimented with utilizing LLMs in different industries like helping companies create brand stories [46] and designing narratives for educational escape rooms [17]. On a personal level, automated story-telling can create stories from photo albums to facilitate experience sharing [31] or assist with entertaining a child with interactive story-telling [10, 57]. LLMs can also write visual novels, like educational narratives on climate change [19].

3 CONCEPT

The core idea of context-adaptive storytelling is to create a new level of spoken narrative immersion for mobile use by weaving elements of the real world into the fictional narrative. This is achieved by generating an audiobook that takes place in real-world locations that the user eventually passes. By extending and shortening, the story is temporally aligned with the elements of the physical environment, allowing the user to visually associate these places with the auditory locations of the story. This effect is further enhanced by enriching the contextual information with descriptions of the

locations, as well as time and weather data. We hypothesize that by blurring the line between the auditory perception of the fictional story and the visual perception of the surrounding real world, the storytelling experience can be taken to a much more immersive level.

3.1 Technical Overview

To implement the concept of context-synchronized storytelling, a robust way of generating custom stories is necessary to allow for incorporating any given contextual information into the plot. For performance reasons, we opted for OpenAI's GPT-4 API[6]. However, since our prototype is based on a prompt-engineering approach, any given LLM could be deployed with minor adjustments to the prompts. To deliver the story to the user, a text-to-speech (TTS) API by Azure AI Speech Studio[7] was used. Regarding the integration of context information, we decided to implement points of interest (POI) with a visual description, weather information, time of day, and season, as well as the planned route and the current position of the user. For each run, the POIs are selected dynamically along the route and according to their *importance score* (see subsection 3.3). Additional data is then gathered from various online APIs. Based on these POIs and their descriptions, we then prompt GPT-4 to generate a story outline structurally following a traditional narrative arc (exposition, rising action, peak, falling action, resolution) [8]. Additionally, four main characters are pre-generated by GPT-4 since this is a popular choice in existing literature that provides enough storytelling material without overcharging the user. Finally, the API is prompted to fully generate the five chapters of the story according to the approximate length of the route (see subsection 3.4 for the full story preparation).

With the complete story in place, the planning phase is complete and the live phase can begin. At this point, the user will be ready to start navigating. The start of the spoken narration is triggered, where the TTS engine reads the prepared story to the user. From now on, the estimated time of arrival (ETA) at the next POI on the route is calculated continuously to make sure that the generated story is aligned with the physical ETA of the user (for a detailed description, see subsection 3.5). Along the journey, this process is repeated for every segment of the route. For the last segment, the end of the story is adjusted to the end of the journey, so the plot has a coherent ending instead of being interrupted abruptly.

3.2 Designing Context for Mobile Storytelling

The selection of relevant context information for context-aware systems is a crucial step in the design of such systems. To break down the complexity of a mobile context-aware system, we have identified three integral parts, which we describe in the following.

Environment. Mobile systems hold a variety of contextual information, both external (e.g., route, travel velocity, surrounding landscape, location) and internal (e.g., emotional state, cognitive load, fatigue, knowledge about the area). These factors could potentially be woven into the narrative. For our system, we decided

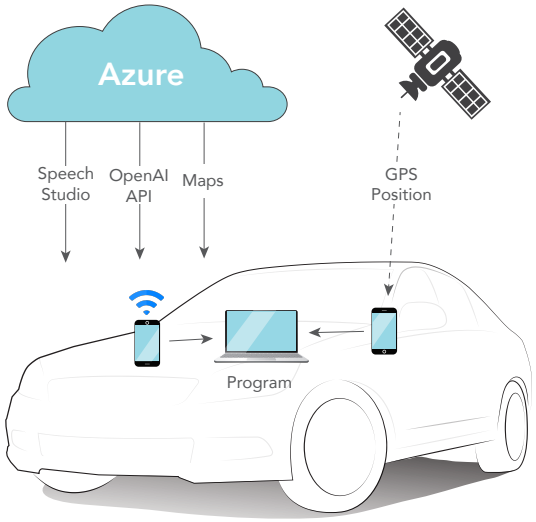


Figure 2: Story-Drivencombines several data streams in a mobile application. The in-vehicle apparatus consists of three mobile devices: one device serves as a mobile router to create a local network, where the main device running the system can communicate with a third, external mobile device for receiving GPS information. Via the cellular connection, several Microsoft Azure APIs are queried for real-time access to a TTS engine, an LLM, and map data.

to focus on a subset of context factors: (1) Incorporating real-world locations into the story was considered to have the greatest impact on the user's perception of the story [43]. We focused on the visual appeal of locations to attract and retain the user's attention, thereby facilitating the visual-auditory connection between the real-world location and the fictional story. These locations are referred to as *points of interest* (POI). (2) Information about the current weather forecast, time of day, and season were selected to align the overall setting of the story to the real world. This data was selected due to its impact on the visual scene that the user would perceive. While it was likely to remain unnoticed by the user due to its subtlety, it rather served the purpose of preventing notable discrepancies between the story setting and the real world (e.g., a story set in a freezing winter narrated on a hot summer day).

Story. Creating a narrative is a complex task and includes an indefinite amount of components that can impact the storytelling experience, e.g., genre, plot, characters, story length, or setting. In the case of prompted, AI-generated stories, we controlled these factors with fixed parameters defined in the prompts. These parameters included the overall genre of the story as well as a plot type from "20 Master Plots" [45] to guide the narrative the LLM would create. Due to the location-based nature of our system, we opted for a *Riddle* plot type as this was most suitable for frequent location changes. Moreover, the story structure followed a traditional

[6]https://openai.com/gpt-4, last accessed on 2024/04/03
[7]https://speech.microsoft.com/portal, last accessed on 2024/04/03
[8]https://www.masterclass.com/articles/what-are-the-elements-of-a-narrative-arc-and-how-do-you-create-one-in-writing, last accessed on 2024/04/03

Figure 3: The context-synchronized storytelling system dynamically selects relevant POIs (highlighted) by calculating a bounding box along the route. All of the POIs overlapping with the bounding box are then filtered according to their importance score and specific threshold values to respect the story's structure, leaving only a relevant selection (marked in green) for the story.

narrative arc to improve comprehensibility. The full story setup is described in subsection 3.4.

Auditory Output. Voice interfaces are led by many parameters, such as tone, accent, reading speed, or emotional emphasis. With recent progress in neural speech synthesis, voice interfaces have become more human-like, allowing for a more natural interaction with computer systems [44]. For the purposes of our context-synchronized storytelling, a human-like state-of-the-art voice was necessary to prevent fatigue and annoyance during longer listening periods. From the Microsoft Azure AI Speech Studio, we used the female voice "Ava", as its "bright, engaging voice" lends itself well to engaging users in a storytelling experience. The neural TTS API is able to synthesize high-quality speech in real-time and automatically emphasize sentences without additional annotation.

3.3 Dynamic POI Selection

From a user's point of view, the initial input to the context-synchronized storytelling system only consists of a route from a starting point to an ending point. Alongside this route, POIs that are relevant to the story are selected through a set of OpenStreetMap (OSM) tags and filtered to fit the story's structure. The quality of an individual POI is determined by their *importance score*, which is calculated by the Nominatim OpenStreetMap API[9]. This score is based on the place's Wikipedia page rank, the overall search rank, and the number of OSM tile views. Additionally, certain threshold values have to be considered during the POI selection: (1) A minimum time interval between two POIs ensures that the storytelling system has enough time to talk about each POI. (2) A maximum temporal distance between two POIs prevents the storytelling system from talking about a location for too long. (3) The first and last chapter of the story (i.e., *exposition* and *resolution*) should not include any real-world locations to allow for an appropriate introduction and conclusion of the story. (4) To guarantee a minimum relevance of every selected location, all POIs below a certain importance threshold are not considered. We empirically found that a value of 0.15 provides great results.

This POI polling process must produce the best possible locations, both in terms of POI quality and story coherence, to allow for an ideal storytelling experience. Therefore, we developed a *global search* procedure that selects the best set of POIs. The algorithm starts by drawing a polygon along the route. Then all relevant, intersecting POIs are queried and filtered. From the remaining POIs, all combinations complying with thresholds (1)-(3) are considered and the set of POIs with the maximum accumulated importance is selected. This approach considers all possible POI combinations and ensures that no POIs with exceptionally high importance scores are left out.

3.4 Story Pre-Generation

Having selected the locations that would be implemented in the context-synchronized storytelling allows for preparing the actual story. OpenAI's GPT-4 API was used for the entire story generation and adjustment. All of the prompts were structured as zero-shot chain-of-thought prompts [26] with the elements of an effective prompt by Giray [18] in mind. To further improve the output of the API, the initial system message passed to the LLM was as follows: "You are a writer of beautiful stories with well-readable phrasing and compelling and creative plots. Your target audience is young adults." We start the story generation by creating a story plan, an approach that was inspired by Yang et al. [54]. They proposed a system that builds a story plan with an LLM to generate coherent long-form story content. We provide all contextual information that we aim to weave into the story (POIs with a description extracted from Wikipedia[10], weather status, time of day, season) to the LLM and instruct it to generate a premise for a story with a specific genre and plot type [45] as well as four central characters to be featured.

Building upon that information, we have the LLM construct a plot outline. In this process, we additionally applied ideas of the Detailed Outline Control (DOC) hierarchical framework [53] for improved scalability. The outline is represented by a hierarchical tree structure that dynamically scales in depth to accommodate longer routes.

[9]https://nominatim.org/release-docs/latest/customize/Importance/-, last accessed on 2024/04/03
[10]https://en.wikipedia.org/, last accessed on 2024/04/03

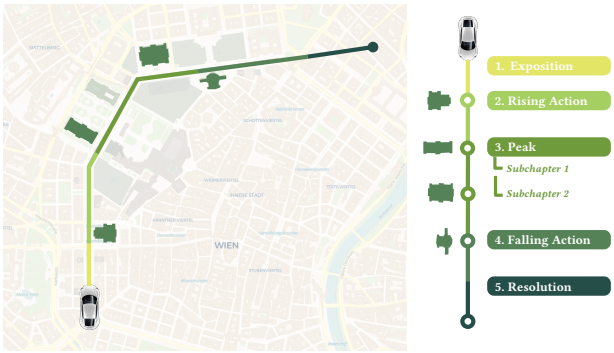


Figure 4: The story's structure follows a hierarchical tree approach [53] to achieve a scalable and flexible storytelling experience. The POIs are divided among the five chapters from the narrative arc and the chapters are split into subchapters if they hold multiple POIs.

The outline is generated with a template pattern [49] to ensure a parsable and reliable output. The pattern provides a five-chapter structure and consists of multiple hierarchies. Each chapter content is described by a short 2-3 sentence summary and the number of POIs that resulted from the dynamic POI selection are divided equally among the three inner chapters of the narrative arc (rising action, peak, falling action). If a chapter holds more than one POI, it is further divided into subchapters, each holding one POI and a short content description that inherits from the chapter's summary. The plot outline was crucial for preserving story coherence later on. The prompts used for this story setup can be found in the Appendix A.1. The duration of a full generation process heavily depended on the API version used and ranged from 40 seconds to five minutes.

After the preparation is completed, the actual story content generation process is initiated. For each chapter, the LLM is given the story's premise, characters, plot outline, contextual information (including POIs), and the summary of the previous chapter. The contextual information also contains a first approximation of the chapter's length, estimated according to the calculated ETA from the POI polling process. If a chapter includes more than one POI, a subchapter is generated for each individual POI. Since our LLM became less reliable with the generation of long text passages, chapters that surpassed a length of about 80 seconds were generated in sections to prevent "LLM laziness", where the API simply would return fewer sentences than requested. Finally, each section, subchapter, and chapter is generated consecutively, returning an "ideal" story that would align story locations with real-world POIs if the pre-calculated ETAs were met during the actual ride. The prompts used for the story generation can be found in the Appendix A.2.

3.5 Real-time Adjustment

Traveling from one place to another is a time-consuming process that is impossible to predict exactly. When traveling by car, unpredictable events such as traffic jams, red lights, and accidents can delay the journey for an indefinite time. Using public transport can be even more unpredictable due to missed connections or train delays. With our context-synchronized storytelling system being intended for mobile use, simply transferring the pre-generated story from subsection 3.4 to the real world would quickly lead to asynchronicity between the story and the real-world POIs. Therefore, the story needs to be adjusted in real-time according to updated information about the user's environment. The synchronicity between the real world and the story is achieved by calculating and comparing two values: The ETA (in seconds) at the real-world POI and the estimated time to mention (ETM, in seconds) of the POI in the story. The difference between the ETA and the ETM of the next POI is periodically calculated during a journey. The ETA is updated every five seconds by recalculating the route from the user's current position to the next POI using the Microsoft Azure Maps API[11]. If the difference to the current ETM surpasses a positive or negative threshold, the system edits the current (sub)chapter.

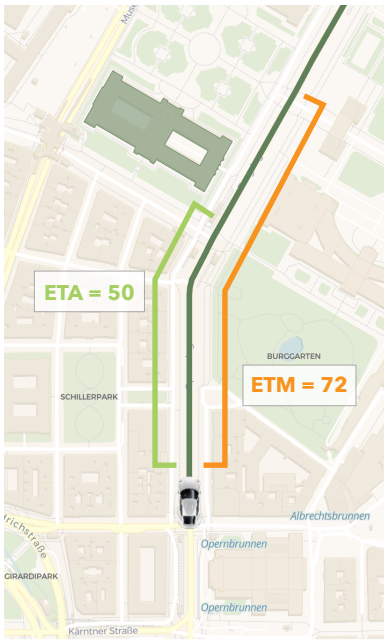
To accomplish this, we have the LLM re-generate the rest of the current passage anew with a number of sentences that will result in a text that implements the desired ETM. We keep track of the sentences in a passage that were already read and edit the remaining text after a padding of two sentences to avoid pauses due to latencies. For this system to work, we utilize a few averages to calculate the number of sentences we need to generate in order to end up with the desired read time. First, we calculated the talking speed of the text-2-speech API in words per second (w_{second}) and the average words per sentence in texts generated by the LLM ($w_{sentence}$). With these averages, we could calculate the average length of a sentence in seconds: $l_{sentence} = w_{sentence}/w_{second}$ When $diff = |ETM - ETA| > l_{sentence}$ we trigger an edit for the current passage and shorten or infill it by $\lfloor diff/l_{sentence} \rfloor$ sentences. We round down because it is more tolerable to come up short in our use case than to overshoot. While the overall generation times varied depending on the amount of text, they never exceeded a duration of roughly seven seconds. An example of the editing method can be found in Figure 5.

Routing APIs like Azure Maps are usually able to predict ETAs accurately to the minute but not to the second. Because our system preemptively edits the current text, we can tolerate minor discrepancies and we also tolerate short pauses right before a POI is approached. It is important, though, that the passage mentioning a POI starts directly before the POI is approached, otherwise it might be missed by the user. To guarantee that timing, we not only consult the routing ETA but also the geographical distance between the GPS position and the POI when the user is in close range. When the distance drops below a certain threshold (0,25km), and the ETA is somewhat small, we can trigger the start of the (sub)chapter mentioning the POI.

4 PRELIMINARY FIELD TEST

In an initial field test in a real-world scenario, the context-synchronized storytelling system was deployed in a car. The goal of this user study (N=5) was to test an in-the-wild deployment of the system under realistic circumstances to record qualitative user feedback as authentic as possible. Therefore, we recruited four participants as drivers of the study vehicle; one additional participant joined

[11] https://azure.microsoft.com/en-us/products/azure-maps, last accessed on 2024/04/03



Max, ever the pragmatist, peered over his laptop, his curiosity piqued by the potential digital footprint of this archaic relic. 5

Elena, her rebellious heart beating in tandem with the diary's secrets, couldn't help but feel the weight of her ancestors' whispers, guiding her from the sepia-toned past. 6

The diary spoke of a hidden treasure, shrouded in the city's cultural tapestry, and protected by riddles that demanded a union of their diverse talents to unravel. 7

~~They exchanged glances, an unspoken agreement passing between them, the thrill of the chase igniting their collective spirit.~~ ~~8~~

~~With the sun casting a gentle glow through the café's stained glass, they stood, ready to step into the labyrinth of Vienna's heart, unaware of the eyes that watched them from the shadows.~~ ~~9~~

~~It was in this moment, under the soft murmur of the Kaffeehaus, that the adventure of a lifetime began, with the city's history as their playing field and the diary as their guide.~~ ~~10~~

Cut content

Their eyes met, a silent pact formed, each aware that the solution to the riddles lay not just in the diary, but in the hidden corners of their beloved Vienna waiting to be deciphered. 8

Edit

Please write the part of the Exposition after the 7. sentence again and make it 1 sentence long.

Figure 5: The context-synchronized storytelling system periodically updates the ETA to the next POI and calculates the difference to the ETM of said POI in the story. Should the difference exceed the threshold value of 10.5 seconds, a re-generation of the rest of the story is triggered. In this example the ETA to the next POI is 50 seconds, but the ETM exceeds it by 22 seconds, so the remaining content of the chapter is removed and replaced by shorter story content.

the user evaluation as a passenger. To prevent loud noises from disturbing the storytelling experience and to reduce the cognitive load for the participants, an electric vehicle with an automatic gearbox was selected for the user evaluation. The context-synchronized storytelling system was deployed using a laptop and a mobile phone, with a third device acting as a mobile data hotspot for the system. The study was recorded by two supervisors seated in the back row of the vehicle to provide an undisturbed driving experience.

4.1 Procedure

The study was conducted in a mid-sized town in Germany. All of the participants were volunteers and were not rewarded with compensation. The participants were instructed to follow the navigation cues given to them by one of the supervisors. The route chosen for the experiment mostly crossed an urban environment with five POIs along the route. Before every run, the participants gave their consent to the anonymized recording and processing of any spoken statements. Afterward, certain demographic data was recorded, including gender, age, knowledge of the English language, and educational level. Moreover, the participant's knowledge of the local area was queried, as well as their recent consumption of written books and audiobooks to record general affinity with storytelling and narration. After the participants had finished a 20-minute round trip with the context-synchronized storytelling system, a semi-structured interview was conducted. The interview mostly focused on qualitative feedback by asking for the participant's favorite and least favorite aspects of the system. Further, the participants were asked to recall as many locations from the story as possible, as well as their memory of the story's setting (i.e., weather and time of day). Finally, their opinion of the sound of the voice interface was queried.

4.2 Participants

Out of the N=5 participants, n=2 self-identified as female, and n=3 as male. The average age of the participants was M=27.2, ranging from 19 to 43. All of the participants reported that their knowledge of the English language was between levels B2 and C2. Their overall knowledge of the area was indicated as *not good* (2), *okay* (1), and *good* (2). While three of the participants had read one book in the last twelve months, the other two had read more than 20 books. Similarly, three participants listened *rarely* or *never* to audiobooks, while the other two indicated to do so *often*.

4.3 Results

The overall response to the system was positive. When asked to recall the locations mentioned in the story, three out of the five participants were able to name three out of five POIs. These three locations were highly visible from the road and visually appealing. The other two POIs were only recalled once each, and neither was well visible from the road or attracted much attention. Regarding the weather, participants had a hard time recalling anything about the weather in the story. Finally, the voice was perceived very well by most of the participants; only one person mentioned that the voice sounded "a little stiff".

The participants' first reaction to the system was very positive. They found the concept of including physical real-world locations

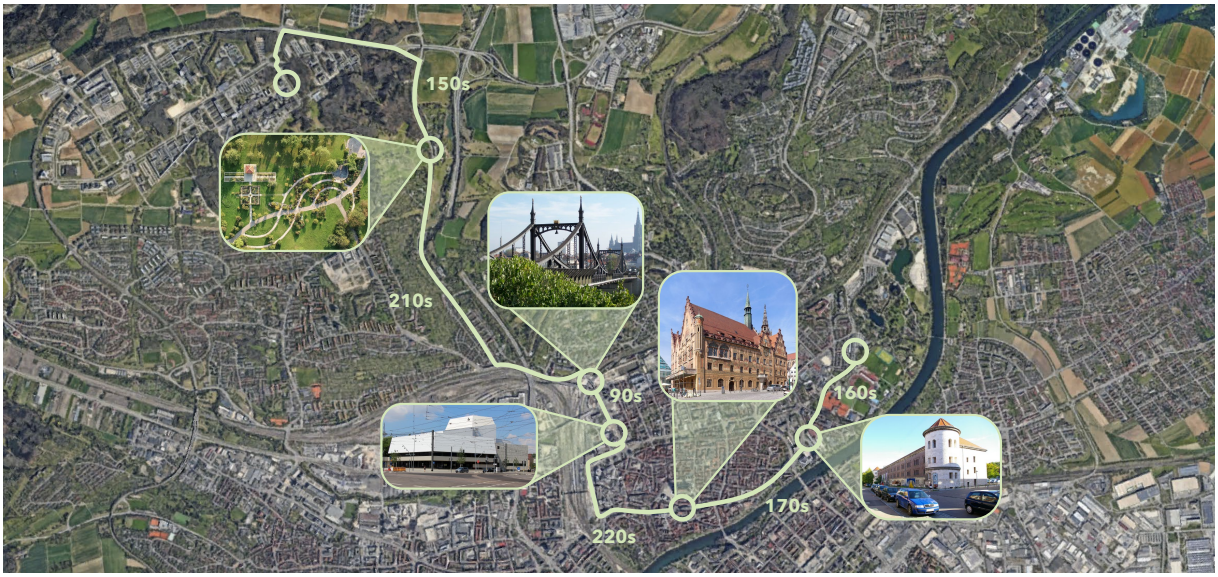


Figure 6: A satellite view of the study route. Along the route, all POIs that were found for the study procedure are highlighted with pictures. Between all of the POIs, the rough ETAs are annotated in seconds for all route segments.

in the virtual story plot "surprising" and promising. Some of the participants reported the moment of realization at the first POI as a "eureka effect" that drew their attention to the story, even if they had been digressing a bit beforehand. A major point of discussion was the issue of driver distraction. While none of the participants complained about being distracted from navigating the vehicle, all of them mentioned having difficulties following the story. Several reasons were stated for this: Since they had to focus on the road, they could not look around freely and hence had problems with finding the mentioned places in the real world. Additionally, some participants claimed to have missed numerous details in the story since they had to focus on the driving task, resulting in an incomplete perception of the storytelling experience. Moreover, when failing to associate fictitious and real elements, participants struggled with following the plot: "If reality and story don't fit together, I can't quite get the world together." Still, all of the participants could figure out the rough plot even when missing some details.

Finally, some weaknesses in the LLM-based approach to automated storytelling were brought up. For example, the plot's lack of motivation was criticized, and participants felt "thrown" into the story. The same also applied to the characters, as they felt shallow and lacked a proper introduction.

5 COMPARATIVE REAL-WORLD STUDY

The preliminary study provided insightful feedback about the initial feedback. Some weaknesses of the system were mentioned, but the main concept of introducing real-world POIs as fictitious elements in the story was met with a highly positive response. Therefore, a larger-scaled real-world user study with N=25 participants was conducted. Because the aforementioned driver distraction prevented participants in the previous study from freely looking around and recognizing locations from the story in the real world, the follow-up field study was conducted with participants in the passenger seat. Moreover, we now compared our context-synchronized storytelling system in a within-subject study design (in the following described as *context-aware condition* with a *baseline condition*), where a simple AI-generated story without any contextual data was read to the participants.

5.1 Procedure

The study aimed to compare two conditions: our context-synchronized storytelling system and a baseline condition. The baseline condition consisted of an AI-generated story without any contextual data. The length of the story, therefore, was fixed, and all of the story locations were fictitious (see Figure 7 for the different locations). The study was conducted in a real-world scenario in a mid-sized town in Germany. Every study run was overseen by two supervisors. The procedure was as follows: After a participant's demographic data was recorded, they were brought to an electric vehicle from our institution. On the way to the vehicle, the participant was given limited information about the system. They were told about the storytelling system being "smart" to a certain degree and that the system was able to include contextual information in the storytelling experience. No specific POIs were mentioned, nor was the existence of a baseline system explained to avoid biasing the participant. Having taken a seat in the front passenger seat, one of the supervisors navigated the vehicle along a fixed route with an average duration of 25 minutes. A second study supervisor was sitting in the backseat row and initiated the respective study condition. During the journey, one of the two conditions was deployed. Arriving at the destination, the participant had to fill out a digital survey regarding the experience during the ride. Afterward, the same route was taken back to the starting point, with the remaining

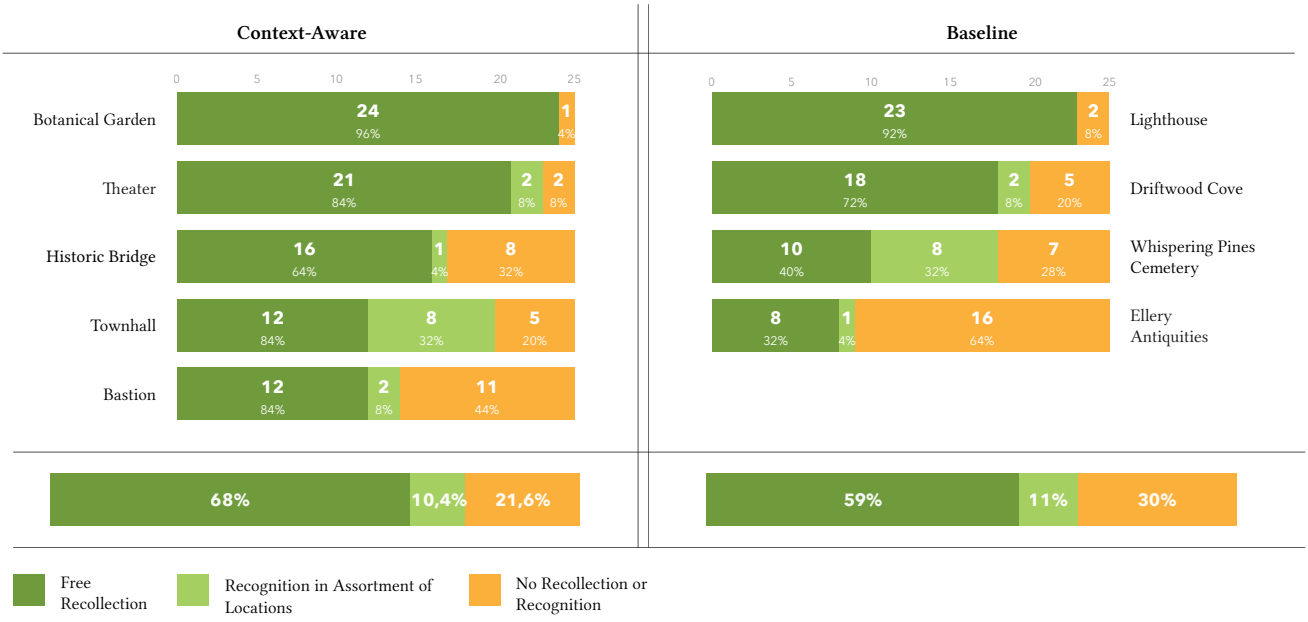


Figure 7: After every story session, we measured participants' ability to recall or recognize locations mentioned in the story (The exact names of the locations were anonymized.). We compared results for the context-aware condition (left) and the baseline (right) and found that participants' recollection was significantly better with the context-aware story (p=0.043).

conditions deployed during the journey. Finally, the post-experience questionnaire was queried again, concluding a full study run. The full route with five POIs along the way can be seen in Figure 6. This route was driven in both directions, with one condition per ride. The two conditions were counterbalanced, i.e., the context-aware condition and the baseline condition were deployed in alternating directions of the route to prevent fatigue or bias from impacting the study results. The goal of the baseline comparison was to identify the delta between Story-Drivenand a practical, real-world use case such as an audiobook. Still, we had to consider the characteristics of AI-generative storytelling (see section 7). Therefore, the baseline condition consisted of a pre-generated story of fixed length (25 minutes, i.e., the average travel duration for the study route) that did not interact with any contextual information. Hence, in some cases, the baseline story had to be interrupted when arriving at the destination, which was not the case for our context-synchronized storytelling system.

5.2 Measures

Similar to the previous driver study, demographic data was collected along with English language level, geographical knowledge, books read, and audiobooks listened to. No physiological data was collected since any changes in, e.g., heart rate or skin conductance, are susceptible to uncontrolled external events, such as traffic events or unpredictable sudden stops of the vehicle. The post-experience questionnaire that was filled out by the participant after every condition measured numerous subjective variables. First, the participants were asked about their estimation of the journey's duration (to be indicated in minutes). This was followed by the Narrative Engagement Scale (NES) [8], a scale that measures the participant's immersion in the story and that consists of four subscales: narrative understanding, attentional focus, narrative presence, and emotional engagement. The NES has been developed and validated with film and television in mind, but already has been successfully transferred to audiobooks [36]. In addition to the NES, participants were asked to recall freely all of the locations they could identify in the story. In a second step, they were asked to indicate all locations from the story given a list of all possible locations. Similarly, the participants were asked to recollect the weather during the story to check for alignment with the weather from the real world. To validate the overall listening experience, the short version of the User Experience Questionnaire (UEQ-S) by Schrepp et al. [41] was used. Moreover, the perception of the narrator's voice was queried. Finally, the participants could enter qualitative feedback when they were asked what they liked and disliked about the system.

5.3 Hypotheses

Considering the nature of our baseline comparison, we can make the assumption that our context-synchronized storytelling system will deliver *better* results than the fixed, pre-generated baseline. Therefore, we establish the following hypotheses:

(1) *H1:* The context-synchronized storytelling system creates a more immersive narrative than the baseline.
(2) *H2:* The context-synchronized storytelling system achieves a better user experience than the baseline.

(3) *H3:* The perceived duration of traveling with the context-synchronized storytelling experience is shorter than with the baseline experience, resulting in an increased loss of sense of time.

5.4 Results

All of the participants volunteered to take part in the study, with no compensation given as a reward. Of the 25 quantitative datasets recorded, one had to be discarded due to data corruption. The quantitative data is evaluated using dependency analyses. Except for the perceived duration of the condition, all records were normally distributed. Because of the sample dependence and the one-directionality of our hypotheses, the quantitative data is checked for significant differences using one-sided paired t-tests. All of the significant differences are reported together with *Cohen's d* for effect size.

Demographic. The participants had an average age of M=26.8 years (SD=4.93), ranging from 22 to 47. 36% of participants self-identified as female, 64% as male. The majority of participants held a college degree (75.0%), four finished college, and two finished vocational training. More than half of all the participants were university students (n=14) and most of the others stated to be employed (n=9). The English language level of the participants ranged from B1 to C2, with the majority (n=14) classifying themselves as C1. Most participants described their knowledge of the study town as "familiar" (n=17), six reported having "little knowledge", and two knew the area "like home". Finally, as already examined in the exploratory driver study (section 4), the participants' affinity with books and audiobooks was queried. On average, the participants read up to 15 books in the last twelve months (M=3.52, SD=3.61). Nine participants reported that they listened to audiobooks "often", seven indicated "sometimes", and five indicated "always". The remaining four participants listened "rarely" or "never" to audiobooks.

Perception of Contextual Information. The free recollection of locations from the story was significantly higher in the context-aware condition than in the baseline condition (p=0.043, medium effect size d-0.73): Participants in the context-aware condition were able to freely recall significantly more locations from the story (M=0.68, SD=0.23) than in the baseline condition (M=0.59, SD=0.20). While the recognition of story locations in a given assortment did not reveal any significant differences, the amount of locations that have not been recognized at all is much lower in the context-aware condition (M=0.22) than in the baseline condition (M=0.30). For a visual description of the location recollection and recognition, see Figure 7. When asked about the story's weather, 76% of the participants correctly recalled the weather in the story that was aligned with the real weather outside.

Narrative Engagement & User Experience. The four different subscales of the NES showed multiple significant differences. While the mean values for the context-aware condition were higher than for the baseline, the subscales *attentional focus* and *narrative presence* showed significant differences. For *attentional focus* (p<0.001, large effect size d=-1.83), the context-aware condition scored significantly higher (M=4.01, SD=1.41) than the baseline condition (M=3.59, SD=1.50). Additionally, the *narrative presence* subscale

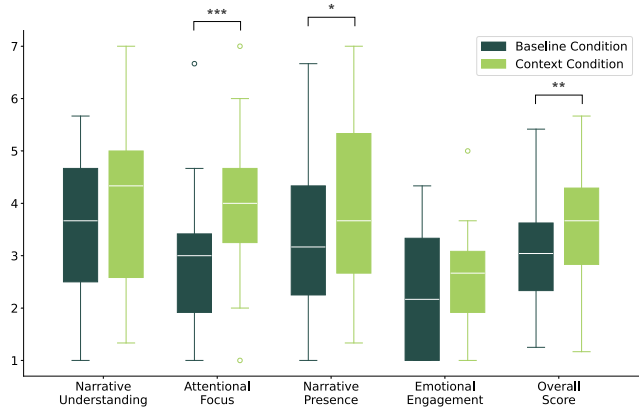


Figure 8: The results of the narrative engagement scale revealed several significant differences for the context-aware condition (light green). The subscales *attentional focus* and *narrative presence* were rated significantly higher than in the baseline condition (dark green).

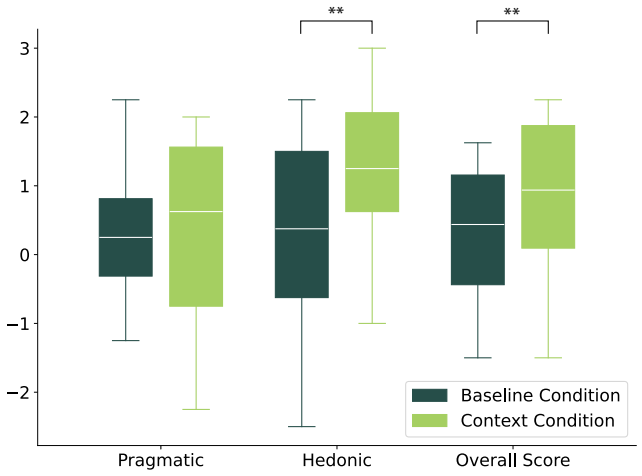


Figure 9: The UEQ-S scores for the baseline (dark green) and the context-aware condition (light green). The hedonic quality was significantly higher (p=0.003) in the context-aware condition than in the baseline condition, which was also reflected in the overall UEQ-S score (p=0.009).

(p=0.04, medium effect size d=-0.74) produced significantly larger values in the context-aware condition (M=3.83, SD=1.65) than in the baseline condition (M=3.32, SD=1.40). This is also reflected in the overall NES score, which was significantly higher (p=0.003, large effect size d=-1.24) in the context-aware condition (M=3.60, SD=1.12) than in the baseline condition (M=3.05, SD=0.99). A visual depiction of the NES results can be found in Figure 8.

Similarly, the UEQ-S revealed significant differences between the baseline and the context-aware condition. The subscale of the hedonic score revealed significantly higher values (p=0.003, large effect size d=-1.22) in the context-aware condition (M=1.17, SD=1.20)

than in the baseline condition (M=0.29, SD=1.39). The overall score was significantly higher (p=0.009, large effect size d=-1.06) in the context-aware condition (M=0.82, SD=1.10) compared to the baseline condition (M=0.31, SD=0.95). A box plot displaying the results can be found in Figure 9.

Qualitative Feedback. The participants were asked to write down a few sentences about what they liked and disliked about the system and any other comments they had. This qualitative feedback was analyzed manually and clustered around different aspects associated with it. Most of the feedback was directed towards the context augmentation of the story. Again, the overall reaction was quite positive, as 19 participants indicated that they liked the inclusion of real locations in the story without specifically being asked for them. The context-synchronized storytelling was described as "a unique experience" that made the baseline story feel "boring" in comparison. Two participants explicitly mentioned that they enjoyed the alignment of the weather in the real world and the story. Moreover, "the system made the ride feel like an adventure, which was quite fun", according to one participant. The time-synchronized alignment of real-world POIs and story locations generally worked out well and had a positive impact on the participants: seven participants specifically expressed that the alignment felt "perfect". The overall integration of the system in the ride felt "smooth" and "seamless". Only one participant actively noted the alignment of the story ending with the end of the ride: "I liked that [the story] ended with the ride and did not stop too early".

Including contextual information was found to increase participants' attention to the story, as four participants mentioned that they listened more actively to the story as a result. Moreover, it helped to draw the participants' attention back to the story when they were distracted: "[...] in instances where I was not paying attention, it brought me back into the story." In comparison, five participants noted that the baseline story lost their attention: "The ride felt a lot longer even though I know the route". Further, an increased cognitive load made following the story more difficult: "Imagining a completely other location, following the story and trying not to get distracted by the actual surroundings was quite demanding". This is also reflected in the story immersion: some participants noted that "places that were familiar were used, which supported the imagination". One participant even reported being fully immersed in the story through the contextual augmentation: "It felt like I was the one leading the story and going around Ulm to solve the riddle." According to the feedback, the high level of immersion made it more exciting and easier to listen. Four participants also reported that the context-synchronized storytelling system promoted their geospatial awareness during the ride.

A major point of criticism was the story's writing style. More than half of the participants (n=13) reported it to be "flowery" and "very descriptive." Four participants had trouble understanding the story at times. Moreover, four participants found the formulation of phrases to be repetitive. Three participants also stated that the writing felt "artificial" and "not human-made." This could have been improved by integrating more dialogues between the characters, according to two participants. The writing style also impacted the understanding of the story content, as eleven participants noted the story to become confusing at times, making it harder to follow the plot. This was also caused by a rushed character introduction, resulting in lackluster motivated characters that are hard to empathize with, according to ten participants. Three participants also complained that the pacing of the story was off; however, opinions differed as to whether it was too fast or too slow. Finally, the story content was often considered boring (n=6), both in the context-aware condition and the baseline condition.

The voice did not interfere with the participant's perception of the story. Most liked the voice (n=14) and reported it as easy to listen to (n=11). Seven participants found the voice to be an appropriate choice for the use case. While some perceived the voice as a bit "drowsy" (n=6), the overall acceptance was quite high, and no participant felt bothered or even distracted. One disadvantage of the English TTS API was the poor pronunciation of foreign words, as noted by two participants. Further comments regarded the use of music and sound effects to underline the plot and to add a few breaks between the chapters. Four participants also wished for personalization of the narrated content: While they liked the overall system, they had a hard time getting used to the story genre and would have preferred to choose their own plot type.

6 DISCUSSION

6.1 Hypotheses

Having found several significant differences between the context-synchronized storytelling experience and the baseline system, arguments can be made for the hypotheses proposed in subsection 5.3.

H1: Story Immersion. The Narrative Engagement Score revealed significantly higher values for the subscales of attentional focus and narrative presence. Particularly the attentional focus showed a three-star significance with a large effect size, which matches a lot of the qualitative feedback. Participants mentioned being drawn back to the story through the inclusion of real-world locations, which could explain an increased attentional focus. The same applies to narrative presence: The participants stated that they were much more present in the story world as a result of merging the virtual and the real world, leading to an improved narrative presence. Therefore, we accept *H1*.

H2: User Experience. The context-synchronized storytelling experience achieved significantly better results in the UEQ-S than the baseline experience. The participants' feedback further confirms this, as the majority of them responded positively to the inclusion of real-world locations in the story. Particularly the temporal synchronization of the locations and the real world was praised as a smooth and seamless integration. Moreover, the experience was described as unique and enjoyable. Therefore, we accept *H2*.

H3: Perceived Travel Duration. Focusing on other things while traveling can make journeys appear much shorter than they actually are. The results of the context-synchronized storytelling system regarding the NES and the UEQ-S, therefore, suggest a loss of sense of time, as the attentional focus is significantly higher than in the baseline condition. However, when asked for an estimation of their travel time, no significant differences were found between the two conditions. Therefore, we reject *H3*.

6.2 User Perspective

The insights gained in the preliminary study (see section 4) compared to the results of the comparison study (section 5) reveal interesting findings about the role of the user. The context-synchronized storytelling experience is based on the visual-auditive association of physical real-world places and narratively mentioned locations in the virtual story world. Therefore, the user is required to have the freedom of looking around. This freedom is not always given to the driver of a vehicle, who needs to focus on the road in ambiguous traffic situations, which limits the system for users operating a vehicle. The early feedback from the preliminary study supports this hypothesis, as many participants stated having missed important details of the story.

These findings also apply to other, more subtle contextual information: The adjustment of weather and time of day were barely noted by many study participants, particularly by drivers participating in the preliminary study. It should be noted that this may have been influenced by the language barrier, as the story was told in English, which was not the native language of any of the participants. However, it can be concluded that the context-synchronized storytelling system is better suited for passengers in a moving vehicle. This brings the topic of mobile, context-synchronous storytelling into the domain of other mobility solutions, such as automated driving and public transportation.

These results also support a second conclusion: The visual appearance of POIs is important for the visual-auditory association process. Both the preliminary and the comparison studies show signs that support this statement: POIs with more visual appeal (such as palaces or churches) were recognized more often than other POIs in the preliminary study. Similarly, in the comparison study, the theater (a large building, see Figure 6) is recognized much more often than the bastion. Therefore, the system could be optimized by including a "visual rating" of POIs in the selection process. However, the constraint remains that in areas with more visually appealing locations, the context-synchronized storytelling experience will be more impactful than in other areas.

7 LIMITATIONS AND FUTURE WORK

The overall concept of context-synchronized storytelling was very well received in our user studies. However, AI-generated storytelling has its limitations and has a long way to go before it can match human storytelling.

The flowery and repetitive writing style and the sometimes repeated patterns applied to several chapters can cause users to lose interest in the story. This also impacts the scalability of generated storytelling experiences, as repetitive phrasing will increasingly tire the user over time. For future research, we would like to investigate whether a generative storytelling system like Story-Driven leads to continuous use, or whether it is used only once, such as during a user's first experience in a new location. The implementation of contextual information in mobile systems always depends on the availability of data. Low POI densities in different locations can strongly affect the storytelling experience. While the system is robust enough to continue working in these cases, the storytelling experience may simply be degraded. Currently, longer story generation times still prevent the use in spontaneous situations, such as deviating from the original route or a change of destination. However, already during the development phase of Story-Driven, the generation times have drastically decreased, which could make spontaneous story generation feasible within only a couple of years.

Well-known issues of LLMs, such as hallucination and laziness, also apply to storytelling applications. The resulting plot inconsistencies and abrupt topic changes can only be avoided by addressing these common problems at the foundational level of LLMs. This also applies to the issue of model bias: stereotypes and other biases present in the LLM's training data also carry over into the storytelling experience. Even worse, biases can be unconsciously transported through the medium of storytelling. Therefore, it is necessary to further investigate the problem of model bias and find strategies or ways to make such problems transparent to the users. Ultimately, the issue of privacy within context-sensitive AI systems must be addressed. Many best-in-class LLMs, like GPT4, are proprietary and depend on the input of data through APIs. Supplying such a service with extensive contextual data could potentially reveal a vast amount of user information. Moreover, the debate around maintaining intellectual property in machine learning methods heavily affects our system since human authors' stories might be represented in the generation of stories through LLMs [25].

In future work, we especially aim to explore new mobility types. For example, traveling with public transport, in an automated driving setting, or with a bicycle might require different incarnations of our method. By incorporating more personalized preferences, including genres, we expect to better align with a user's interests. Our explorative field study also demonstrates the need to thoroughly examine how this would affect the user's primary task (e.g., steering or paying attention to the surroundings).

Finally, there are societal concerns that need to be examined more closely about the creation of media that is closely tailored to an individual. This can lead to societal polarization as people become less exposed to diverse perspectives.

8 CONCLUSION

We presented Story-Driven, a system for generative storytelling in mobile environments, such as in a vehicle. Story-Driven has the ability to generate stories that incorporate contextual information, such as POIs and the current weather, and weave them into the final story. To achieve the best possible user experience, we introduced a new method for synchronizing the arrival at POIs with the actual utterances in the story itself. We evaluated our approach in two real-world user studies and found that compared to a baseline system, Story-Driven increased the level of immersion in the story and improved the overall user experience. Participants found the user experience to be "unique" and "engaging" with a "eureka moment" when a POI came into view. As the field of machine learning continues to advance at a rapid pace, we believe that new ways to combine the real world with fictional stories will be a very exciting area to explore, not only for audiobooks but also for other types of media such as videos or even Mixed Reality.

REFERENCES

[1] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. 1999. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium,*

HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1. Springer, 304–307.

[2] Mieke Bal. [n. d.]. *Narratology: introduction to the theory of narrative* (3. ed ed.). University of Toronto Press.

[3] Matthias Baldauf, Schahram Dustdar, and Florian Rosenberg. 2007. A survey on context-aware systems. *International Journal of ad Hoc and ubiquitous Computing* 2, 4 (2007), 263–277.

[4] Roland Barthes and Roland Barthes. [n. d.]. *S-Z* (nachdr ed.). Number 687 in Suhrkamp-Taschenbuch Wissenschaft. Suhrkamp.

[5] Nina Begus. 2023. Experimental Narratives: A Comparison of Human Crowdsourced Storytelling and AI Storytelling. *arXiv preprint arXiv:2310.12902* (2023).

[6] David Bethge, Daniel Bulanda, Adam Kozlowski, Thomas Kosch, Albrecht Schmidt, and Tobias Grosse-Puppendahl. 2024. HappyRouting: Learning Emotion-Aware Route Trajectories for Scalable In-The-Wild Navigation. arXiv:2401.15695 [cs.HC]

[7] Arthur Brisbane. [n. d.]. Newspaper Copy That People Must Read, Advertising's Relation to the Growth of Reading Ability—the Thunderstorm and "Yellow" Journalism—an Example of the Power of Comparison in Writing. ([n. d.]), 17.

[8] Rick Busselle and Helena Bilandzic. 2009. Measuring narrative engagement. *Media psychology* 12, 4 (2009), 321–347.

[9] Guanling Chen and David Kotz. 2000. A survey of context-aware mobile computing research. (2000).

[10] Zexin Chen, Eric Zhou, Kenneth Eaton, Xiangyu Peng, and Mark Riedl. 2023. Ambient Adventures: Teaching ChatGPT on Developing Complex Stories. *arXiv preprint arXiv:2308.01734* (2023).

[11] Haoran Chu and Sixiao Liu. [n. d.]. Can AI Tell Good Stories? Narrative Transportation and Persuasion with ChatGPT. ([n. d.]). https://psyarxiv.com/c3549/download?format=pdf

[12] John Joon Young Chung, Wooseok 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* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. https://doi.org/10.1145/3491102.3501819

[13] Andy Coenen, Luke Davis, Daphne Ippolito, Emily Reif, and Ann Yuan. 2021. Wordcraft: a human-ai collaborative editor for story writing. *arXiv preprint arXiv:2107.07430* (2021).

[14] Lajos Matyas Csepregi. 2021. The effect of context-aware llm-based npc conversations on player engagement in role-playing video games. *Unpublished manuscript* (2021).

[15] Anind K Dey. 1998. Context-aware computing: The CyberDesk project. In *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments*. AAAI Press Menlo Park, CA, 51–54.

[16] Fiona Draxler, Daniel Buschek, Mikke Tavast, Perttu Hämäläinen, Albrecht Schmidt, Juhi Kulshrestha, and Robin Welsch. 2023. Gender, age, and technology education influence the adoption and appropriation of LLMs. *arXiv preprint arXiv:2310.06556* (2023).

[17] Panagiotis Fotaris, Theodoros Mastoras, and Petros Lameras. 2023. Designing Educational Escape Rooms With Generative AI: A Framework and ChatGPT Prompt Engineering Guide. In *17th European Conference on Games Based Learning*.

[18] Louie Giray. 2023. Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering* (2023), 1–5.

[19] Mustafa Can Gursesli, Pittawat Taveekitworachai, Febri Abdullah, Mury F Dewantoro, Antonio Lanata, Andrea Guazzini, Van Khôi Lê, Adrien Villars, and Ruck Thawonmas. 2023. The Chronicles of ChatGPT: Generating and Evaluating Visual Novel Narratives on Climate Change Through ChatGPT. In *International Conference on Interactive Digital Storytelling*. Springer, 181–194.

[20] Anton Gustafsson, John Bichard, Liselott Brunnberg, Oskar Juhlin, and Marco Combetto. [n. d.]. Believable Environments: Generating Interactive Storytelling in Vast Location-Based Pervasive Games. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (New York, NY, USA, 2006-06-14) *(ACE '06)*. Association for Computing Machinery, 24–es. https://doi.org/10.1145/1178823.1178852

[21] Chi-yang Hsu, Yun-Wei Chu, Ting-Hao Huang, and Lun-Wei Ku. [n. d.]. Plot and Rework: Modeling Storylines for Visual Storytelling. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (Online, 2021-08), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, 4443–4453. https://doi.org/10.18653/v1/2021.findings-acl.390

[22] Ting-Hao Kenneth Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh, Lucy Vanderwende, Michel Galley, and Margaret Mitchell. [n. d.]. Visual Storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (San Diego, California, 2016-06), Kevin Knight, Ani Nenkova, and Owen Rambow (Eds.). Association for Computational Linguistics, 1233–1239. https://doi.org/10.18653/v1/N16-1147

[23] Catherine Kanellopoulou, Katia Lida Kermanidis, and Andreas Giannakoulopoulos. [n. d.]. The Dual-Coding and Multimedia Learning Theories: Film Subtitles as a Vocabulary Teaching Tool. 9, 3 ([n. d.]), 210. https://doi.org/10.3390/educsci9030210 Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[24] Mohamed Kari, Tobias Grosse-Puppendahl, Alexander Jagaciak, David Bethge, Reinhard Schütte, and Christian Holz. [n. d.]. SoundsRide: Affordance-Synchronized Music Mixing for In-Car Audio Augmented Reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event USA, 2021-10-10). ACM, 118–133. https://doi.org/10.1145/3472749.3474739

[25] Taewook Kim, Hyomin Han, Eytan Adar, Matthew Kay, and John Joon Young Chung. [n. d.]. Authors' Values and Attitudes Towards AI-bridged Scalable Personalization of Creative Language Arts. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA, 2024-05-11). ACM, 1–16. https://doi.org/10.1145/3613904.3642529 arXiv:2403.00439 [cs]

[26] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems* 35 (2022), 22199–22213.

[27] Vikram Kumaran, Jonathan Rowe, Bradford Mott, and James Lester. 2023. SCENECRAFT: automating interactive narrative scene generation in digital games with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 19. 86–96.

[28] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A Alghamdi, et al. 2024. A Design Space for Intelligent and Interactive Writing Assistants. *arXiv preprint arXiv:2403.14117* (2024).

[29] Zhiyu Lin and Mark Riedl. 2021. Plug-and-blend: A framework for controllable story generation with blended control codes. *arXiv preprint arXiv:2104.04039* (2021).

[30] Richard E. Mayer. [n. d.]. *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164603 Pages: xi, 210.

[31] Munan Ning, Yujia Xie, Dongdong Chen, Zeyin Song, Lu Yuan, Yonghong Tian, Qixiang Ye, and Li Yuan. 2023. Album Storytelling with Iterative Story-aware Captioning and Large Language Models. *arXiv preprint arXiv:2305.12943* (2023).

[32] Allan Paivio. [n. d.]. Dual coding theory: Retrospect and current status. 45, 3 ([n. d.]), 255–287. https://doi.org/10.1037/h0084295 Place: Canada Publisher: Canadian Psychological Association.

[33] Jeongyoon Park, Jumin Shin, Gayeon Kim, and Byung-Chull Bae. 2023. Designing a Language Model-Based Authoring Tool Prototype for Interactive Storytelling. In *Interactive Storytelling: 16th International Conference on Interactive Digital Storytelling, ICIDS 2023, Kobe, Japan, November 11–15, 2023, Proceedings, Part II* (Kobe, Japan). Springer-Verlag, Berlin, Heidelberg, 239–245. https://doi.org/10.1007/978-3-031-47658-7_22

[34] Bastian Pfleging, Maurice Rang, and Nora Broy. [n. d.]. Investigating User Needs for Non-Driving-Related Activities during Automated Driving. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (New York, NY, USA, 2016-12-12) *(MUM '16)*. Association for Computing Machinery, 91–99. https://doi.org/10.1145/3012709.3012735

[35] Hannah Rashkin, Asli Celikyilmaz, Yejin Choi, and Jianfeng Gao. 2020. Plotmachines: Outline-conditioned generation with dynamic plot state tracking. *arXiv preprint arXiv:2004.14967* (2020).

[36] Daniel C Richardson, Nicole K Griffin, Lara Zaki, Auburn Stephenson, Jiachen Yan, Thomas Curry, Richard Noble, John Hogan, Jeremy I Skipper, and Joseph T Devlin. 2020. Engagement in video and audio narratives: Contrasting self-report and physiological measures. *Scientific Reports* 10, 1 (2020), 11298.

[37] Alejandro Rivero-Rodriguez, Paolo Pileggi, and Ossi Antero Nykänen. 2016. Mobile context-aware systems: technologies, resources and applications. *International Journal of Interactive Mobile Technologies* 10, 2 (2016), 25–32.

[38] Marie-Laure Ryan. [n. d.]. *Narrative as virtual reality: immersion and interactivity in literature and electronic media* (transferred to digital print. 2001 - [im kolophon: milton keynes: lightning source, 2010] ed.). Johns Hopkins Univ. Press.

[39] Nick Ryan, Jason Pascoe, and David Morse. 1999. Enhanced reality fieldwork: the context aware archaeological assistant. (1999).

[40] Bill N Schilit and Marvin M Theimer. 1994. Disseminating active map information to mobile hosts. *IEEE network* 8, 5 (1994), 22–32.

[41] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.* (2017).

[42] Nisha Simon and Christian Muise. 2022. TattleTale: Storytelling with Planning and Large Language Models. In *ICAPS Workshop on Scheduling and Planning Applications*.

[43] Nicholas Suwono, Justin Chen, Tun Hung, Ting-Hao Huang, I-Bin Liao, Yung-Hui Li, Lun-Wei Ku, and Shao-Hua Sun. [n. d.]. Location-Aware Visual Question Generation with Lightweight Models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, 2023-12), Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, 1415–1432. https://doi.org/10.18653/v1/2023.emnlp-main.88

[44] Xu Tan, Tao Qin, Frank Soong, and Tie-Yan Liu. 2021. A survey on neural speech synthesis. *arXiv preprint arXiv:2106.15561* (2021).
[45] Ronald B Tobias. 2012. *20 master plots: And how to build them.* Penguin.
[46] Runyu Wang, Keng Leng Siau, and Zili Zhang. 2023. Using AI and ChatGPT in Brand Storytelling. (2023).
[47] Yichen Wang, Kevin Yang, Xiaoming Liu, and Dan Klein. 2023. Improving Pacing in Long-Form Story Planning. *arXiv preprint arXiv:2311.04459* (2023).
[48] Christian Weiß. 2011. V2X communication in Europe–From research projects towards standardization and field testing of vehicle communication technology. *Computer Networks* 55, 14 (2011), 3103–3119.
[49] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).
[50] Polly W. Wiessner. [n. d.]. Embers of society: Firelight talk among the Ju/'hoansi Bushmen. 111, 39 ([n. d.]), 14027–14035. https://doi.org/10.1073/pnas.1404212111 Publisher: Proceedings of the National Academy of Sciences.
[51] Bob G. Witmer and Michael J. Singer. [n. d.]. Measuring Presence in Virtual Environments: A Presence Questionnaire. 7, 3 ([n. d.]), 225–240. https://doi.org/10.1162/105474698565686
[52] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, Vol. 10. CEUR-WS Team.
[53] Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. 2022. Doc: Improving long story coherence with detailed outline control. *arXiv preprint arXiv:2212.10077* (2022).
[54] Kevin Yang, Yuandong Tian, Nanyun Peng, and Dan Klein. [n. d.]. *Re3: Generating Longer Stories With Recursive Reprompting and Revision.* https://doi.org/10.48550/arXiv.2210.06774 arXiv:2210.06774 [cs]
[55] Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019. Plan-and-write: Towards better automatic storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7378–7385.
[56] Min-Hsuan Yeh, Vincent Chen, Ting-Hao Huang, and Lun-Wei Ku. [n. d.]. Multi-VQG: Generating Engaging Questions for Multiple Images. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (Abu Dhabi, United Arab Emirates, 2022-12), Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, 277–290. https://doi.org/10.18653/v1/2022.emnlp-main.19
[57] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–21.

A APPENDIX

A.1 Story Setup Prompts

We used the following prompts for creating a story plan as described in subsection 3.4. First, we generate a premise, then four characters with descriptions, and lastly a plot outline. We use a template pattern [49] to receive parsable output. The plot outline prompt creates a plot outline for all chapters with a one-level hierarchy. The chapter outline prompt is used to further plan subchapters of a chapter if it contains more than one POI. With this output, we build the second level of the plot outline.

Premise Prompt

```
We are planning a GENRE story with this
plot type: PLOT_TYPE.
Please write a premise for such a story in
2-3 sentences.
The story will be a journey along multiple
locations in CITY (COUNTRY) and the
locations are an integral part of the story
structure. The locations are the following:

LOCATIONS_WITH_DETAILS

Please do not use any other locations apart
from the ones I provide to you and
incorporate them all into the story in the
same order as they are provided to you.
Also, consider the following contextual
information:

SITUATIONAL_INFORMATIONS

Premise:
```

Character Prompt:

```
List the names and details of all major
characters for this story, but use a
maximum of 4 characters. Define their
details in one sentence each.

1. Full Name:
Details:
```

Plot Outline Prompt:

```
Outline the main plot points of the story.
Generate one plot point for each of the
following story parts: Exposition, Rising
Action, Peak, Falling Action, and
Resolution. In the Exposition and
Resolution, the story's setting is
somewhere in the city of CITY. Do not use
any of the specific locations I provided
within the Exposition and Resolution. For
the other three chapters, I provide you
with the locations that should be featured
in each one. Generate one point at a time.
Every point should be 1-2 sentences long.
Make sure that the story has a clear ending
in the Resolution.

I am providing you with a template for your
output. The placeholder for the content is
CONTENT. Please preserve the formatting.
This is the template:

Exposition: CONTENT

Rising Action: CONTENT
Locations: LOCATIONS_WITH_DETAILS

Peak: CONTENT
Locations: LOCATIONS_WITH_DETAILS

Falling Action: CONTENT
Locations: LOCATIONS_WITH_DETAILS
```

```
Resolution: CONTENT
```

Chapter Outline Prompt:

```
Please outline the main plot points of the
CHAPTER chapter. Generate one plot point
for each of the locations featured in the
chapter, every point should be 1-2
sentences long.

Premise of the chapter: CHAPTER_PLOT
Locations: LOCATIONS_WITH_DETAILS

Please use this format:

LOCATION: CONTENT
...
```

A.2 Story Generation Prompts

After creating a story plan, we used the following prompt to generate the entire story. The chapter prompt is an example of generating a chapter with one POI. The prompt for generating a subchapter, if the corresponding chapter is divided into subchapters, is equivalent to the chapter prompt. This example prompt also assumes that the chapter is long enough to warrant generating it in sections. When that is not the case, we don't prompt for sections, but rather the whole chapter in one go.

Chapter Prompt:

```
Please write a GENRE story with this plot
type: PLOT_TYPE.

Premise: PREMISE
Characters: CHARACTERS
Outline: PLOT_OUTLINE
Contextual information:
CONTEXTUAL_INFORMATIONS
Summary of the previous chapter: SUMMARY

In the chapter CHAPTER, the story setting
is at this location:
LOCATION_WITH_DETAILS
Do not use any other locations in this
chapter.

Please write the 1. section out of SECTIONS
sections of the CHAPTER chapter in
SENTENCES sentences and make sure that you
divide the plot of the CHAPTER among all
sections.
Generate one sentence at a time and output
them as coherent text. Call it Chapter
CHAPTER_NUMBER, give it a name, and put a
full stop after the title. Only use the
plot described for this chapter in the
outline.
```

The edit prompt is used for editing (sub)chapters during the story session to adjust the timing of the story to the real-world scenario. Again, this prompt assumes that the chapter to be edited was generated in sections.

Edit Prompt:

```
Please write the part after the
INFILL_POSITION. sentence of the
CURRENT_SECTION. section of the CHAPTER
again. The title of a chapter is counted as
a sentence if it is contained in this
section.
Just output the newly generated text that
picks up after the INFILL_POSITION.
sentence of the original text and make it
NEW_SENTENCES sentences long. Generate one
sentence at a time and output them as
coherent text. Keep in mind which
characters were already introduced.
```